TCGA BRCA SV Analysis

Asad

5/7/2023

```
library(TCGAbiolinks)
library(TCGAmutations)
```

```
## Loading required package: maftools
```

```
## Loading required package: data.table
```

```
library(maftools)
```

```
setwd('E:/TCGA Data Analysis/CNV and SNV Analysis')
```

#Create TCGA object and read maf data

```
tcga_brca=tcga_load(study = "BRCA")
```

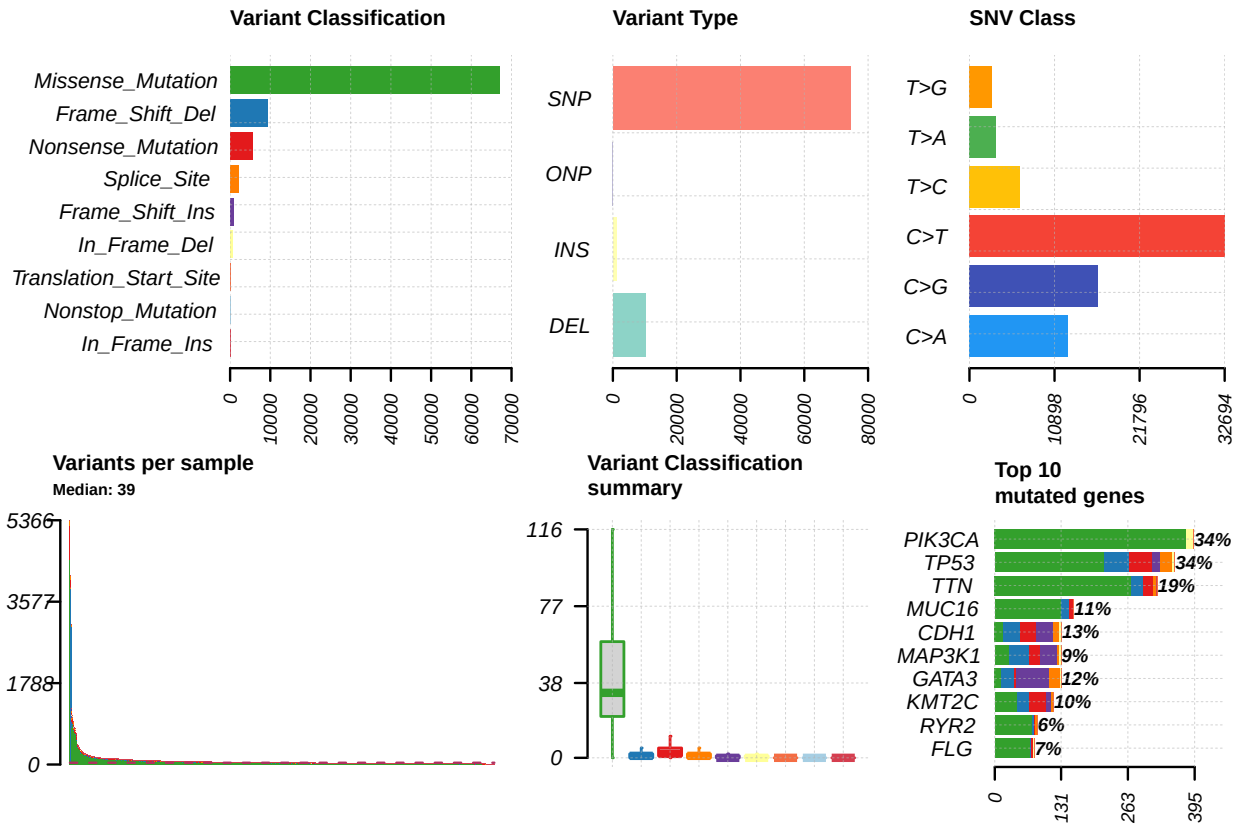
```
## Loading BRCA. Please cite: https://doi.org/10.1016/j.cels.2018.03.002 for reference
```

```
BRCA = read.maf(maf = tcga_brca@data, clinicalData = tcga_brca@clinical.data)
```

```
## -Validating
## --Removed 2639 duplicated variants
## -Summarizing
## --Possible FLAGS among top ten genes:
##   TTN
##   MUC16
##   FLG
## -Processing clinical data
## -Finished in 9.960s elapsed (9.010s cpu)
```

