



Clase 6:

Librería Maftools para interpretación de variantes.

Evelin González F.
evelyn.gonzalez@uoh.cl

Organización de las clases: Parte 2

Clase 5: Introducción de R.

Clase 6: Librería Maftools para interpretación de variantes.

Clase 7: Análisis de clustering/PCA y categorización de variantes patogénicas, visualización de los datos.

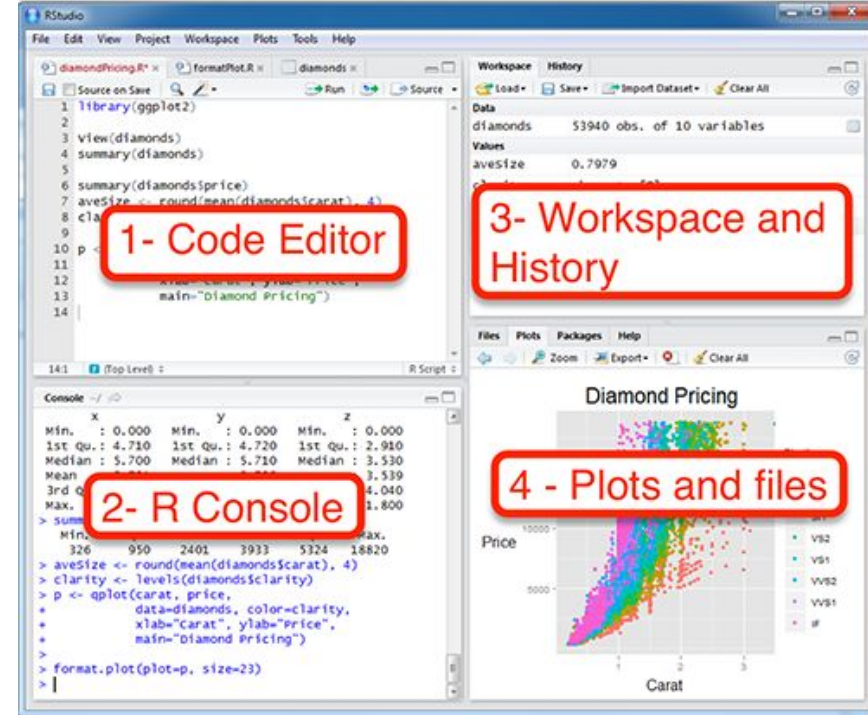
Introducción a R y Rstudio

R es un lenguaje de programación y entorno para el análisis estadístico y la visualización de datos

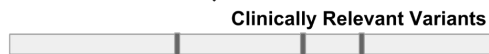
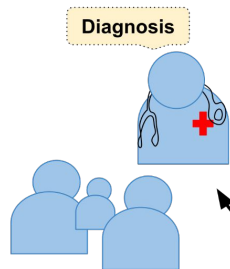
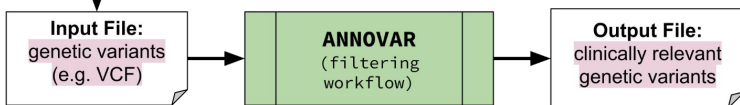
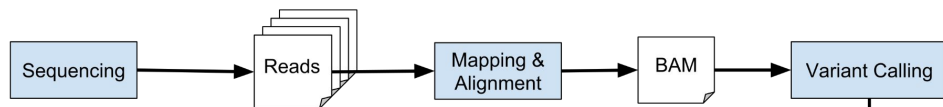
Open Source, multiplataforma y extensible

Por qué usar R?

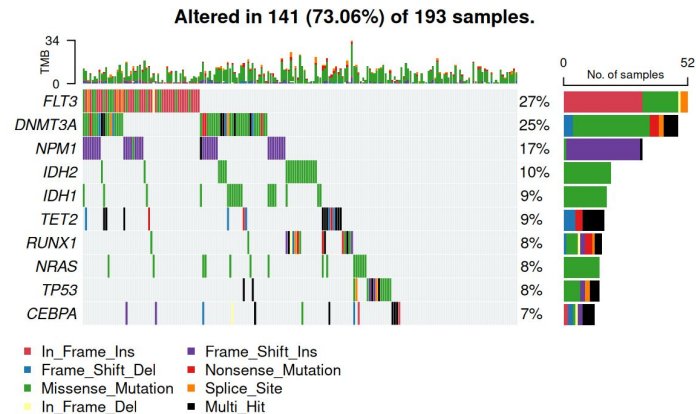
- Amplio soporte para **estadística**: desde análisis básicos hasta modelado avanzado.
- Potente para **gráficos y visualizaciones**.
- Gran cantidad de paquetes en **CRAN** y **Bioconductor** para aplicaciones específicas.



