

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

Marie Cariou

Mars 2021

Contents

1	Data	2
2	Comparison Bats	2
2.1	Cooper-bats results VS DGINN-bats results	2
2.2	Cooper-bats VS Hawkins-bats and DGINN-bats VS Hawkins-bats	3
3	Overlap	3
3.1	Data	3
3.2	Mondrian	4
3.3	subsetR	6

1 Data

Analysis were formatted by the script covid_comp_script0_table.Rnw.

```
home<-" /home/adminmarie/Documents/"
workdir<-paste0(home,"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] 332 141

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

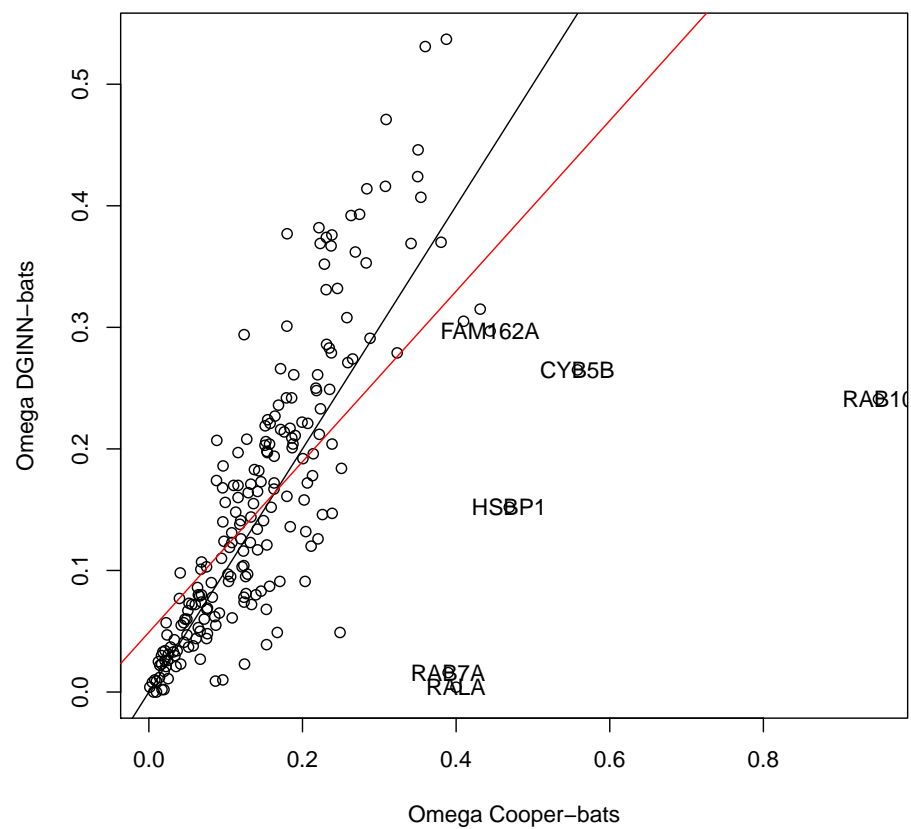
2 Comparison Bats

2.1 Cooper-bats results VS DGINN-bats results

```
tab$bats_omegaM0codeml[tab$bats_omegaM0codeml=="na"]<-NA

plot(tab$cooper.batsAverage_dNdS,
  as.numeric(as.character(tab$bats_omegaM0codeml)),
  xlab="Omega Cooper-bats",
  ylab="Omega DGINN-bats")
abline(0,1)
abline(lm(as.numeric(as.character(tab$bats_omegaM0codeml))~
  tab$cooper.batsAverage_dNdS),
  col="red")

outlier<-tab[tab$cooper.batsAverage_dNdS>0.35 &
  as.numeric(as.character(tab$bats_omegaM0codeml))<0.3,]
text(x=outlier$cooper.batsAverage_dNdS,
y=as.numeric(as.character(outlier$bats_omegaM0codeml)),
outlier$Gene.name)
```



2.2 Cooper-bats VS Hawkins-bats and DGINN-bats VS Hawkins-bats

3 Overlap

3.1 Data

```
tmp<-na.omit(tab[,c("Gene.name", "bats_codemlM7M8_p.value",
                    "hawkins_Positive.Selection..M8vM8a.p.value",
                    "cooper.batsM7.M8_p_value", "bats_BUSTED",
                    "bats_BppM1M2", "bats_BppM7M8", "bats_codemlM1M2",
                    "bats_codemlM7M8")])
```