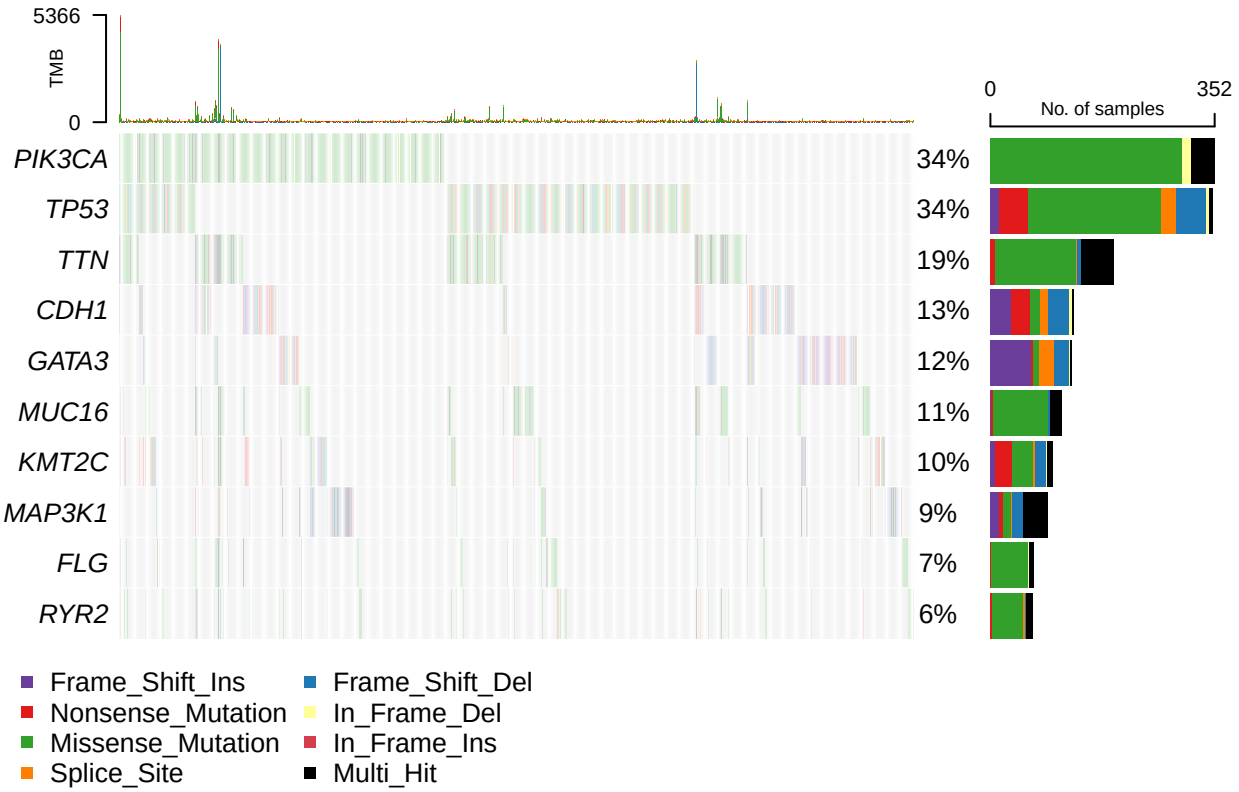
#Take a overview of the data

```
plotmafSummary(maf = BRCA, rmOutlier = TRUE, addStat = 'median', dashboard = TRUE, titvRaw = T)
```



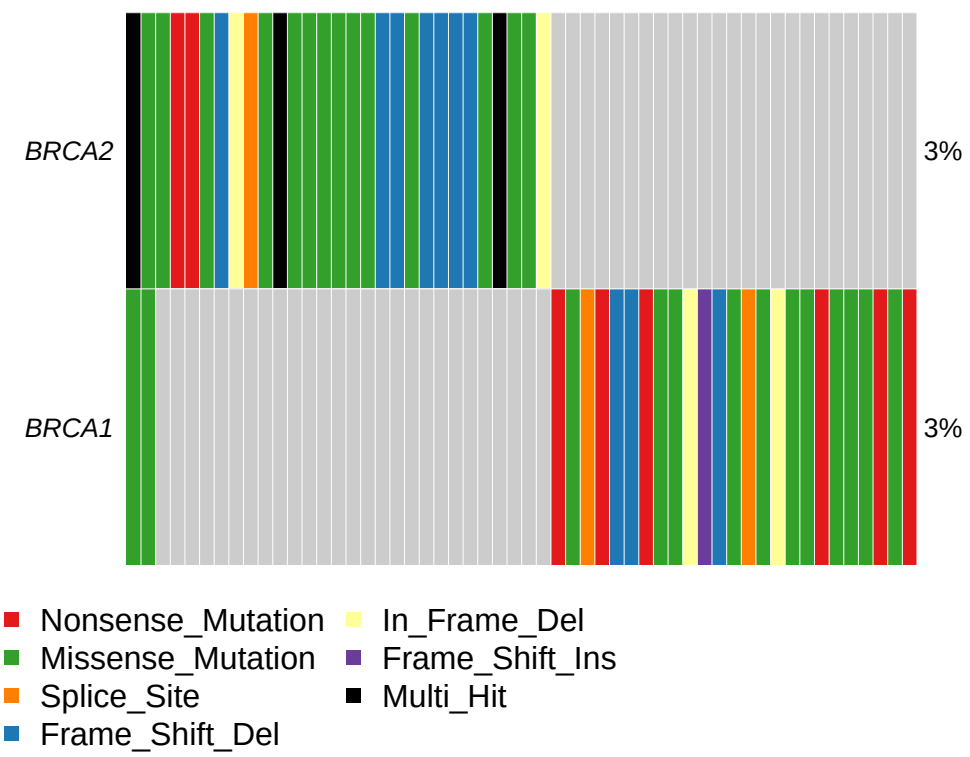
#Make oncoplot of top 10 mutated genes in BRCA

```
oncoplot(maf = BRCA, top = 10)
```



#Let's make an oncoprint of BRCA1 and BRCA2 genes

```
oncostrip(maf = BRCA, genes = c('BRCA1','BRCA2'))
```