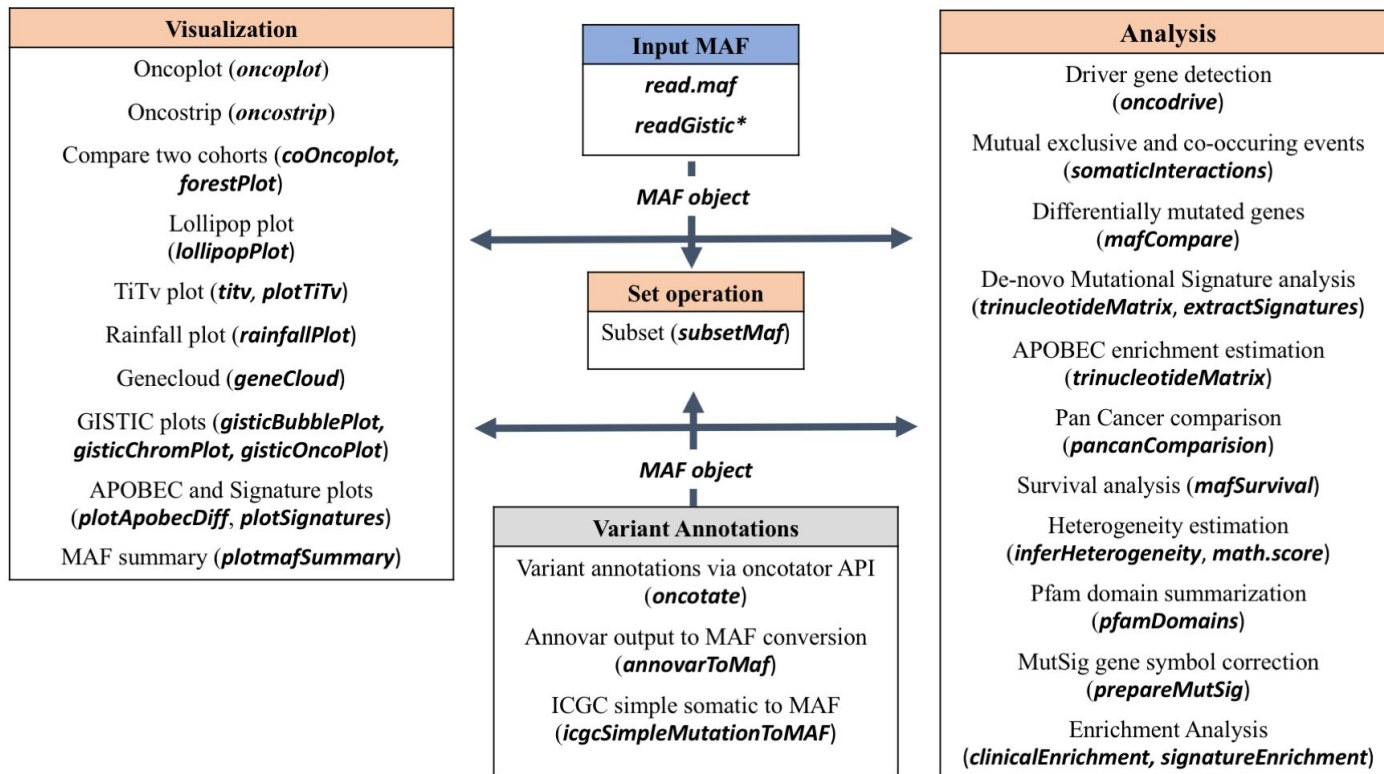
Maftools: Un paquete de R para resumir, analizar y visualizar archivos MAF



Clinical report



Maftools, para la interpretación de variantes oncológicas



MAF (*Mutation Annotation Format*)

Requisitos de los campos MAF

Los archivos MAF contienen numerosos campos, desde nombres de cromosomas hasta anotaciones de COSMIC. Sin embargo, la mayoría de los análisis en **maftools** utiliza los siguientes campos:

Campos obligatorios:

- Hugo_Symbol
- Chromosome
- Start_Position
- End_Position
- Reference_Allele
- Tumor_Seq_Allele2
- Variant_Classification
- Variant_Type
- Tumor_Sample_Barcode

```
if (!require("BiocManager"))  
  install.packages("BiocManager")  
BiocManager::install("maftools")
```

Campos opcionales recomendados:

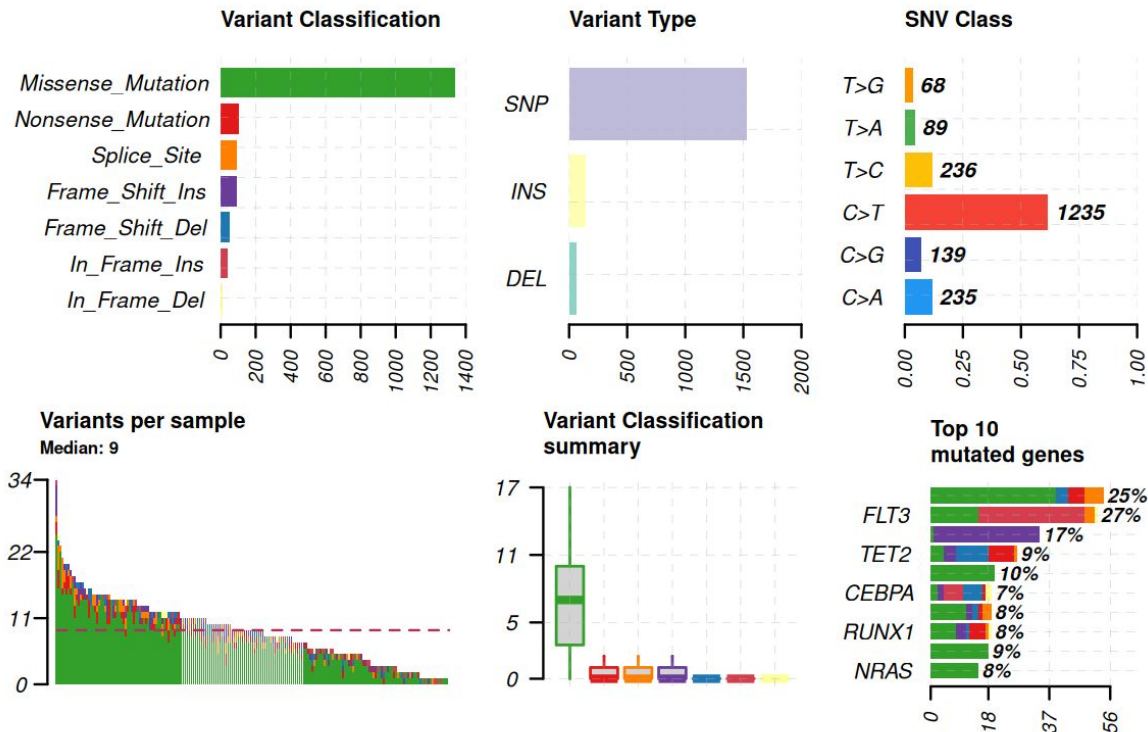
- Campos no específicos de MAF que contengan información sobre VAF (frecuencia del alelo variante) y cambios en aminoácidos.

