Mathematical Foundations of Bioinformatics
Problem Set 4
Fateme Vafaee Sharbaf

Programming: TCGA datasets

Download the MAF les for the following TCGA projects UCEC, ACC, SARC, BRCA using the TCGABi-
olinks package in R.
(a) Save all datasets as RData les in the separate folders and les.
(b) Use the maftools package to plot three plots of your choice and save the gures (in PDF and PNG) in a
neat folder structure.
(c) Make yourself familiar with the R Markdown and generate a short report using the generated PDFs in
the previous step.

Program Guide Line

In Order to have a progarm describing in the previous section, we imply two steps:

1. Load And Save

2. Plots

In the following we try to explain more about steps.

1. Load And Save

In this section we try to download datasets from TCGA web site (for more information click On TCGA Web
site) using the ''TCGAbiolinks'' .
We also consider a specific directory pattern for the downloaded data.

DataSets	ACC	SARC	BRCA	UCEC
Plots	Lollipop	PlotVaf	plotmafSummary	

```
Save_Load=function(folder_name,DataSet_Names)
{
  if(folder_name=="Input")
  {
    dir.create("Input")
    dir.create("Input/Maf")
  }
  if(folder_name=="Input/Maf")
    dir.create("Input/Maf")
```

```
  dir_name="Input/Maf/";
  DataSet_FullPath="dir"
  for(i in 1:length(DataSet_Names))
  {
    MainDir=paste(dir_name,DataSet_Names[i],sep = "")
    Maf_Data_Name=paste(DataSet_Names[i],'_M',sep = "")
    RDate_Data_Name=paste(DataSet_Names[i],'_M.RData',sep = "")
    DataSet_FullPath[i]=paste(MainDir,"/",RDate_Data_Name,sep = "")
    if(!dir.exists(MainDir))
    {
      dir.create(MainDir)
    }else{
      if(file.exists(DataSet_FullPath[i]))
        next
    }

    maf <- GDCquery_Maf(DataSet_Names[i], pipelines = "muse")
    save(maf,file=DataSet_FullPath[i])

  }
return(DataSet_FullPath)
}
```

2. Plots

THe main aim in Plots function is to draw three different plot for datasets and save (in pdf and png format) the in the Output folder.

```
Plots=function(DataSet_Dir,DataSet_Names)
{
  memory.limit(size = 30000)
  if(!dir.exists("Outputs"))
    dir.create("Outputs")
  for(i in 1:length(DataSet_Names))
  {
    load(DataSet_Dir[i])
    maf=read.maf(maf = maf)
    #Loc=strsplit(DataSet_Dir[i],'.RData')[[1]]
    Loc=paste("Outputs/",DataSet_Names[i],sep = "")
    if(!dir.exists(Loc))
      dir.create(Loc)
    Loc=paste(Loc,"/",DataSet_Names[i],sep = "")
    #------------Lollipop------------------

    pngName=paste(Loc,'_lollipopPlot.png',sep = "")
    png(file = pngName, width = 1200,height = 500)
    lollipopPlot(maf = maf, gene = 'TP53', AACol = 'HGVSp_Short', refSeqID = 'NM_001126112')
    dev.off()

    pdfName=paste(Loc,'_lollipopPlot.pdf',sep = "")
    pdf(file = pdfName, width = 10,height = 5)
    lollipopPlot(maf = maf, gene = 'TP53', AACol = 'HGVSp_Short', refSeqID = 'NM_001126112')
```

```
  dev.off()
  #-------------PlotVaf-------------------

  pngName=paste(Loc,'_PlotVaf.png',sep = "")
  png(file = pngName, width = 1200,height = 500)
  laml.titv = titv(maf = maf, useSyn = TRUE)
  plotTiTv(laml.titv)
  dev.off()

  pdfName=paste(Loc,'_PlotVAF.pdf',sep = "")
  pdf(file = pdfName, width = 10,height = 5)
  laml.titv = titv(maf = maf, useSyn = TRUE)
  plotTiTv(laml.titv)
  dev.off()

  #-------------plotmafSummary----------------

  pngName=paste(Loc,'_plotmafSummary.png',sep="")
  png(file = pngName, width = 1200,height = 500)
  plotmafSummary(maf = maf, addStat = 'median')
  dev.off()

  pdfName=paste(Loc,'_plotmafSummary.pdf',sep="")
  pdf(file = pdfName, width = 10,height = 5)
  plotmafSummary(maf = maf, addStat = 'median')
  dev.off()
  }
}
```