```

tmp$bats_codemlM7M8_p.value[tmp$bats_codemlM7M8_p.value=="na"]<-NA
tmp$bats_codemlM7M8_p.value<-as.numeric(
  as.character(tmp$bats_codemlM7M8_p.value))
dim(tmp)

## [1] 174    9

```

170 genes (present in the 3 experiments)

3.2 Mondrian

```

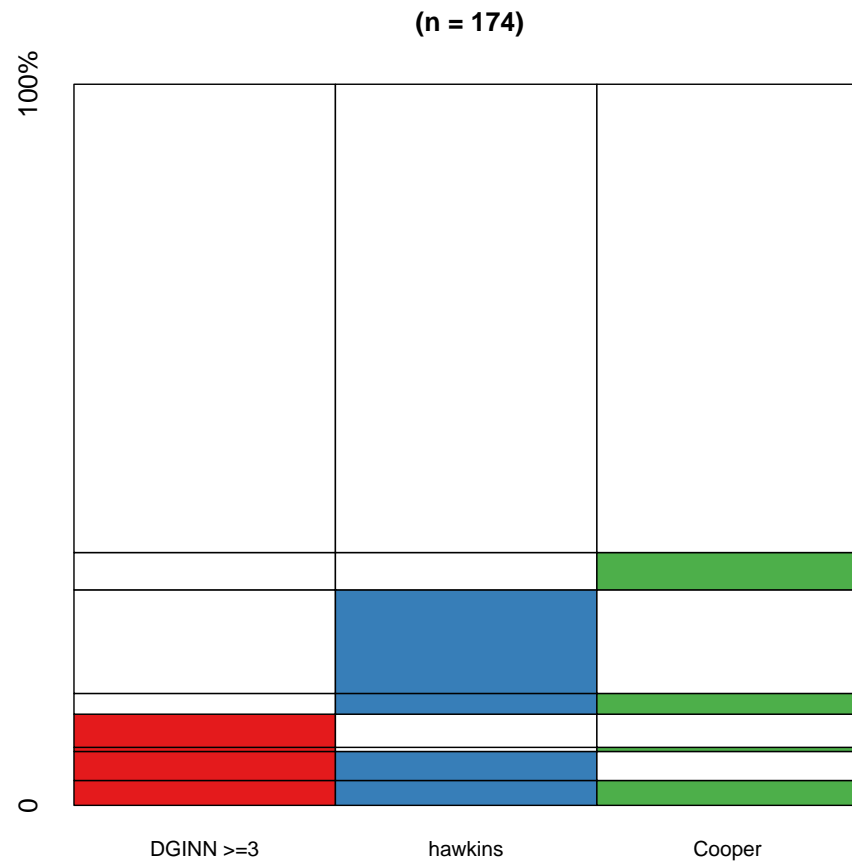
library(Mondrian)

monddata<-as.data.frame(tmp$Gene.name)
monddata$bats_hawkins<-ifelse(
  tmp$hawkins_Positive.Selection..M8vM8a.p.value<0.05, 1, 0)
monddata$bats_cooper<-ifelse(
  tmp$cooper.batsM7.M8_p_value<0.05, 1, 0)

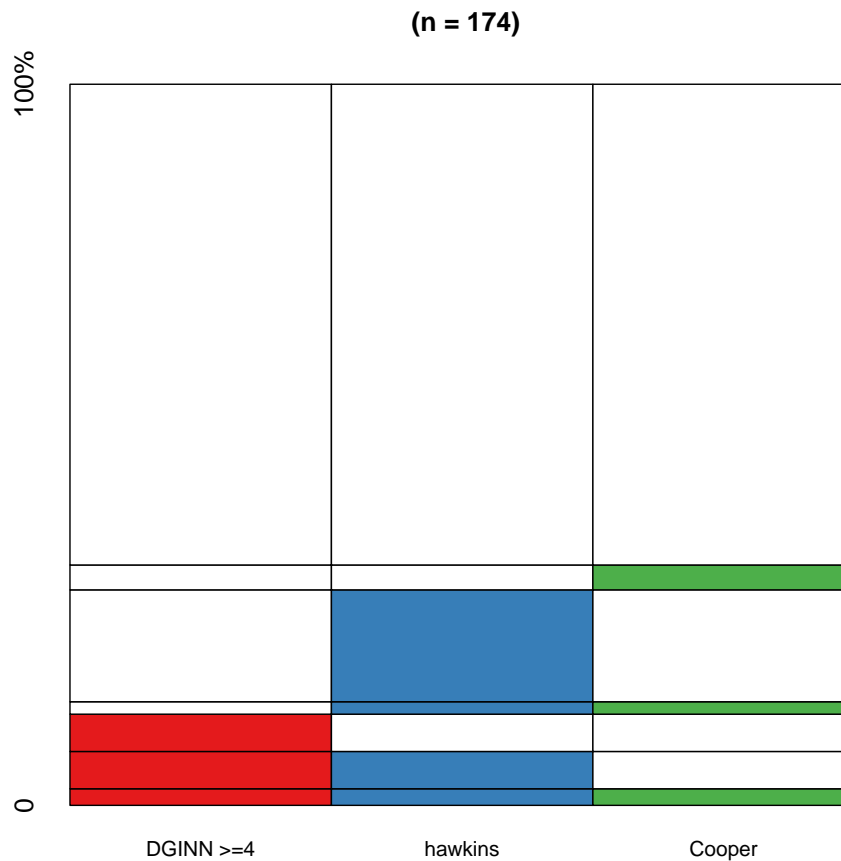
dginntmp<-rowSums(cbind(tmp$bats_codemlM1M2=="Y",
                        tmp$bats_codemlM7M8=="Y",
                        tmp$bats_BppM1M2=="Y",
                        tmp$bats_BppM7M8=="Y",
                        tmp$bats_BUSTED=="Y"))

