

# TCGA MultiOmics Data Analysis

Pablo Riesgo ([priesgo@uoc.edu](mailto:priesgo@uoc.edu))

08/11/2015

# The dataset

## Retrieval

- TCGA normalized level 3 datasets for RNASeq and RPPA
- For tumor types breast cancer and ovary cancer

```
library("RTCGAToolbox")
firehose_datasets = getFirehoseDatasets()
firehose_dates = getFirehoseRunningDates()
BRCA.dataset = getFirehoseData (dataset="BRCA", runDate=firehose_dates)
OV.dataset = getFirehoseData (dataset="OV", runDate=firehose_dates)
```

# The dataset

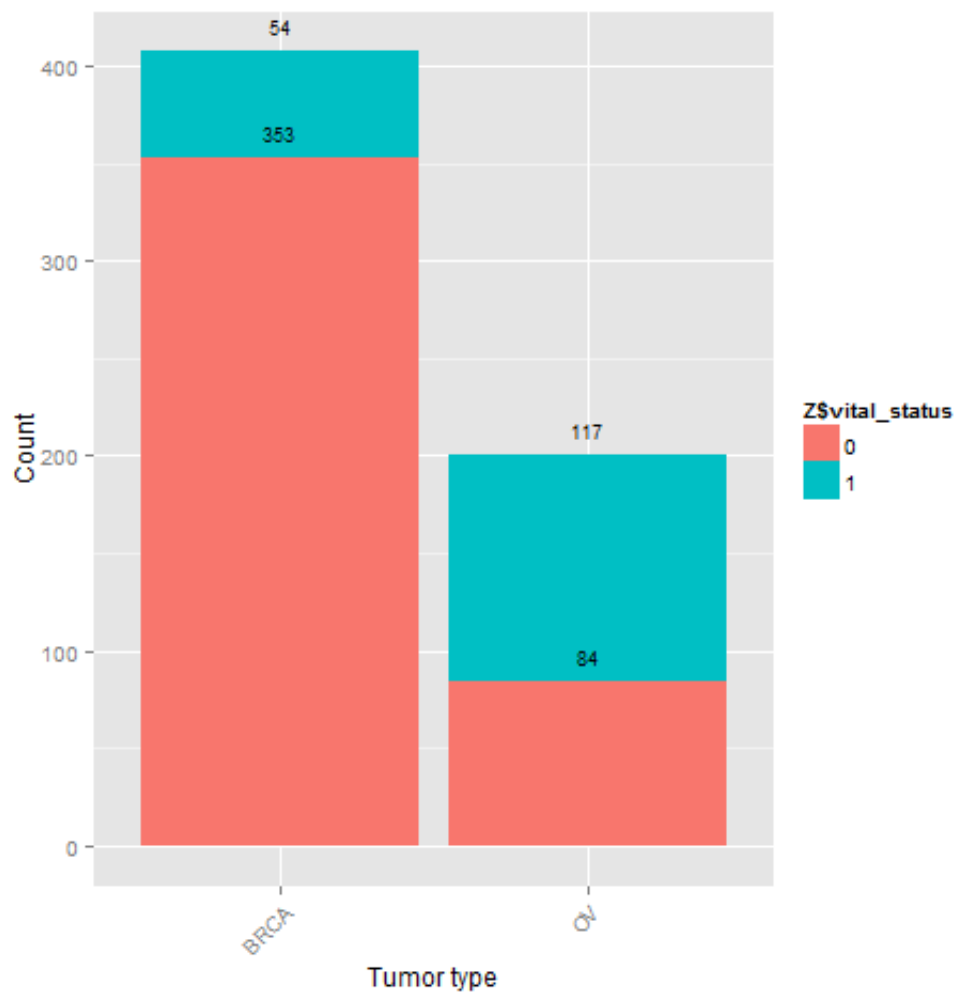
## Preprocessing

- Removed samples not having data for RNASeq and RPPA
- Removed variables for genes not matching with HGNC gene names
- Removed RPPA variables for antibodies matching the same gene, only the antibody with the maximum absolute average expression was conserved
- Scaling
- Removed variables with zero or near zero variance
- Removed variables with missing values
- Removed variables with correlation over 0.7

# The dataset

Samples

407 BRCA + 201 ovary cancer samples



# The dataset

## Variables

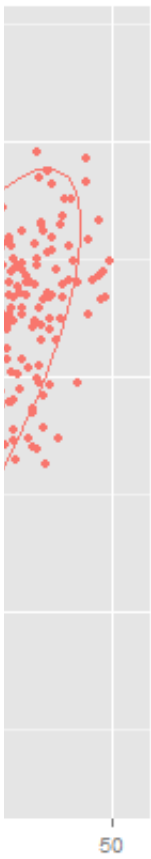
12880 gene expression + 55 protein expression variables are far too many to plot...