Análisis exploratorio, funciones básicas de maftools

```
#Shows sample summary.  
getSampleSummary(laml)  
#Shows gene summary.  
getGeneSummary(laml)  
#shows clinical data associated with samples  
getClinicalData(laml)  
#Shows all fields in MAF  
getFields(laml)  
#Writes maf summary to an output file with basename laml.  
write.mafSummary(maf = laml, basename = 'laml')
```

Graficando el Resumen del MAF

```
plotmafSummary(maf = lam1, rmOutlier = TRUE, addStat = 'median', dashboard = TRUE, titvRaw = FALSE)
```



Maftools : Oncoplots

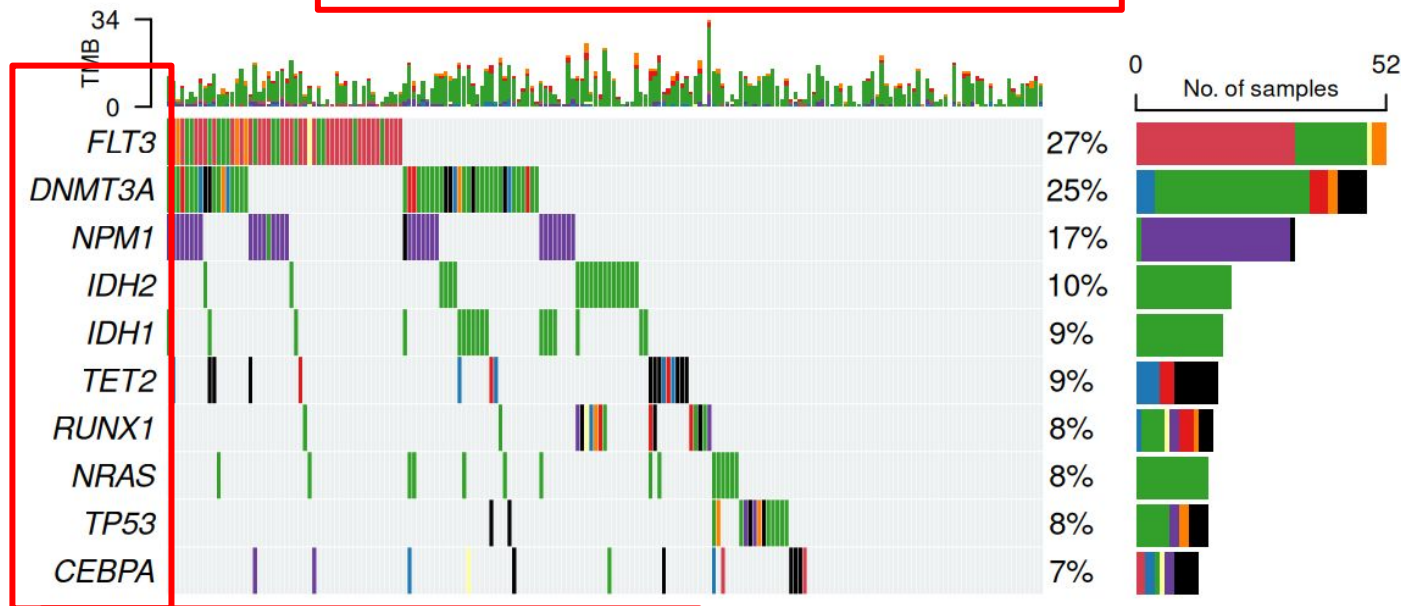
```
#oncoplot for top ten mutated genes.
```

```
oncoplot(maf = lam1, top = 10)
```

Top 10 de genes
con mayor número
de alteraciones

Altered in 141 (73.06%) of 193 samples.

