

Positive selection on genes interacting with SARS-Cov2, comparison of different analysis

Marie Cariou

October 2020

Contents

1	Data	2
2	Comparisons Primates	2
2.1	Janet Young's results (Young-primate) VS DGINN-full's results	2
2.2	Janet Young's results (Young-primate) VS Cooper's result . .	3
2.3	Cooper's results (Cooper-primate) VS DGINN-full's results .	4
3	Overlap	6
3.1	Mondrian	6
3.2	subsetR	8
4	Gene List	10
5	Shiny like	12

1 Data

Analysis were formatted by the script covid_comp_script0_table.Rnw.

```
workdir<-"/home/adminmarie/Documents/CIRI_BIBS_projects/2020_05_Etienne_covid/"

tab<-read.delim(paste0(workdir,
  "covid_comp/covid_comp_complete.txt"), h=T, sep="\t")
dim(tab)

## [1] 332 139

tab$Gene.name<-as.character(tab$Gene.name)
tab$Gene.name[tab$PreyGene=="MTARC1"]<-"MTARC1"
```

2 Comparisons Primates

2.1 Janet Young's results (Young-primate) VS DGINN-full's results

Comparaison des Omega: colonne L "whole.gene.dN.dS.model.0" VS colonne "omega" dans la sortie de dginn.

```
plot(tab$whole.gene.dN.dS.model.0, as.numeric(as.character(tab$dginn.primate_omegaMOBpp)),
      xlab="Omega Young-primate", ylab="DGINN-full's")

## Warning in xy.coords(x, y, xlabel, ylabel, log): NAs introduits
lors de la conversion automatique

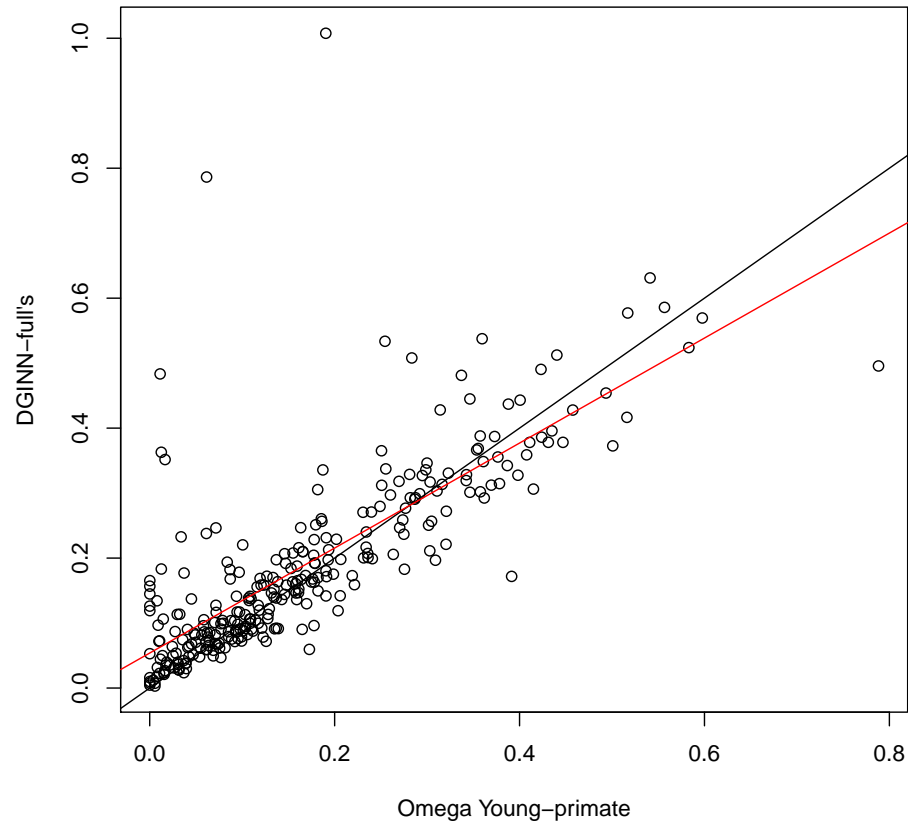
abline(0,1)
abline(lm(as.numeric(as.character(tab$dginn.primate_omegaMOBpp))~tab$whole.gene.dN.dS.model.0))

## Warning in eval(predvars, data, env): NAs introduits lors de
la conversion automatique

outlier<-tab[tab$whole.gene.dN.dS.model.0<0.4 & tab$dginn.primate_omegaMOBpp>0.5,]

## Warning in Ops.factor(tab$dginn.primate_omegaMOBpp, 0.5): '>'
not meaningful for factors

text(x=outlier$whole.gene.dN.dS.model.0,
     y=outlier$dginn.primate_omegaMOBpp,
     outlier$Gene.name)
```