# Principal component Analysis

Gene expression



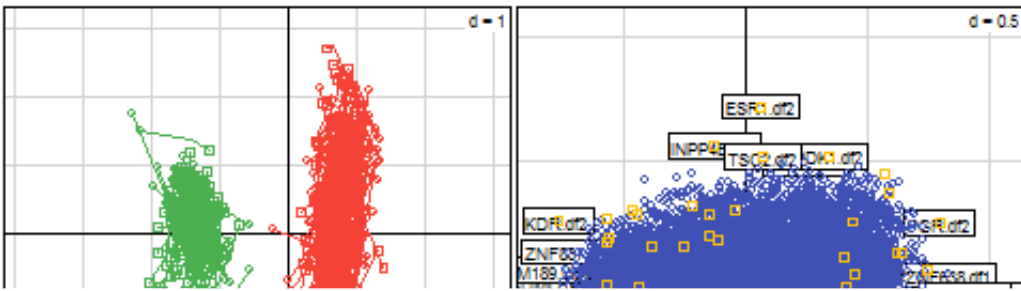
# it Analysis



# Multiple Co-Inertia Analysis (MCIA)

The 2 first PCs explain 34.05% and 11.55% of the variance

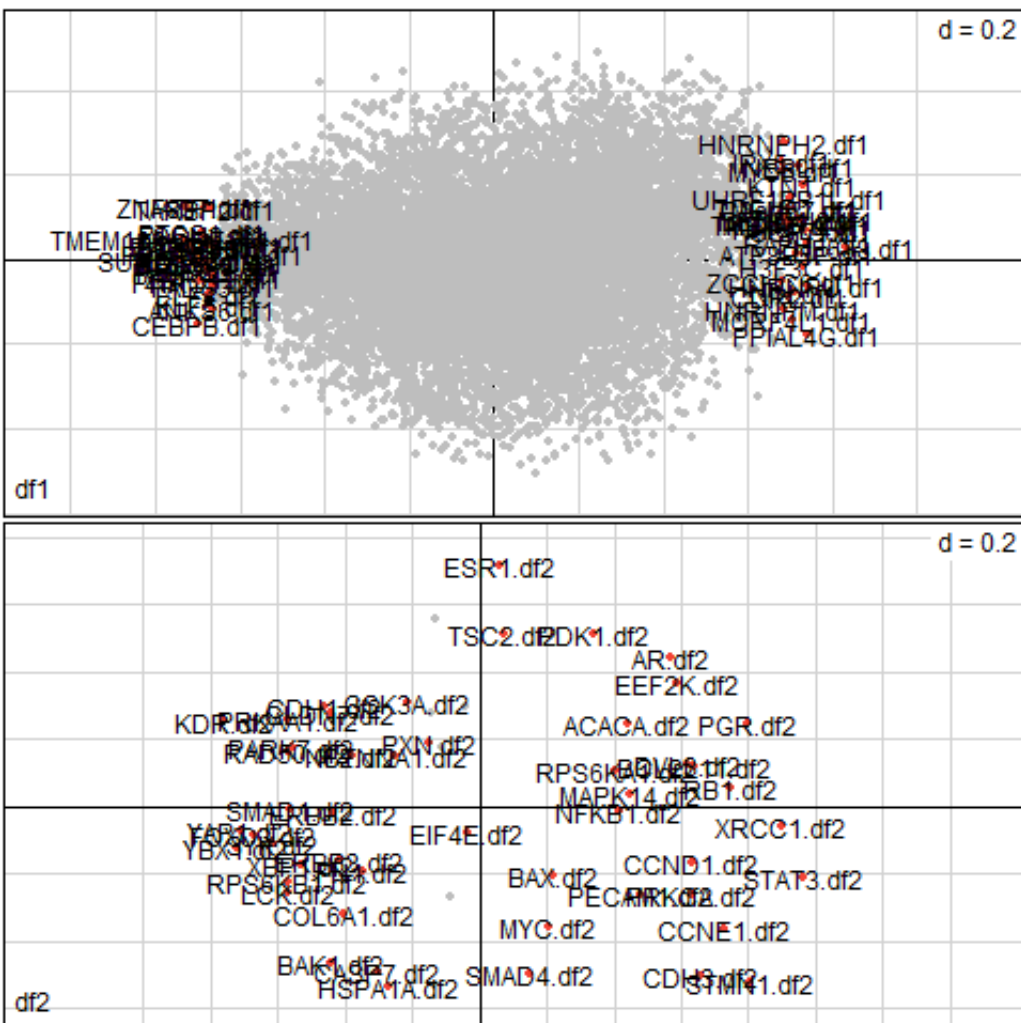




10/18


# Multiple Co-Inertia Analysis (MCIA)

Labeled items are the top 100 variables



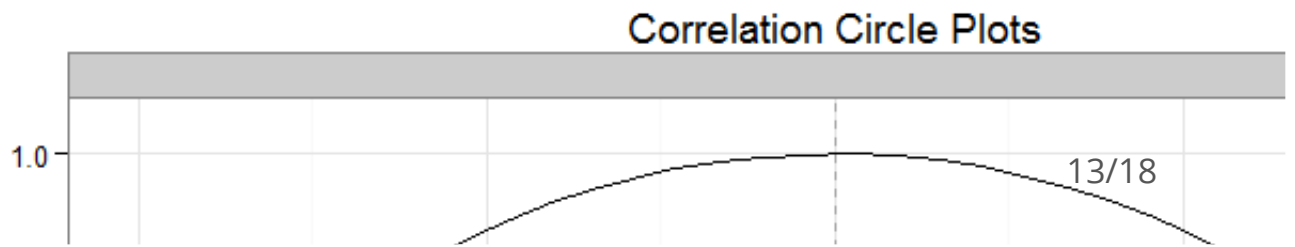
# sparse Least Square Regression (sPLS)

Samples

plotIndiv			
Block: X			
			12/18

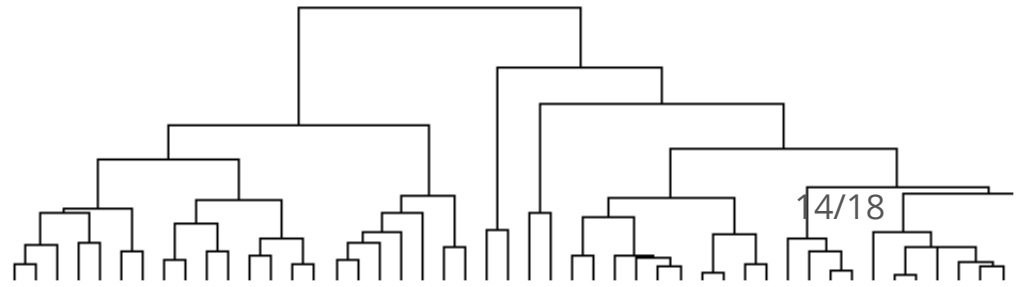
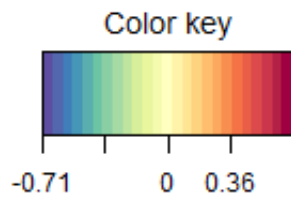