Número de pacientes
con alteración



■ In_Frame_Ins ■ Frame_Shift_Ins
■ Frame_Shift_Del ■ Nonsense_Mutation
■ Missense_Mutation ■ Splice_Site
■ In_Frame_Del ■ Multi_Hit

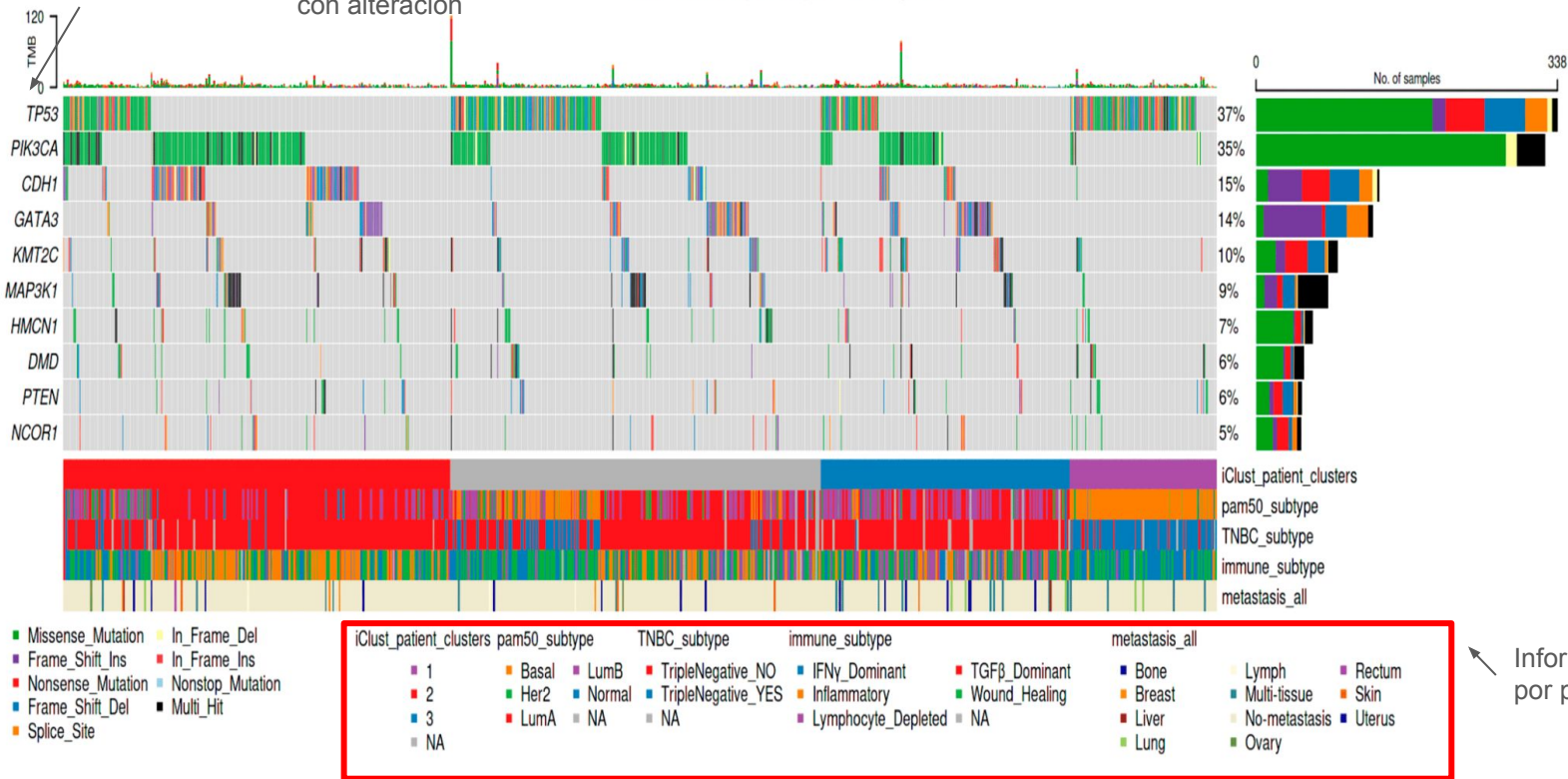
Tipo de mutación.

Maftools : Oncoplots

Top 10 de genes
con mayor número
de alteraciones

Número de pacientes
con alteración

Altered in 809 (87.93%) of 920 samples.



Información Clínica
por paciente

Tipo de mutación.

Caso de uso: Estudio de mutaciones somáticas de 4 tipos de cáncer en pacientes chilenos