2.2 Janet Young's results (Young-primate) VS Cooper's result

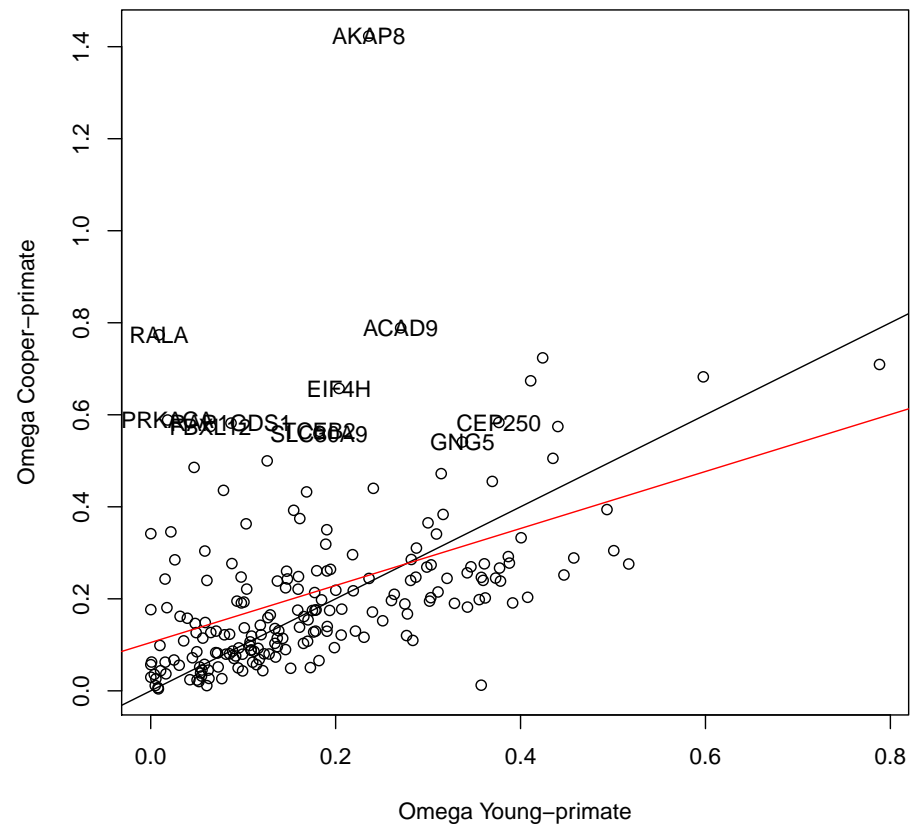
Comparaison des Omega: colonne L "whole.gene.dN.dS.model.0" VS colonne "cooper.primates.Average_dNdS".

```
plot(tab$whole.gene.dN.dS.model.0, as.numeric(as.character(tab$cooper.primates.Average_dNdS)),
     xlab="Omega Young-primate", ylab="Omega Cooper-primate")
abline(0,1)
abline(lm(as.numeric(as.character(tab$cooper.primates.Average_dNdS))~tab$whole.gene.dN.dS.model.0))
```

```

outlier<-tab[tab$whole.gene.dN.dS.model.0<0.4 & tab$cooper.primates.Average_dNdS>0.5,
text(x=outlier$whole.gene.dN.dS.model.0,
y=outlier$cooper.primates.Average_dNdS,
outlier$Gene.name)

