

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

Marie Cariou

March 2021

## Contents

<b>1</b>	<b>Data</b>	<b>2</b>
1.1	Environment . . . . .	2
1.2	Necessary packages . . . . .	2
1.3	Read DGINN Young table . . . . .	2
<b>2</b>	<b>Omega comparisons</b>	<b>3</b>
<b>3</b>	<b>Mondrian</b>	<b>5</b>

# 1 Data

## 1.1 Environment

Analysis were formatted by the script covid.comp\_script0\_table.Rnw.

```
home<-"/home/adminmarie/Documents/CIRI_BIBS_projects/"
workdir<-paste0(home, "2020_05_Etienne_covid/2020_dginn_covid19/")
```

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

## [1] 332 141

tab$Gene.name<-as.character(tab$Gene.name.x)
tab$Gene.name[tab$PreyGene=="MARC1"]<-"MARC1"
```

## 1.2 Necessary packages

```
library(Mondrian)
library(UpSetR)
```

## 1.3 Read DGINN Young table

DGINN-Young-primate table correspond to DGINN results, on the SAME alignment as Young-primate.

I will merge the 2 tables.

```
dginnY<-read.delim(paste0(workdir,
  "data/summary_primate_young.res"),
  fill=T, h=T)
dim(dginnY)

## [1] 1992 7

names(dginnY)

## [1] "Gene" "Omega" "Method" "PosSel" "PValue" "NbSites"
## [7] "PSS"
```

```

add_col<-function(method="PamLM1M2"){

tmp<-dginnY[dginnY$Method==method,
             c("Gene", "Omega", "PosSel", "PValue", "NbSites", "PSS")]

names(tmp)<-c("Gene.name", paste0("Omega_", method),
             paste0("PosSel_", method), paste0("PValue_", method),
             paste0("NbSites_", method), paste0("PSS_", method))

tab<-merge(tab, tmp, by="Gene.name")

return(tab)
}

tab<-add_col("PamLM1M2")
tab<-add_col("PamLM7M8")
tab<-add_col("BppM1M2")
tab<-add_col("BppM7M8")

# Manip pour la colonne BUSTED

tmp<-dginnY[dginnY$Method=="BUSTED",c("Gene", "Omega", "PosSel", "PValue")]
names(tmp)<-c("Gene.name", "Omega_BUSTED", "PosSel_BUSTED", "PValue_BUSTED")
tab<-merge(tab, tmp, by="Gene.name")

tmp<-dginnY[dginnY$Method=="MEME",c("Gene", "NbSites", "PSS")]
names(tmp)<-c("Gene.name", "NbSites_MEME", "PSS_MEME")
tab<-merge(tab, tmp, by="Gene.name")

dim(tab)

## [1] 332 167

```

## 2 Omega comparisons

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.

```

tab$whole.gene.dN.dS.model.0<-as.numeric(as.character(
  tab$whole.gene.dN.dS.model.0))
tab$'dginn.primate_omegaMOBpp'[tab$'dginn.primate_omegaMOBpp'=="na"]<-NA
tab$'dginn.primate_omegaMOBpp'<-as.numeric(as.character(
  tab$'dginn.primate_omegaMOBpp'))

plot(tab$whole.gene.dN.dS.model.0,
      tab$'dginn.primate_omegaMOBpp',
      xlab="Omega Young-primate", ylab="DGINN-full's")
abline(0,1)
abline(lm(tab$'dginn.primate_omegaMOBpp'~tab$whole.gene.dN.dS.model.0), col="red")

outgenes<-c("TCEB1", "HSBP1", "REEP6", "RPL36")

outlier<-tab[tab$Gene.name.x %in% outgenes,]
text(x=outlier$whole.gene.dN.dS.model.0+0.05,
y=outlier$'dginn.primate_omegaMOBpp'+0.02,
outlier$Gene.name.x)

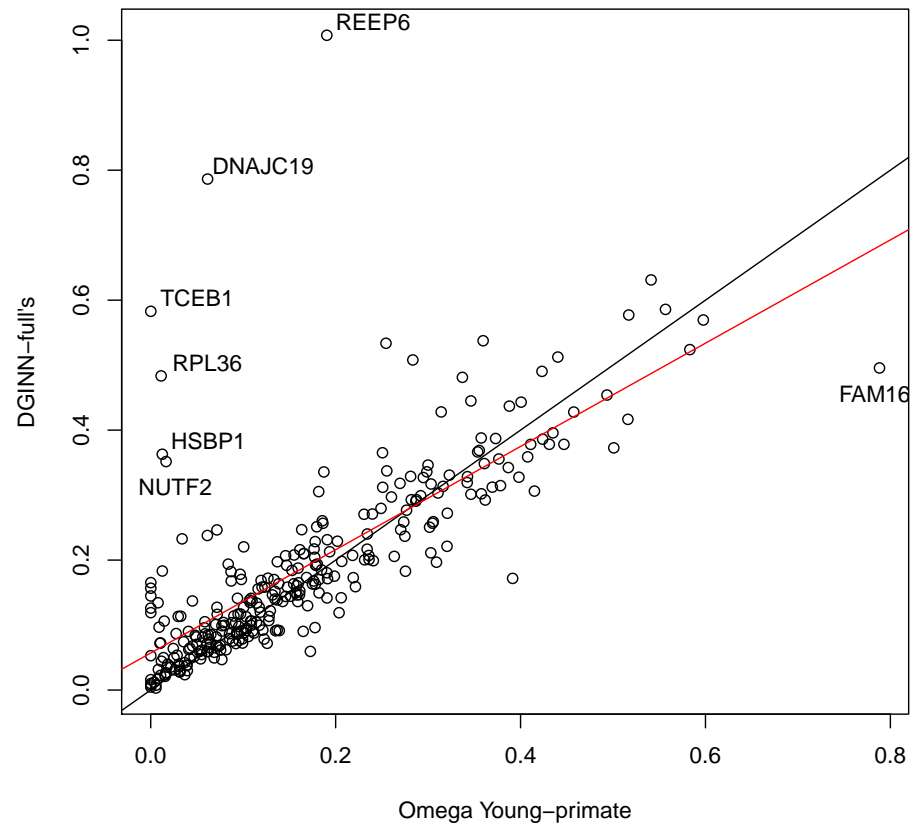
outgenes<-c("DNAJC19")