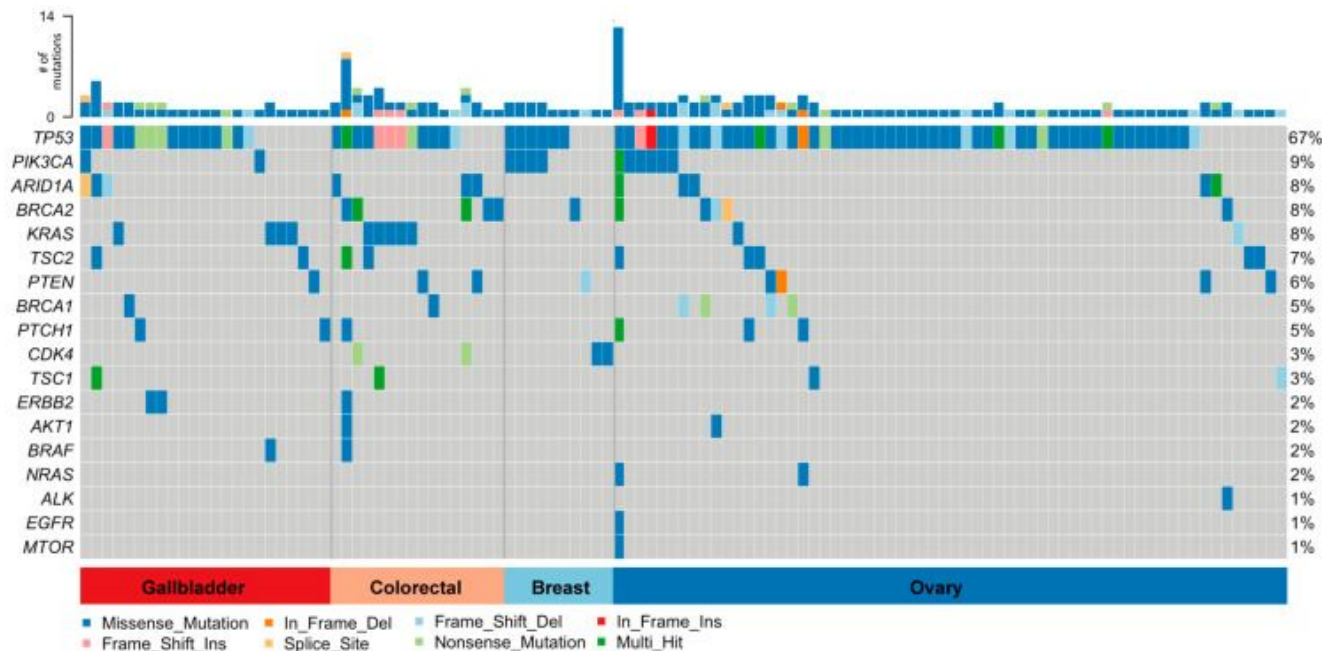


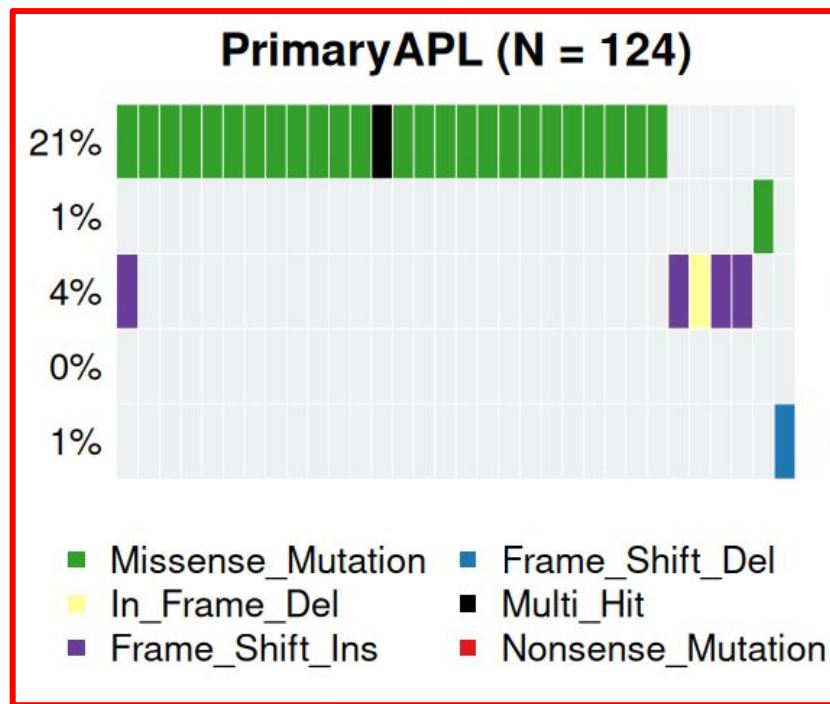
Figura: Variantes somáticas no sinónimas detectadas en todas las muestras tumorales; 197 variantes somáticas no sinónimas. Variantes detectadas en 111 pacientes.

```
#oncoplot for top ten mutated genes.  
oncoplot(maf = lam1, top = 10)
```

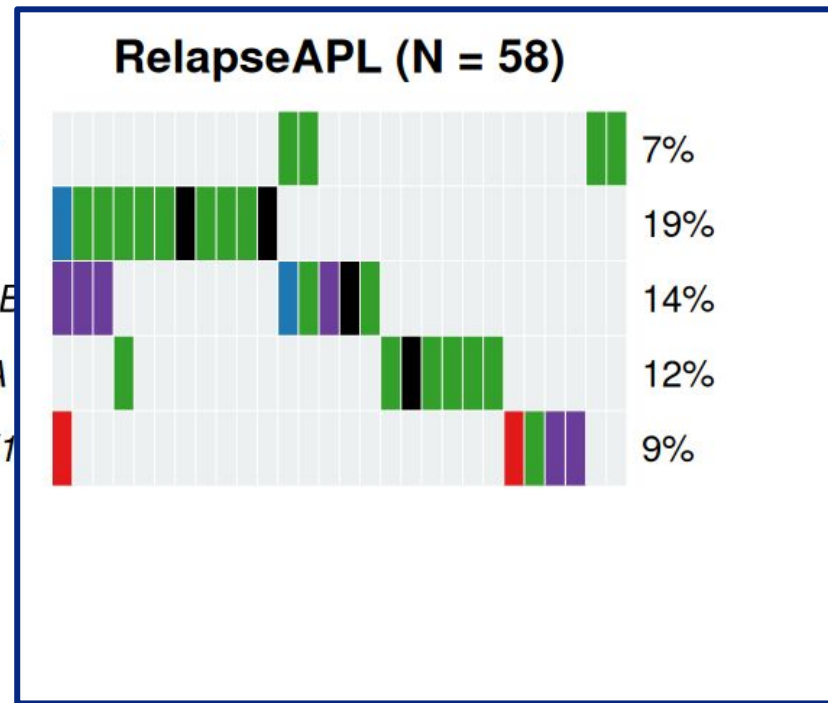
Maftools: Co-oncoplots

```
genes = c("PML", "RARA", "RUNX1", "ARID1B", "FLT3")
```

```
coOncoplot(m1 = primary.apl, m2 = relapse.apl, m1Name =  
'PrimaryAPL', m2Name = 'RelapseAPL', genes = genes,  
removeNonMutated = TRUE)
```