```



2.3 Cooper's results (Cooper-primate) VS DGINN-full's results

Comparaison des Omega: colonne "cooper.primates.Average_dNdS" VS colonne "omega" dans la sortie de dginn.

```

plot(tab$cooper.primates.Average_dNd, as.numeric(as.character(tab$dginn.primate_omegaMOBpp)),
      xlab="Omega Cooper-primate", ylab="DGINN-full's")

## Warning in xy.coords(x, y, xlabel, ylabel, log): NAs introduits
lors de la conversion automatique

abline(0,1)
abline(lm(as.numeric(as.character(tab$dginn.primate_omegaMOBpp))~tab$cooper.primates.Average_dNd))

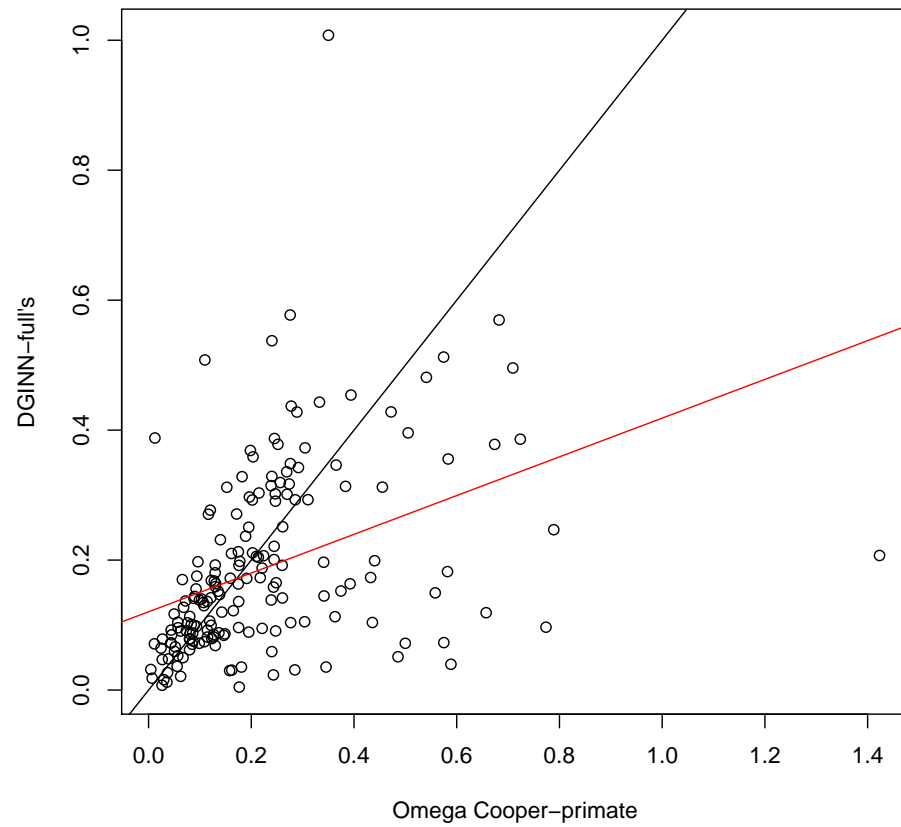
## Warning in eval(predvars, data, env): NAs introduits lors de
la conversion automatique

outlier<-tab[tab$cooper.primates.Average_dNd<0.4 & tab$dginn.primate_omegaMOBpp>0.5,]

## Warning in Ops.factor(tab$dginn.primate_omegaMOBpp, 0.5): '>'
not meaningful for factors

text(x=outlier$cooper.primates.Average_dNd,
     y=outlier$dginn.primate_omegaMOBpp,
     outlier$Gene.name)