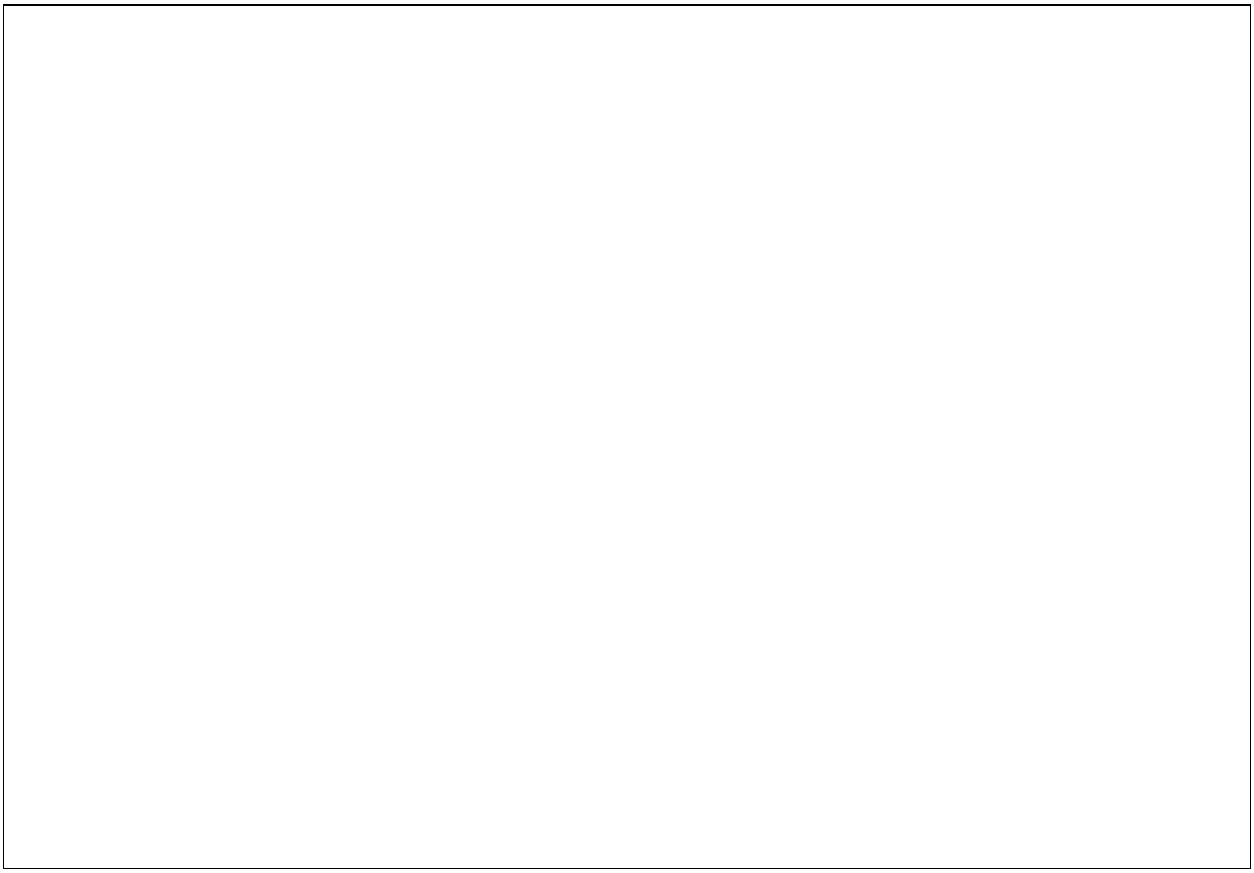



```
dev.off()
```

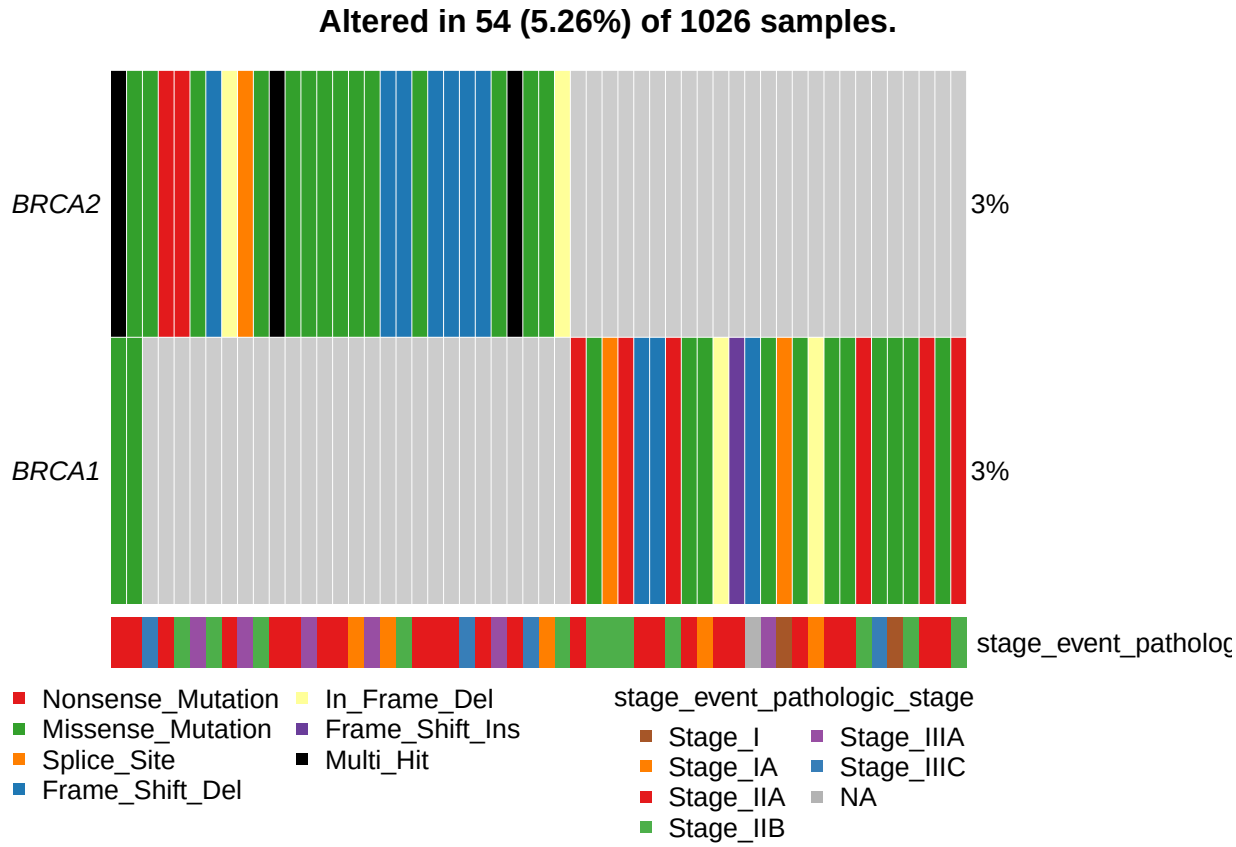
```
## null device
##           1
```

#Let's add mutation type strip

```
library(grid)
grid.rect(gp=gpar(fill=0),vp=viewport(layout.pos.col=500,layout.pos.row=300))
```