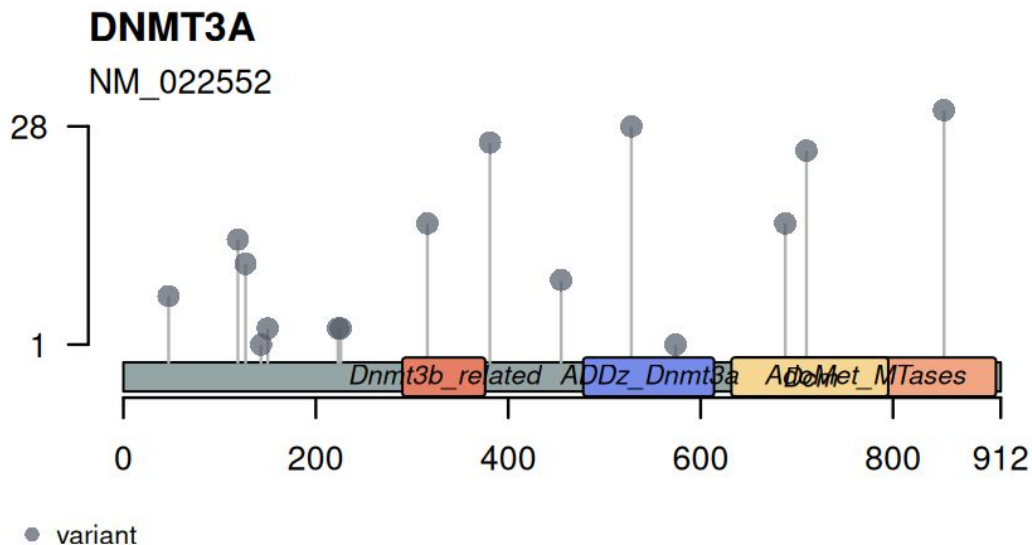
GRUPO1



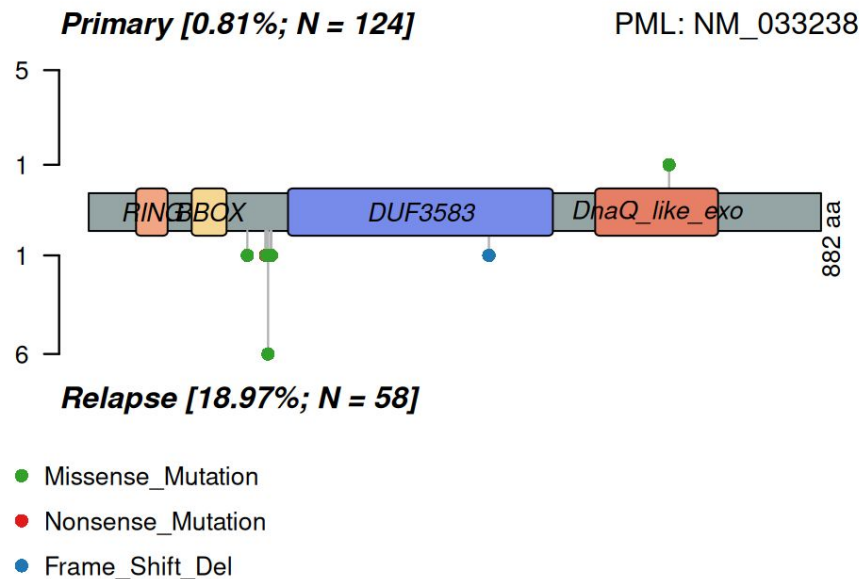
GRUPO2

Maftools: Lollipop Plot

```
#lollipop plot for DNMT3A, which is one of the most frequent mutated gene in Leukemia.  
lollipopPlot(  
  maf = lam1,  
  gene = 'DNMT3A',  
  AACol = 'Protein_Change',  
  showMutationRate = TRUE,  
  labelPos = 882  
)
```



maftools: Lollipop Plot2



```
lollipopPlot2(m1 = primary.apl,
```

```
m2 = relapse.apl,
```

```
gene = "PML",
```