```



3 Overlap

3.1 Mondrian

```
library(Mondrian)

#####

monddata<-as.data.frame(tab$Gene.name)
dim(monddata)

## [1] 332 1
```

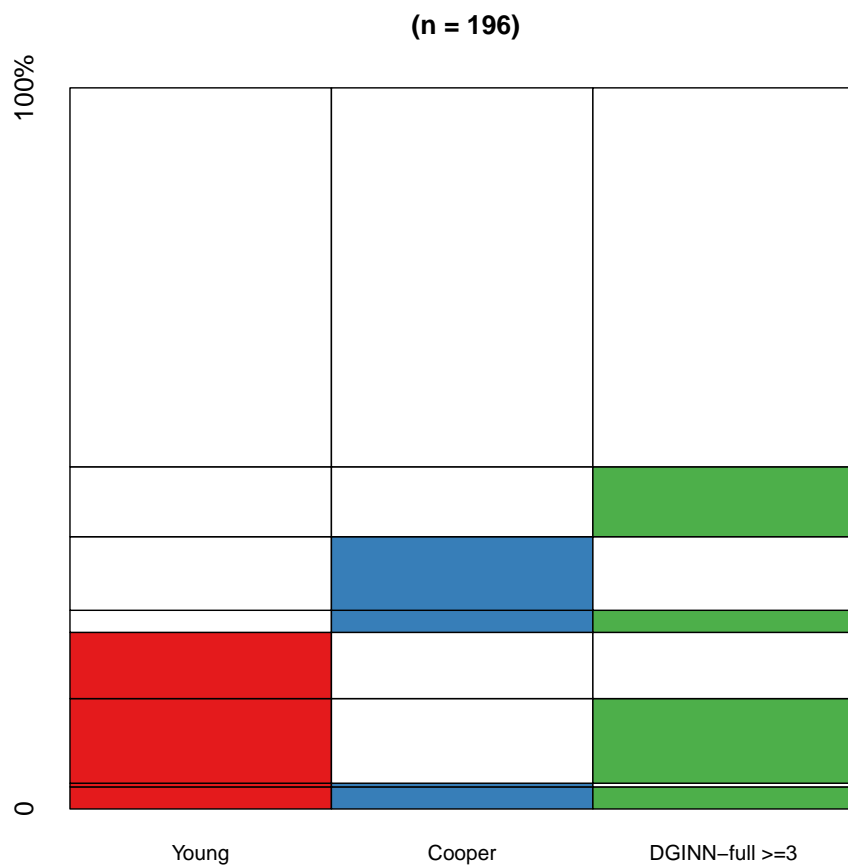
```

dginnfulltmp<-rowSums(cbind(tab$dginn.primate_BUSTED=="Y", tab$dginn.primate_BppM1M2=
tab$dginn.primate_BppM7M8=="Y", tab$dginn.primate_codemlM1M2=="Y", tab$dginn.primate_

monddata$primates_young<-ifelse(tab$pVal.M8vsM7<0.05, 1, 0)
monddata$primate_cooper<-ifelse(tab$cooper.primates.M7.M8_p_value<0.05, 1, 0)
monddata$primates_dginn_full<-ifelse(dginnfulltmp>=3, 1,0)

mondrian(na.omit(monddata[,2:4]), labels=c("Young", "Cooper", "DGINN-full >=3" ))

```

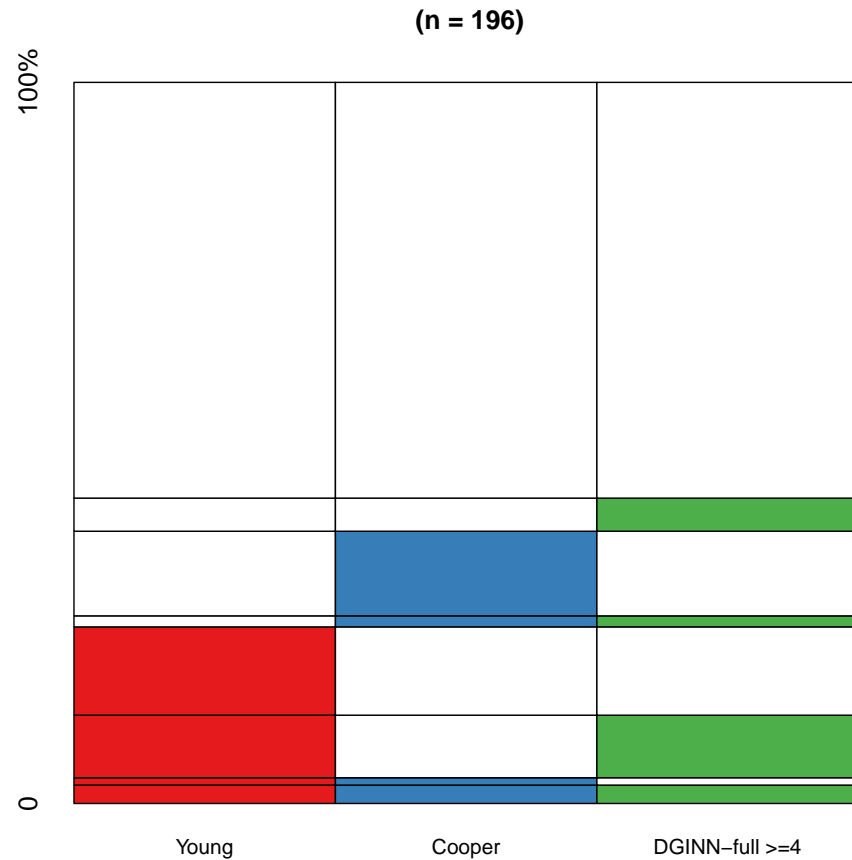


```

#####
monddata$primates_dginn_full<-ifelse(dginnfulltmp>=4, 1,0)