outlier<-tab[tab$Gene.name.x %in% outgenes,]
text(x=outlier$whole.gene.dN.dS.model.0+0.06,
y=outlier$'dginn.primate_omegaMOBpp'+0.02,
outlier$Gene.name.x)

outgenes<-c("NUTF2", "FAM162A")

outlier<-tab[tab$Gene.name.x %in% outgenes,]
text(x=outlier$whole.gene.dN.dS.model.0+0.01,
y=outlier$'dginn.primate_omegaMOBpp'-0.04,
outlier$Gene.name.x)

```



### 3 Mondrian

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

## [1] 332 1

tab$pVal.M8vsM7<-as.numeric(as.character(tab$pVal.M8vsM7))

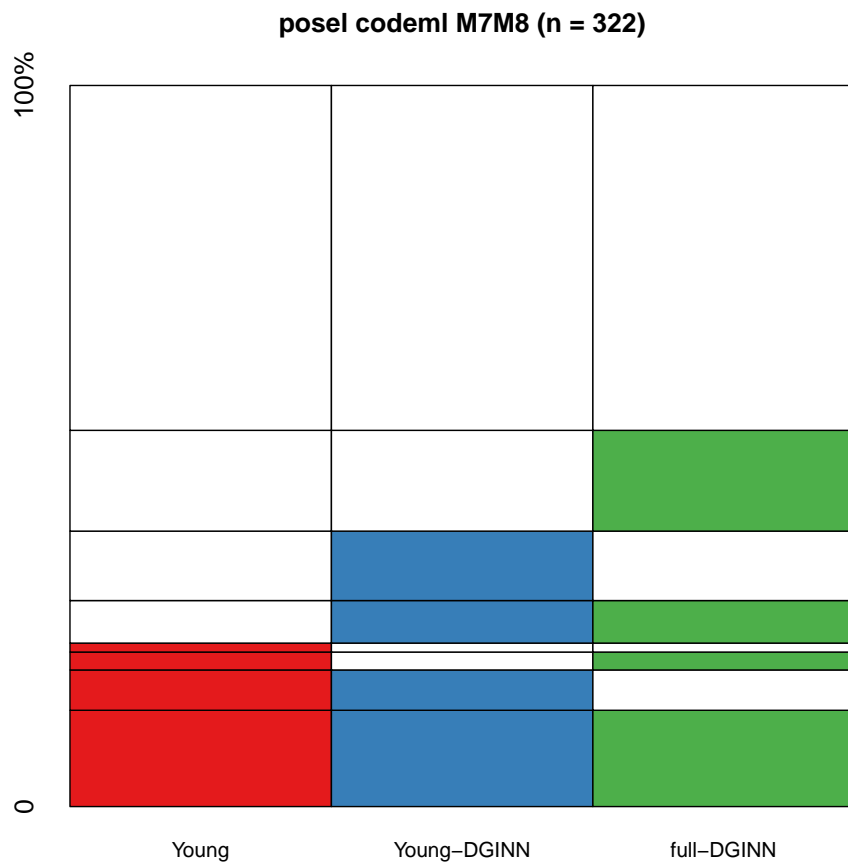
monddata$primates_young<-ifelse(tab$pVal.M8vsM7<0.05, 1, 0)
```

```

monddata$primates_dginn_young<-tab$PosSel_BppM7M8=="Y"
monddata$primates_dginn_full<-tab$'dginn.primate_codemlM7M8'=="Y"

mondrian(na.omit(monddata[,2:4]),
          labels=c("Young", "Young-DGINN", "full-DGINN"),
          main="posel codeml M7M8")

```



```

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

## [1] 332 1

```

```

tab$pVal.M2vsM1<-as.numeric(as.character(tab$pVal.M2vsM1))

monddata$primates_young<-ifelse(tab$pVal.M8vsM7.adj<0.05, 1, 0)

monddata$primates_dginn_young<-tab$PosSel_BppM7M8=="Y"
monddata$primates_dginn_full<-tab$'dginn.primate_codemlM7M8'=="Y"

mondrian(na.omit(monddata[,2:4]),
          labels=c("Young", "Young-DGINN", "full-DGINN"),
          main="posel codeml M1M2")

```