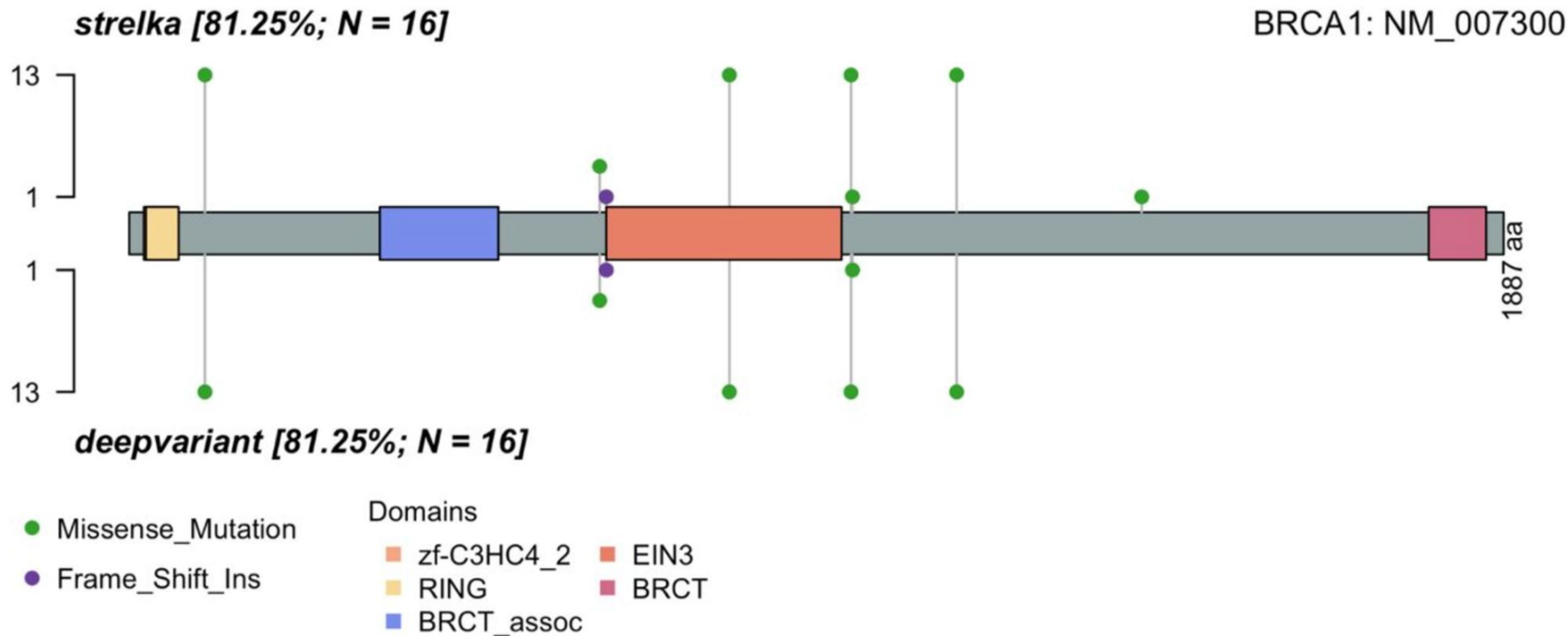
```
AACol1 = "amino_acid_change",
```

```
AACol2 = "amino_acid_change",
```

```
m1_name = "Primary",
```