```

```
mondrian(na.omit(monddata[,2:4]), labels=c("Young", "Cooper", "DGINN-full >=4"))
```



3.2 subsetR

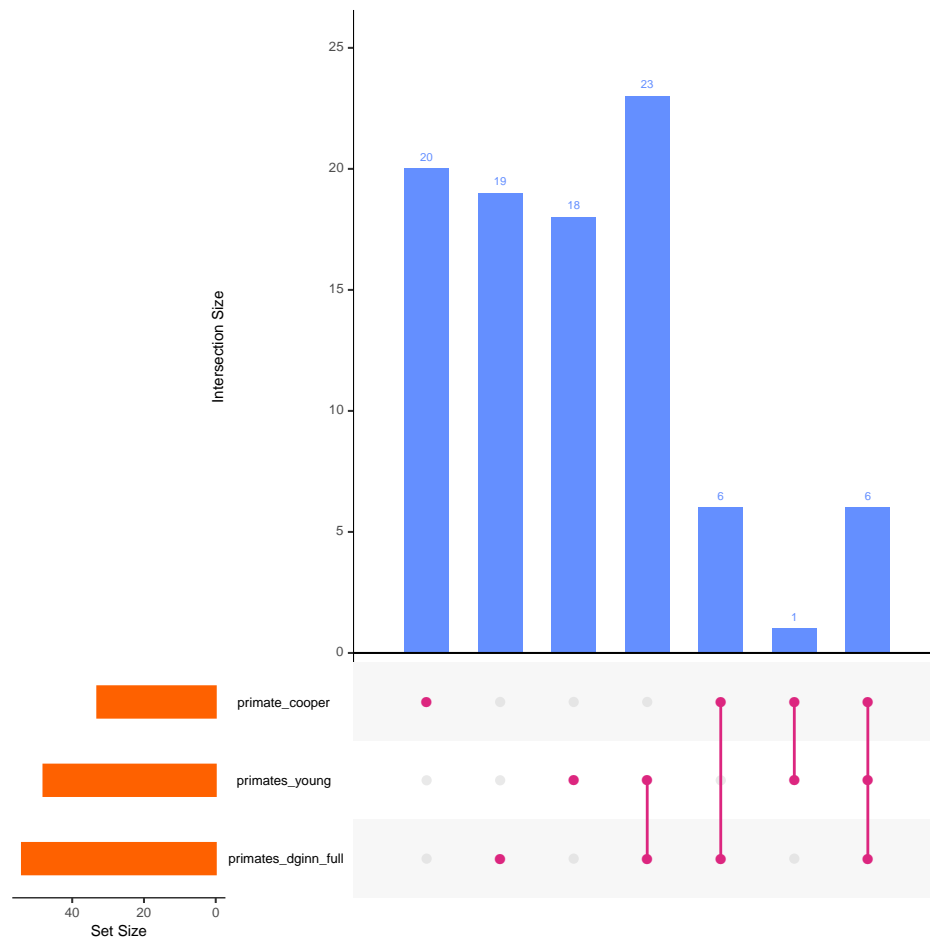
Just another representation of the same result.

```
library(UpSetR)
upsetdata<-as.data.frame(tab$Gene.name)

upsetdata$primates_young<-ifelse(tab$pVal.M8vsM7<0.05, 1, 0)
upsetdata$primate_cooper<-ifelse(tab$cooper.primates.M7.M8_p_value<0.05, 1, 0)
upsetdata$primates_dginn_full<-ifelse(dginnfulltmp>=3, 1,0)
```