# sparse Least Square Regression (sPLS)

Variables



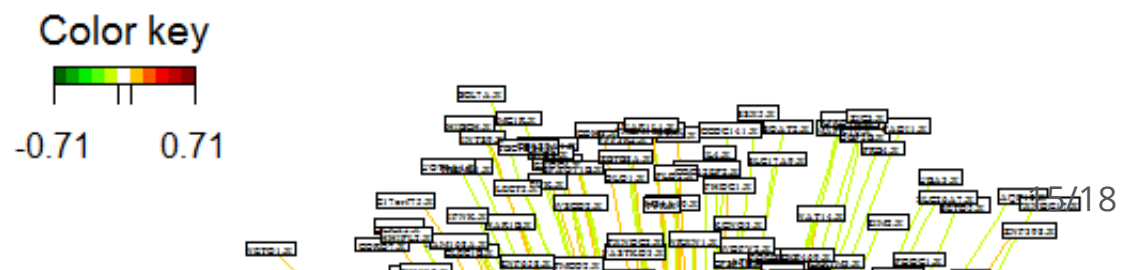
# sparse Least Square Regression (sPLS)

Heatmap



# sparse Least Square Regression (sPLS)

Network



# Results

Matches between MCIA and sPLS top 100 and known oncogenes and tumor suppressor genes



# Results

Matches between MCIA and sPLS enriched pathways in Reactome

# Other methods to study

- Canonical Correlation Analysis
- Regularized Generalized CCA (error with dataset)
- Regularized CCA (long computation)
- CCA with Elastic Net (did not found implementation)
- Procrustes (pending)