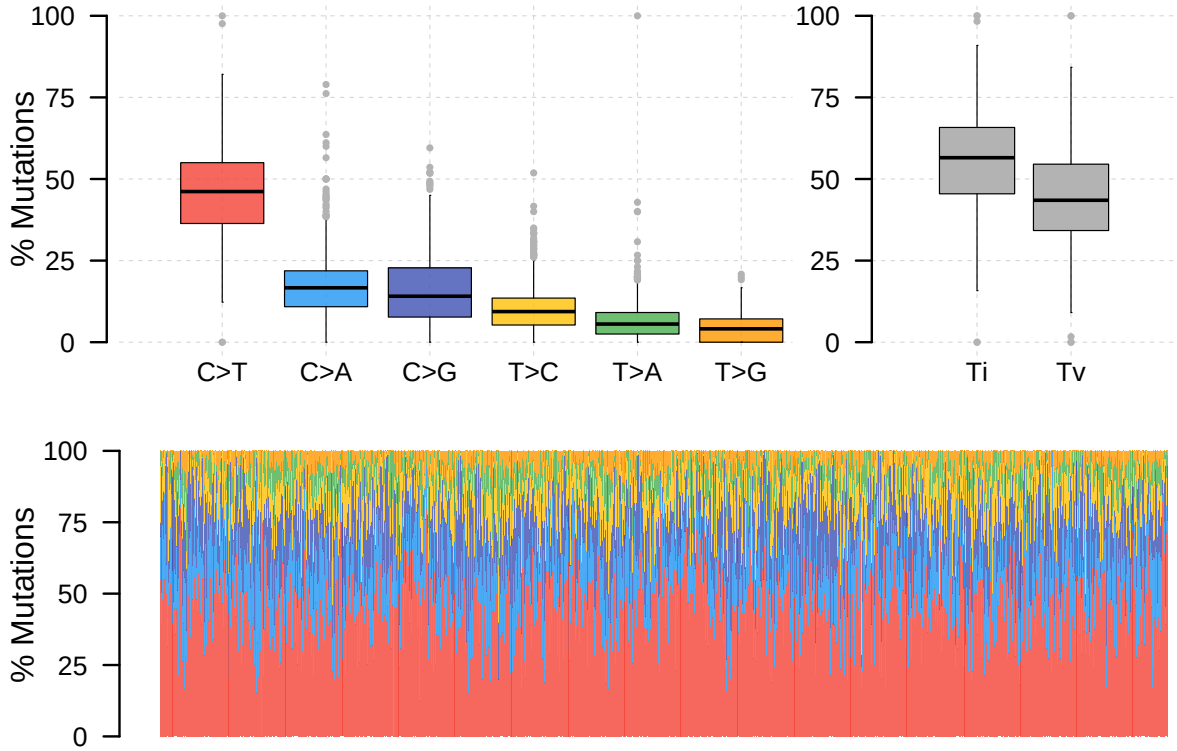


```
oncostrip(maf = BRCA, genes = c('BRCA1','BRCA2'),clinicalFeatures = 'stage_event_pathologic_stage',widt
```



#Let's check the total transition and transersion mutation frequency

```
titv = titv(maf = BRCA, plot = T)
```