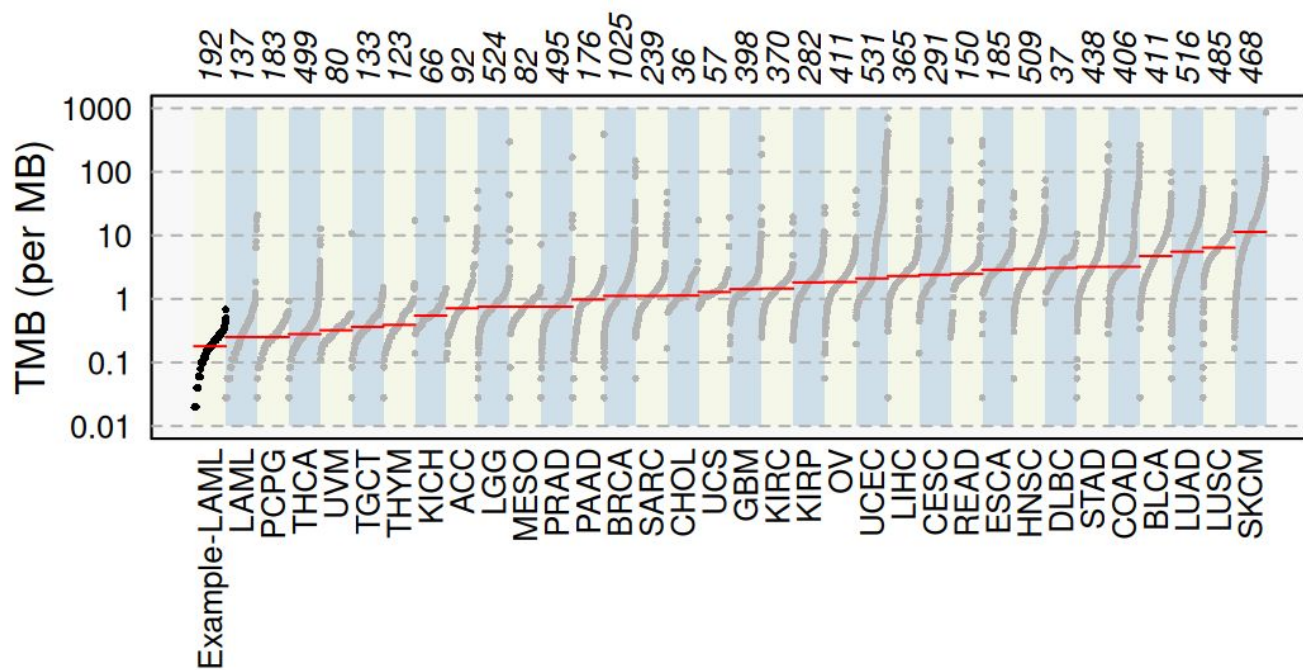
```
m2_name = "Relapse")
```

Caso de uso: comparación entre llamadores de variantes

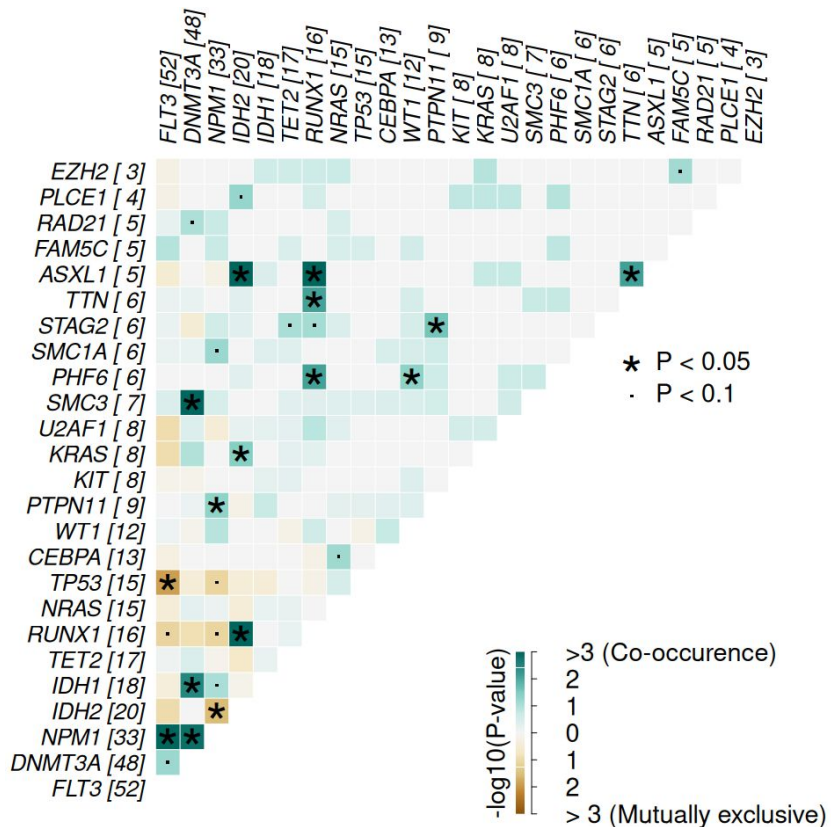


Cálculo de TMB por tipo de câncer

```
laml.mutload = tcgaCompare(maf = laml, cohortName = 'Example-LAML', logscale = TRUE, capture_size = 50)
```



maftools: Co-ocurrencia y exclusión mutua.



Interacciones Somáticas

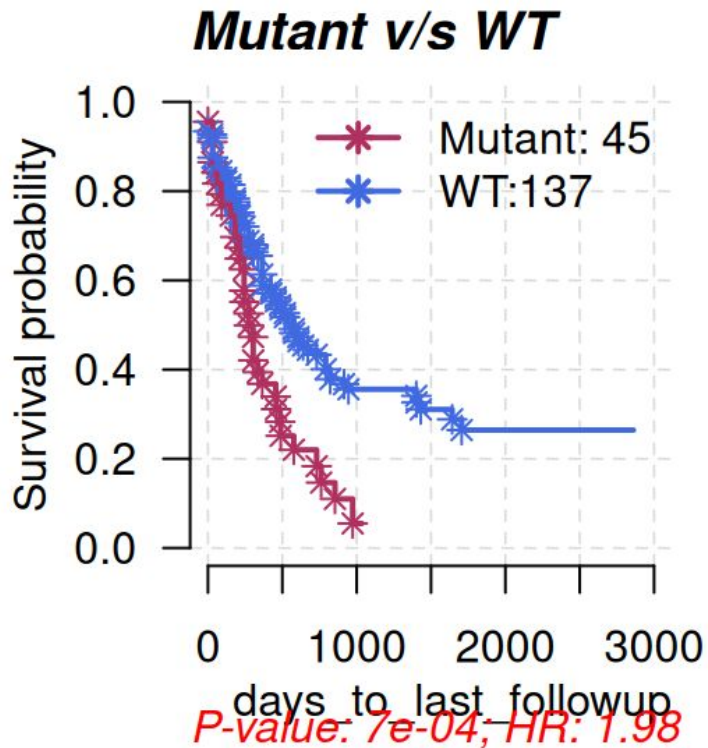
Conjuntos de genes mutuamente excluyentes o que co-ocurren pueden detectarse utilizando la función `somaticInteractions`, la cual realiza pruebas exactas de Fisher por pares para identificar pares significativos de genes.

```
#exclusive/co-occurrence event analysis on top 10 mutated genes.
```

```
somaticInteractions(maf = lam1, top = 25, pvalue = c(0.05, 0.1))
```

Análisis de Supervivencia

```
mafSurvival(maf = lam1, genes = 'DNMT3A', time =  
'days_to_last_followup', Status = 'Overall_Survival_Status',  
isTCGA = TRUE)
```



El análisis de supervivencia es una parte esencial de los proyectos de secuenciación basados en cohortes. La función `mafSurvive` realiza este análisis y genera **curvas de Kaplan-Meier** agrupando las muestras según el estado de mutación de genes definidos por el usuario o grupos de muestras proporcionados manualmente.

Esta función requiere que los datos de entrada incluyan:

- **Tumor_Sample_Barcode:** Asegúrate de que coincidan con los del archivo MAF.
- **Evento binario (1/0):** Indica si ocurrió el evento (1) o no (0).
- **Tiempo hasta el evento:** Tiempo transcurrido hasta el evento o hasta el final del seguimiento.

Si ya cuentas con datos de supervivencia en los datos de anotación, la función los utilizará automáticamente. En caso de que los datos de supervivencia se encuentren en una tabla separada, puedes proporcionarlos utilizando el argumento `clinicalData`.