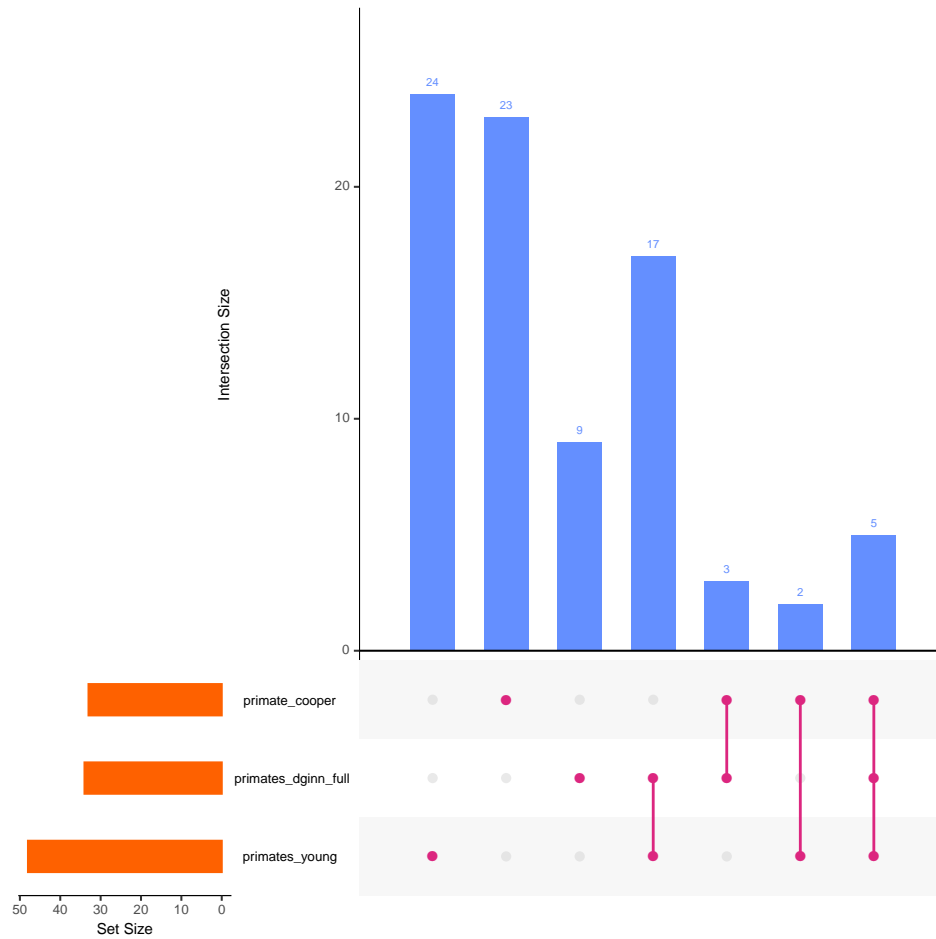


```
upset(na.omit(upsetdata), nsets = 3, matrix.color = "#DC267F",
      main.bar.color = "#648FFF", sets.bar.color = "#FE6100")
```



```
###
upsetdata$primates_dginn_full<-ifelse(dginnfulltmp>=4, 1,0)

upset(na.omit(upsetdata), nsets = 3, matrix.color = "#DC267F",
      main.bar.color = "#648FFF", sets.bar.color = "#FE6100")
```