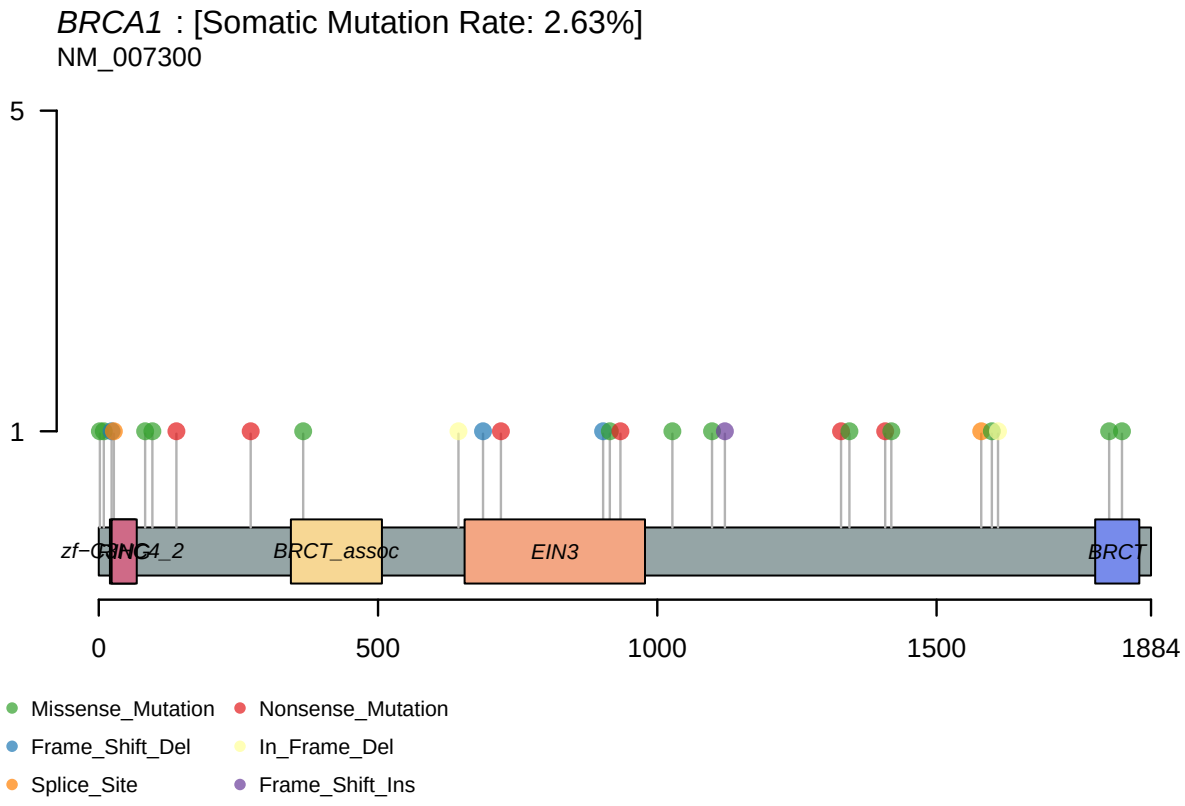
#lollipop plot for BRCA1 gene in breast cancer.

```
lollipopPlot(maf = BRCA, gene = 'BRCA1', AACol = 'HGVSp_Short', showMutationRate = TRUE)
```

```
## 5 transcripts available. Use arguments refSeqID or proteinID to manually specify tx name.
```

```
##     HGNC refseq.ID protein.ID aa.length
## 1: BRCA1 NM_007294  NP_009225      1863
## 2: BRCA1 NM_007297  NP_009228      1816
## 3: BRCA1 NM_007298  NP_009229       759
## 4: BRCA1 NM_007299  NP_009230       699
## 5: BRCA1 NM_007300  NP_009231      1884
```

```
## Using longer transcript NM_007300 for now.
```