Outputs Plots

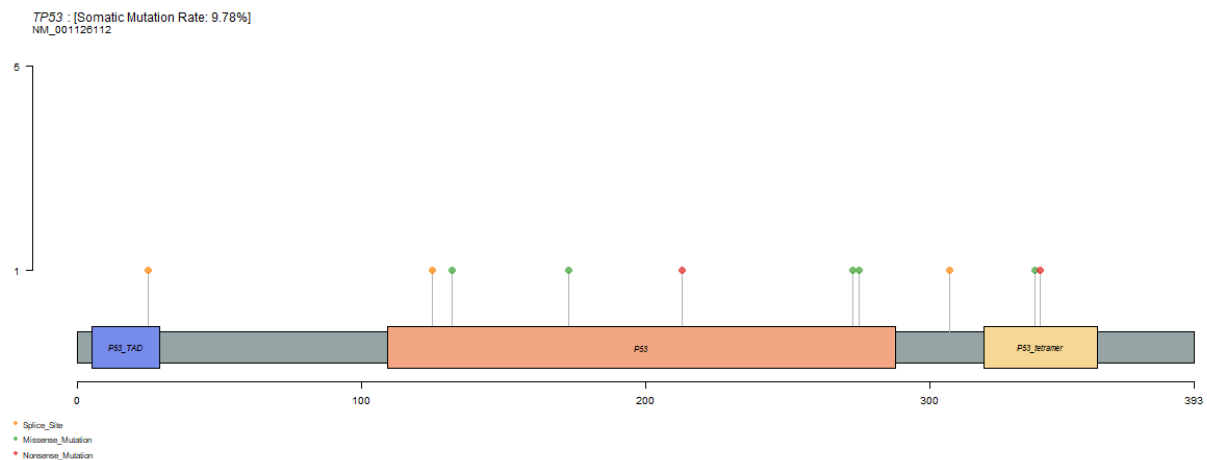Here we depict three main plots, drawing for datasets.
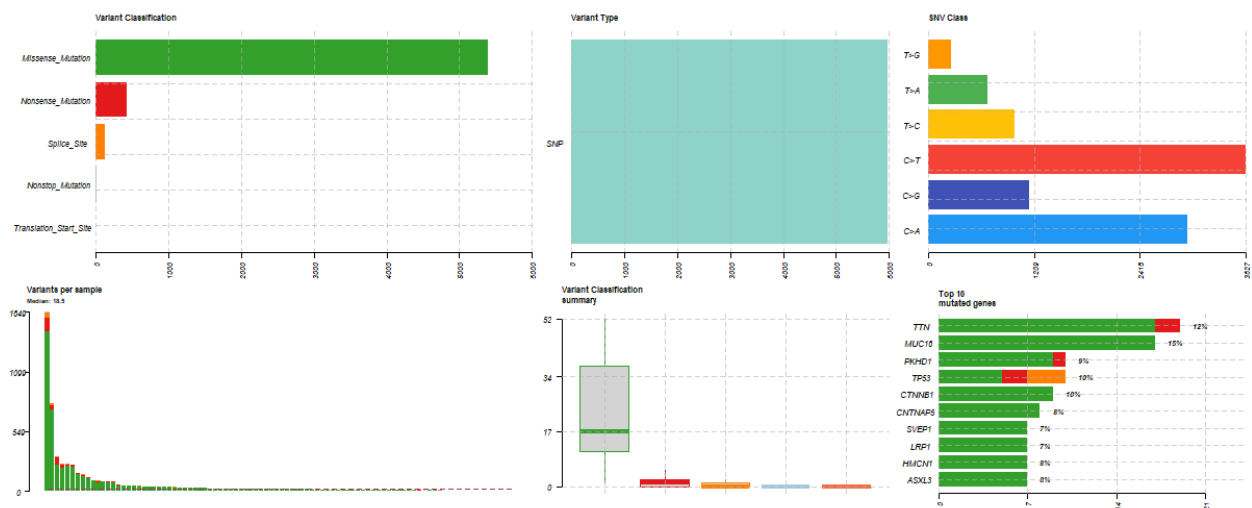

Figure 1: Lollipop Plot For ACC

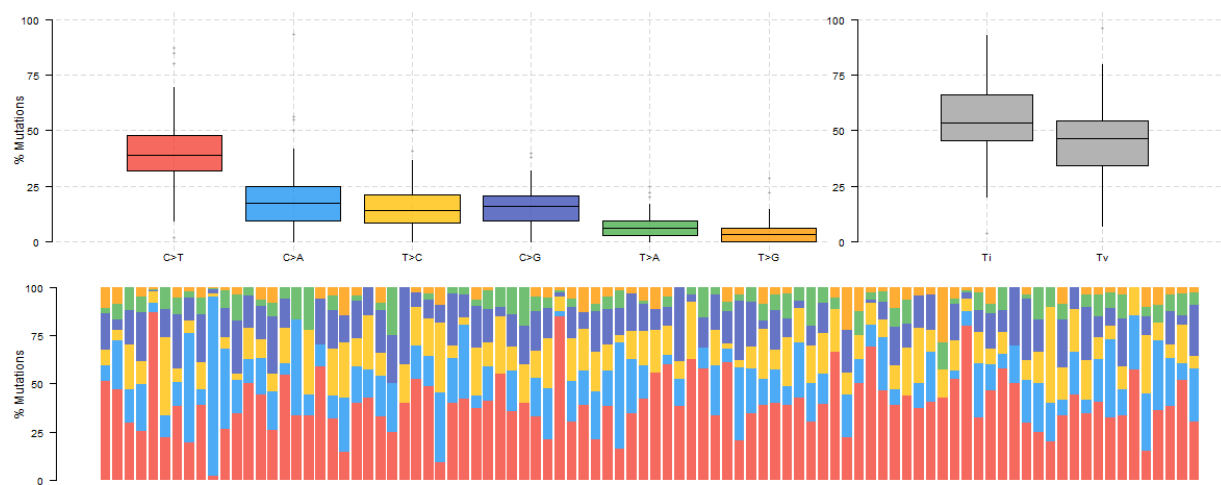Figure 2: plotmafSummary Plot For ACC

Figure 3: PlotVaf Plot For ACC