4 Gene List

Genes under positive selection for at least 4 methods.

```
dginnfulltmp<-rowSums(cbind(tab$dginn.primate_BUSTED=="Y",
  tab$dginn.primate_BppM1M2=="Y",
  tab$dginn.primate_BppM7M8=="Y",
  tab$dginn.primate_codemlM1M2=="Y",
  tab$dginn.primate_codemlM7M8=="Y"))

tab$Gene.name[dginnfulltmp>=4 & is.na(dginnfulltmp)==F]
```

```
## [1] "ACADM"      "BCS1L"      "BRD4"      "CDK5RAP2"   "CEP135"
## [6] "CEP68"      "CLIP4"      "DNMT1"     "DPH5"       "EMC1"
## [11] "FYC01"     "GCC2"      "GGH"       "GHITM"      "GIGYF2"
## [16] "GLA"       "GOLGA7"    "HECTD1"   "IDE"        "ITGB1"
## [21] "LARP1"     "LARP4B"    "LMAN2"    "MARK1"     "MIPOL1"
## [26] "MPHOSPH10" "MYCBP2"    "NDUFAF2"  "NDUFB9"    "PCNT"
## [31] "POLA1"     "PRIM2"     "PRKAR2A"  "PVR"       "REEP6"
## [36] "RIPK1"     "SAAL1"     "SEPSECS"  "SIRT5"     "SLC25A21"
## [41] "SLC27A2"   "TMEM39B"   "TOR1AIP1" "TUBGCP2"   "UBAP2"
## [46] "UGGT2"     "VPS39"     "ZNF318"

tab$Gene.name[dginnfulltmp>=3 & is.na(dginnfulltmp)==F]

## [1] "ACADM"      "ADAM9"      "AP2A2"     "ATE1"       "BCS1L"
## [6] "BRD4"      "BZW2"      "CDK5RAP2"  "CEP135"     "CEP68"
## [11] "CLIP4"     "CNTRL"     "DNMT1"     "DPH5"       "EDEM3"
## [16] "EIF4E2"    "EMC1"      "EXOSC2"    "FYC01"      "GCC2"
## [21] "GGH"       "GHITM"     "GIGYF2"    "GLA"        "GOLGA7"
## [26] "GOLGB1"    "GORASP1"   "HDAC2"     "HECTD1"     "HS6ST2"
## [31] "IDE"       "ITGB1"     "LARP1"     "LARP4B"     "LARP7"
## [36] "LMAN2"     "MARK1"     "MDN1"     "MIPOL1"     "MOV10"
## [41] "MPHOSPH10" "MRPS5"     "MYCBP2"    "NAT14"      "NDUFAF2"
## [46] "NDUFB9"    "NGLY1"     "NPC2"      "PCNT"       "PITRM1"
## [51] "PLAT"      "PLOD2"     "PMPCB"     "POLA1"      "POR"
## [56] "PRIM2"     "PRKAR2A"   "PTBP2"     "PVR"        "RAB14"
## [61] "RAB1A"     "RAB2A"     "RAP1GDS1"  "RBX1"       "REEP6"
## [66] "RIPK1"     "RPL36"     "SAAL1"     "SCCPDH"     "SEPSECS"
## [71] "SIRT5"     "SLC25A21"  "SLC27A2"   "STOM"       "TIMM8B"
## [76] "TMEM39B"   "TOR1AIP1"  "TRIM59"    "TRMT1"      "TUBGCP2"
## [81] "UBAP2"     "UGGT2"     "USP54"     "VPS39"      "ZNF318"

tmp<-tab[dginnfulltmp>=4 & is.na(dginnfulltmp)==F,
c("Gene.name", "dginn.primate_BUSTED", "dginn.primate_BppM1M2",
  "dginn.primate_BppM7M8", "dginn.primate_codemlM1M2", "dginn.primate_codemlM7M8")]

write.table(tmp, "geneList_DGINN_full_primate_pos4.txt", row.names=F, quote=F)
```