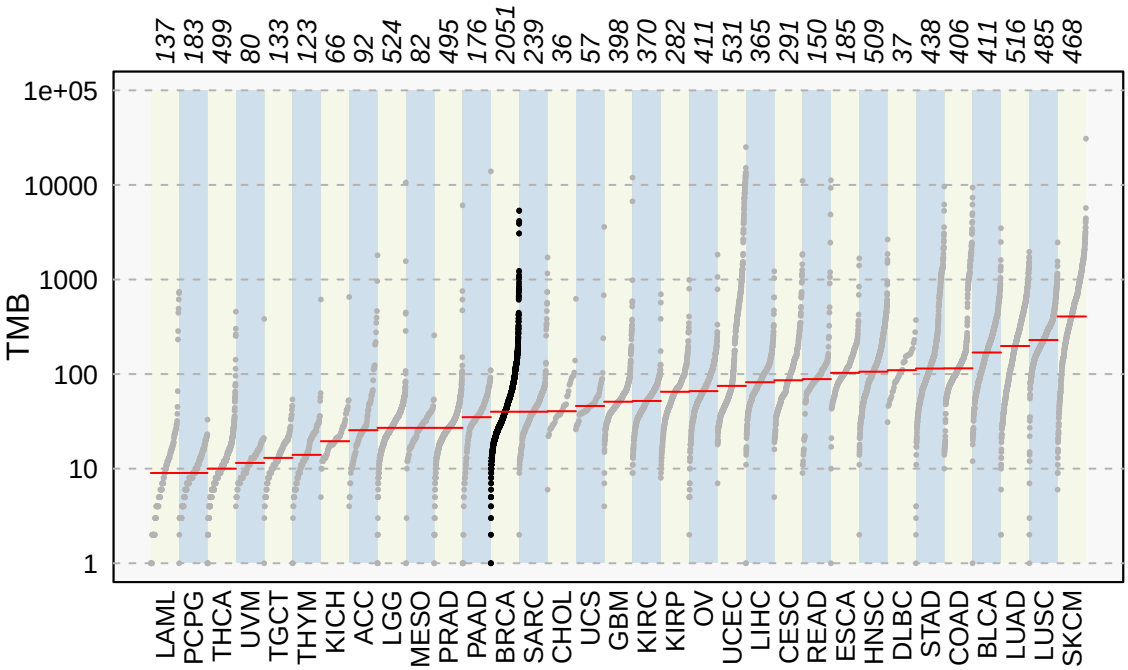


#Compare mutation load with other TCGA cohort

```
BRCA.mutload = tcgaCompare(maf = BRCA, cohortName = 'BRCA')
```

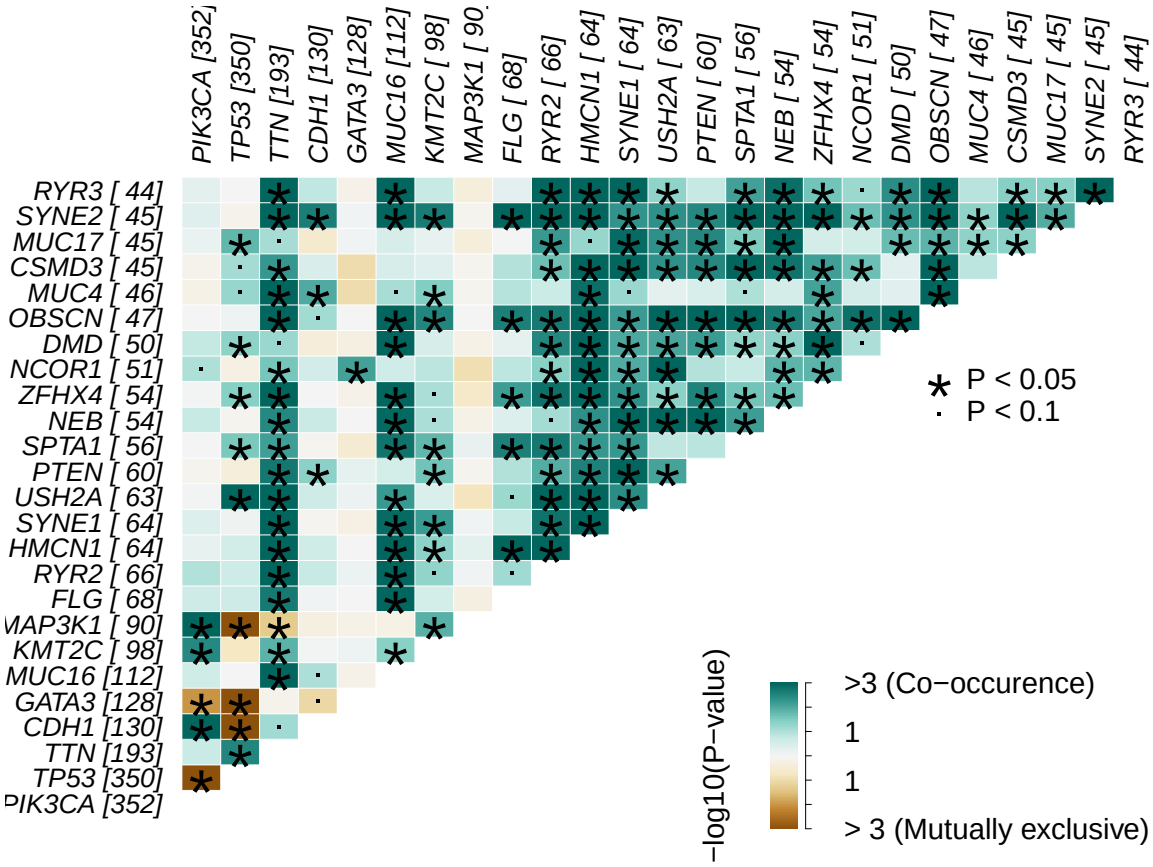
```
## Warning in FUN(X[[i]], ...): Removed 0 samples with zero mutations.
```

```
## Performing pairwise t-test for differences in mutation burden..
```



#Visualizing top co-occuring mutated genes

```
somaticInteractions(maf = BRCA, top = 25, pvalue = c(0.05, 0.1))
```




```
##       gene1  gene2       pValue oddsRatio  00 11  01  10             Event
##   1:   TP53   CDH1 1.487319e-14 0.1212479 555  9 121 341 Mutually_Exclusive
##   2:  GATA3   TP53 1.511579e-12 0.1552365 559 11 339 117 Mutually_Exclusive
##   3:  HMCN1    TTN 2.268967e-08 4.6293475 800 31 162  33       Co_Occurence
##   4:  MUC16    TTN 6.316428e-08 3.3171099 765 44 149  68       Co_Occurence
##   5: PIK3CA   TP53 8.709609e-08 0.4604239 406 82 268 270 Mutually_Exclusive
##  ---
## 296:    FLG  GATA3 1.000000e+00 0.9311548 838  8 120  60 Mutually_Exclusive
## 297:  HMCN1  GATA3 1.000000e+00 1.0023787 842  8 120  56       Co_Occurence
## 298:    NEB  GATA3 1.000000e+00 0.8710224 850  6 122  48 Mutually_Exclusive
## 299:  OBSCN MAP3K1 1.000000e+00 0.9659580 893  4  86  43 Mutually_Exclusive
## 300:  MUC17    FLG 1.000000e+00 1.0065873 916  3  65  42       Co_Occurence
##               pair event_ratio
##   1:    CDH1, TP53       9/462
##   2:   GATA3, TP53      11/456
##   3:    HMCN1, TTN      31/195
##   4:    MUC16, TTN      44/217
##   5:  PIK3CA, TP53      82/538
##  ---
## 296:    FLG, GATA3       8/180
## 297:  GATA3, HMCN1       8/176
## 298:    GATA3, NEB       6/170
## 299: MAP3K1, OBSCN       4/129
## 300:    FLG, MUC17       3/107
```

#Detecting cancer driver genes that are frequently mutated

```
BRCA.sig = oncodrive(maf = BRCA, AACol = 'HGVSp_Short', minMut = 5, pvalMethod = 'zscore')
```

```
## Warning in oncodrive(maf = BRCA, AACol = "HGVSp_Short", minMut = 5, pvalMethod
## = "zscore"): Oncodrive has been superseeded by OncodriveCLUSTL. See
## http://bg.upf.edu/group/projects/oncodrive-clust.php
```

```
## No syn mutations found! Skipping background estimation. Using predefined values. (Mean = 0.279; SD =
```

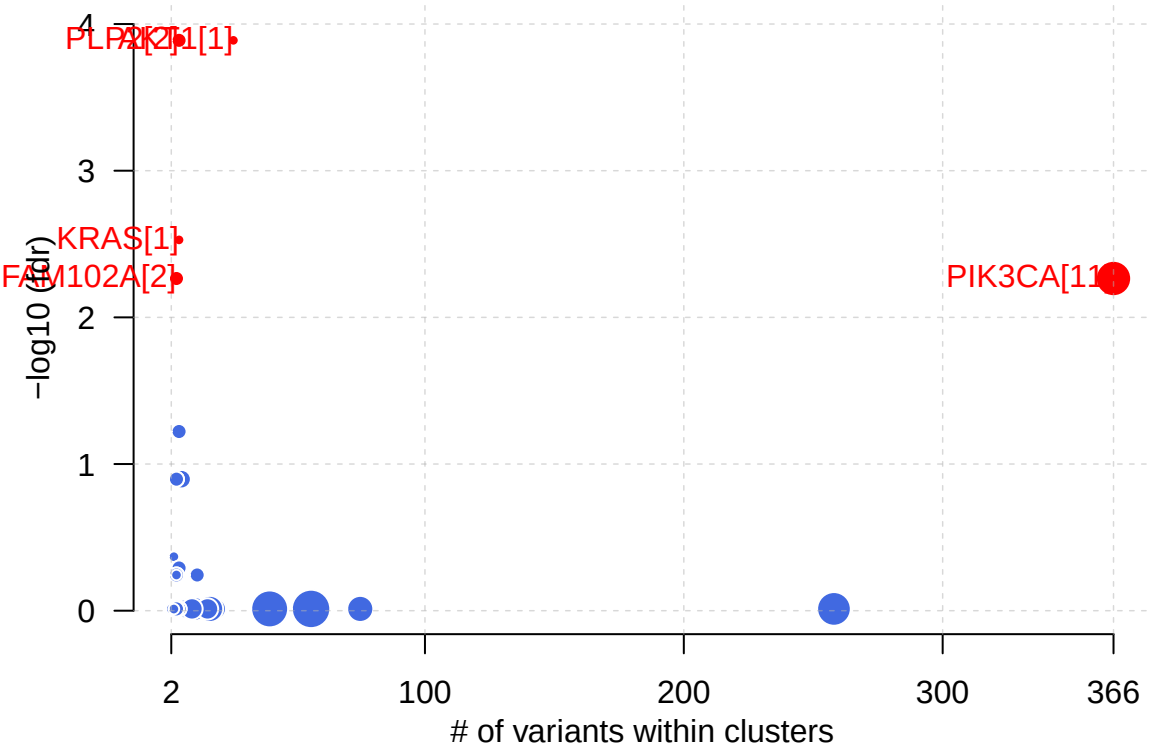
```
## Estimating cluster scores from non-syn variants..
```

```
##   |                                                                              |
```

```
head(BRCA.sig)
```