5 Shiny like

```
makeFig1 <- function(df){  
  
  # prepare data for colors etc  
  colMethods <- c("deepskyblue4", "darkorange", "deepskyblue3", "mediumseagreen",  
  nameMethods <- c("BUSTED", "BppM1M2", "BppM7M8", "codemlM1M2", "codemlM7M8", "MEME"  
  metColor <- data.frame(Name = nameMethods, Col = colMethods, stringsAsFactors = F)  
  
  # subset for this specific figure  
  #df <- df[df$nbY >= 1, ] # to drop genes found by 0 methods (big datasets)  
  xt <- df[, c("BUSTED", "BppM1M2", "BppM7M8", "codemlM1M2", "codemlM7M8")]  
  xt$Gene <- df$Gene  
  nbrMeth <- 5  
  # reverse order of dataframe so that genes with the most Y are at the bottom (to be  
  xt[,1:5] <- ifelse(xt[,1:5] == "Y", 1, 0)  
  # sort and Filter the 0 lines  
  xt<-xt[order(rowSums(xt[,1:5])),]  
  xt<-xt[rowSums(xt[,1:5])>2,]  
  
  row.names(xt)<-xt$Gene  
  xt<-xt[,1:5]  
  
  colFig1 <- metColor[which(metColor$Name %in% colnames(xt)), ]  
  
  ##### PART 1 : NUMBER OF METHODS  
  par(xpd = NA, mar=c(2,7,4,0), oma = c(0,0,0,0), mgp = c(3,0.3,0))  
  
  h = barplot(  
    t(xt),  
    border = NA,  
    axes = F,  
    col = adjustcolor(colFig1$Col, alpha.f = 1),  
    horiz = T,  
    las = 2,  
    main = "Methods detecting positive selection",  
    cex.main = 0.85,  
    cex.names = min(50/nrow(xt), 1.5)  
  )
```

```

axis(3, line = 0, at = c(0:nbrMeth), label = c("0", rep("", nbrMeth - 1), nbrMeth),

legend("bottomleft",
      horiz = T,
      border = colFig1$Col,
      legend = colFig1$Name,
      fill = colFig1$Col,
      cex = 0.8,
      bty = "n",
      xpd = NA
    )
}

```

```
source("covid_comp_shiny.R")
```

```

df<-read.delim(paste0(workdir,
"/data/DGINN_202005281649summary_cleaned.csv"),
      fill=T, h=T, sep=",")

```

```
names(df)
```

```

## [1] "File"          "Name"          "Gene"
## [4] "GeneSize"      "NbSpecies"     "omegaM0Bpp"
## [7] "omegaM0codeml" "BUSTED"        "BUSTED.p.value"
## [10] "MEME.NbSites"  "MEME.PSS"      "BppM1M2"
## [13] "BppM1M2.p.value" "BppM1M2.NbSites" "BppM1M2.PSS"
## [16] "BppM7M8"       "BppM7M8.p.value" "BppM7M8.NbSites"
## [19] "BppM7M8.PSS"   "codemlM1M2"     "codemlM1M2.p.value"
## [22] "codemlM1M2.NbSites" "codemlM1M2.PSS" "codemlM7M8"
## [25] "codemlM7M8.p.value" "codemlM7M8.NbSites" "codemlM7M8.PSS"

```

```

dftmp<-tab[,c("File", "Name", "Gene.name",
"GeneSize", "dginn.primate_NbSpecies", "dginn.primate_omegaM0Bpp",
"dginn.primate_omegaM0codeml", "dginn.primate_BUSTED", "dginn.primate_BUSTED.p",
"dginn.primate_MEME.NbSites", "dginn.primate_MEME.PSS", "dginn.primate_BppM1M2",
"dginn.primate_BppM1M2.p.value", "dginn.primate_BppM1M2.NbSites", "dginn.prima",
"dginn.primate_BppM7M8", "dginn.primate_BppM7M8.p.value", "dginn.primate_BppM7M8.NbSites", "dginn.primate_BppM7M8.PSS", "dginn.primate_BppM7M8.p.value", "dginn.primate_BppM7M8.NbSites", "dginn.primate_BppM7M8.PSS")

```

```
"dginn.primate_BppM7M8.PSS", "dginn.primate_codemlM1M2", "dginn.primate_codemlM1M2.NbSites", "dginn.primate_codemlM1M2.PSS", "dginn.primate_codemlM7M8.p.value", "dginn.primate_codemlM7M8.NbSites" , "dginn.primate_codemlM7M8.PSS"
```

```
names(dftmp)<-names(df)
makeFig1(dftmp)
```