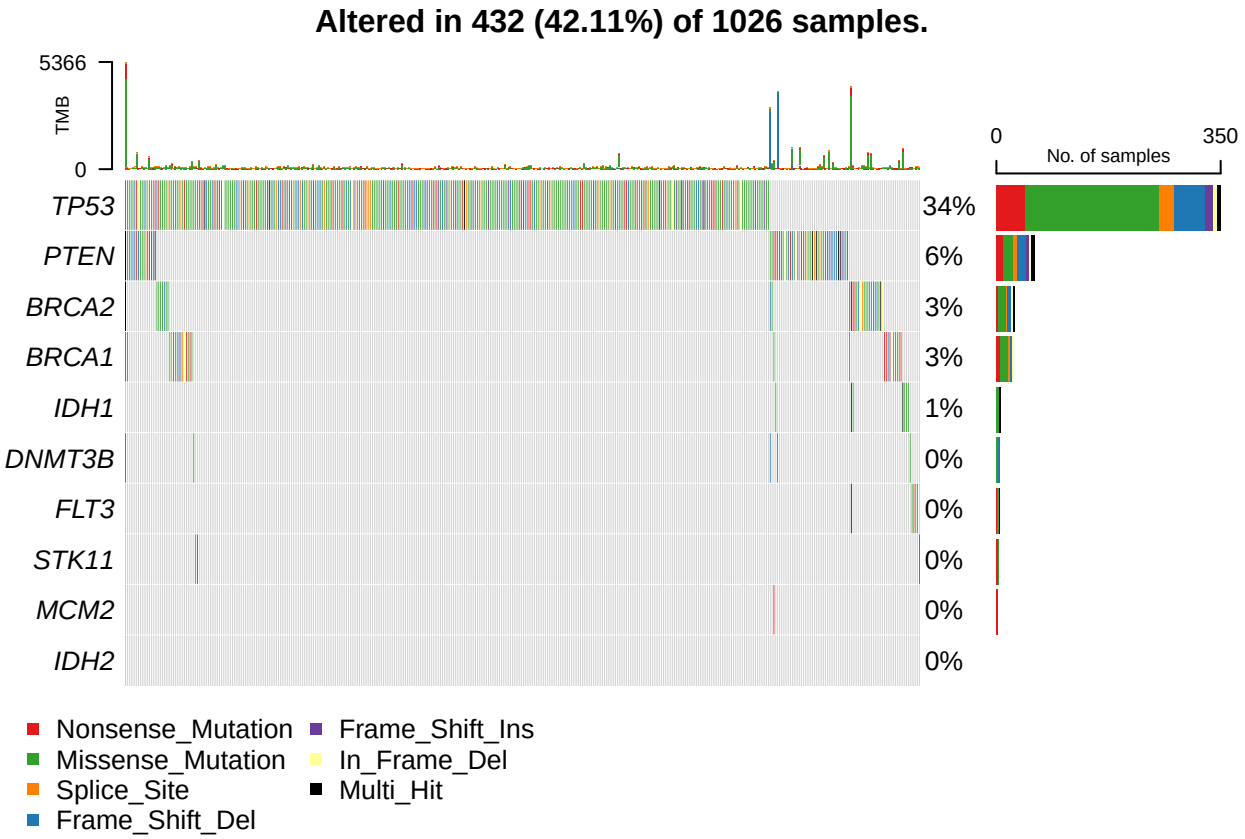
```
##    Hugo_Symbol Frame_Shift_Del Frame_Shift_Ins In_Frame_Del In_Frame_Ins
## 1:        AKT1               0               0            0            0
## 2:        PLP2               0               0            0            0
## 3:        KRAS               0               0            0            0
## 4:     FAM102A               0               0            0            0
## 5:      PIK3CA               0               0           15            0
## 6:      PIK3R3               3               0            0            0
##    Missense_Mutation Nonsense_Mutation Nonstop_Mutation Splice_Site
## 1:                28                 0                0          0
## 2:                 5                 0                0          0
## 3:                 6                 0                0          0
## 4:                 4                 1                0          0
## 5:               378                 0                0          0
## 6:                 4                 0                0          0
##    Translation_Start_Site total MutatedSamples AlteredSamples clusters
## 1:                      0    28             28             28        1
## 2:                      0     5              3              3        2
## 3:                      0     6              6              6        1
## 4:                      0     5              4              4        2
## 5:                      0   393            352            352       11
## 6:                      0     7              7              7        2
##    muts_in_clusters clusterScores protLen   zscore         pval          fdr
## 1:               26     0.9285714     480 4.996703 2.915935e-07 0.0001291759
## 2:                5     0.9414214     152 5.095549 1.738661e-07 0.0001291759
## 3:                5     0.8333333     189 4.264103 1.003536e-05 0.0029637769
## 4:                4     0.8000000     384 4.007692 3.065747e-05 0.0054325028
## 5:              366     0.8059869    1068 4.053746 2.520201e-05 0.0054325028
## 6:                5     0.7142857     461 3.348352 4.064690e-04 0.0600219154
##    fract_muts_in_clusters
## 1:              0.9285714
## 2:              1.0000000
## 3:              0.8333333
## 4:              0.8000000
## 5:              0.9312977
## 6:              0.7142857
```

```
plotOncodrive(BRCA.sig)
```



#Visualizing multiple genes

```
multi_genes = c("TP53", "MCM2", "BRCA1", "BRCA2", "DNMT3B", "PTEN", "STK11", "IDH1", "IDH2", "FLT3")
oncoplot(maf = BRCA, genes = multi_genes)
```


Altered in 432 (42.11%) of 1026 samples.
5366
TMB
0
0
350
No. of samples
TP53
PTEN
BRCA2
BRCA1
IDH1
DNMT3B
FLT3
STK11
MCM2
IDH2
34%
6%
3%
3%
1%
0%
0%
0%
0%
0%
Nonsense_Mutation
Missense_Mutation
Splice_Site
Frame_Shift_Del
Frame_Shift_Ins
In_Frame_Del
Multi_Hit