monddata$bats_dginn<-ifelse(dginntmp>=3, 1,0)
mondrian(monddata[,2:4],
  labels=c("DGINN >=3", "hawkins", "Cooper"))

```



```
monddata$bats_dginn<-ifelse(dginntmp>=4, 1,0)
mondrian(monddata[,2:4],
         labels=c("DGINN >=4", "hawkins", "Cooper"))
```



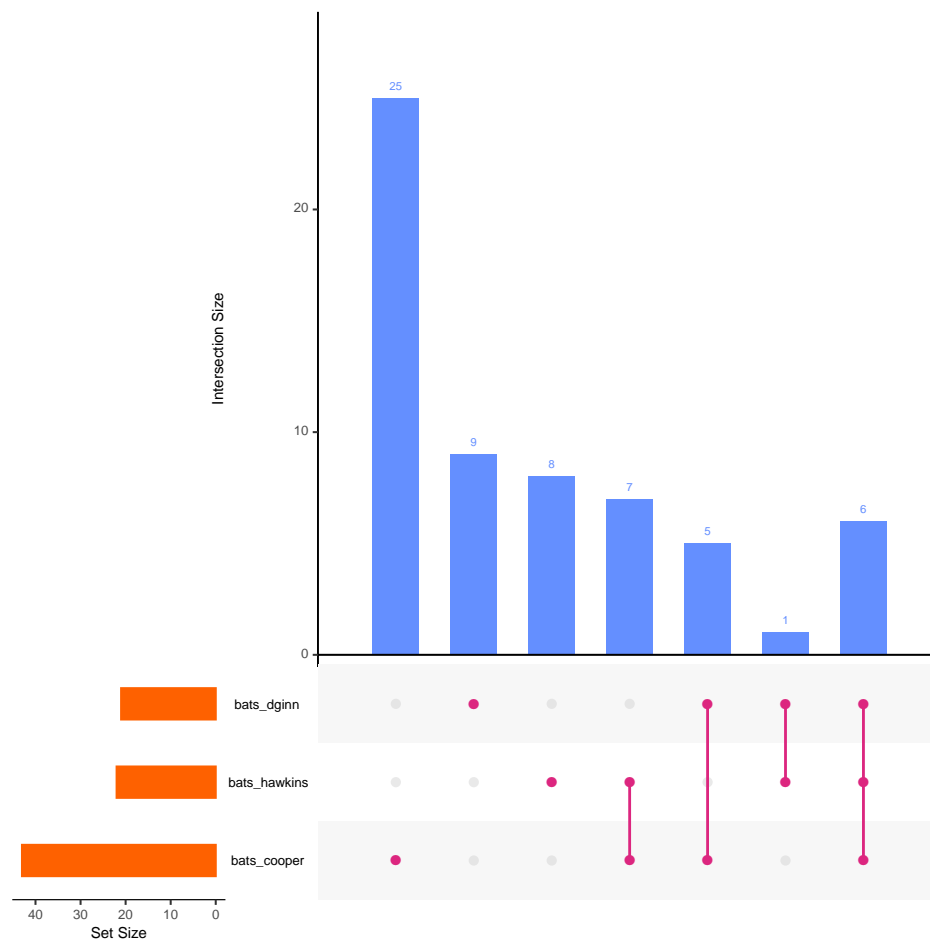
3.3 subsetR

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

upsetdata$bats_hawkins<-ifelse(
  tmp$hawkins_Positive.Selection..M8vM8a.p.value<0.05, 1, 0)
upsetdata$bats_cooper<-ifelse(
  tmp$cooper.batsM7.M8_p_value<0.05, 1, 0)

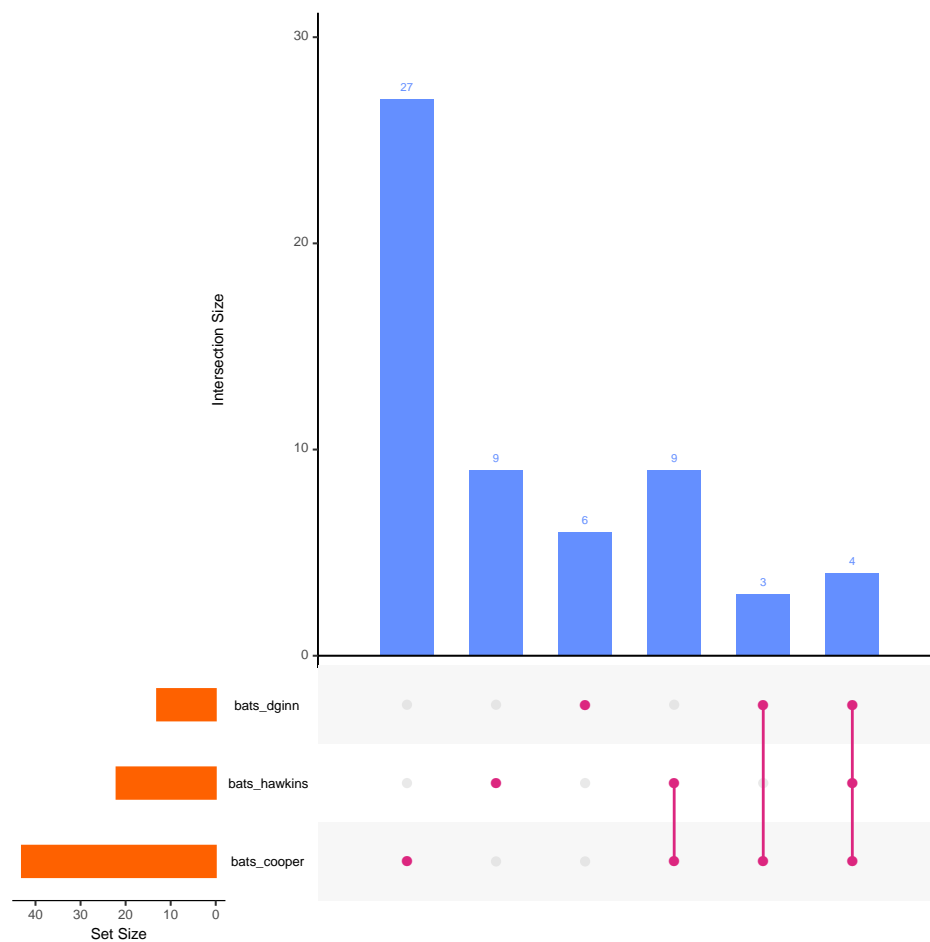
upsetdata$bats_dginn<-ifelse(dginn$tmp>=3, 1,0)
```

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



```
upsetdata$bats_dginn<-ifelse(dginntmp>=4, 1,0)

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



```
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("bats_File", "bats_Name",
  "Gene.name", "bats_GeneSize",
  "bats_NbSpecies", "bats_omegaM0Bpp",
  "bats_omegaM0codeml", "bats_BUSTED",
  "bats_BUSTED_p.value", "bats_MEME_NbSites",
  "bats_MEME_PSS", "bats_BppM1M2",
  "bats_BppM1M2_p.value", "bats_BppM1M2_NbSites",
  "bats_BppM1M2_PSS", "bats_BppM7M8",
  "bats_BppM7M8_p.value", "bats_BppM7M8_NbSites",
  "bats_BppM7M8_PSS", "bats_codemlM1M2",
  "bats_codemlM1M2_p.value", "bats_codemlM1M2_NbSites",
  "bats_codemlM1M2_PSS", "bats_codemlM7M8",
  "bats_codemlM7M8_p.value", "bats_codemlM7M8_NbSites" ,
  "bats_codemlM7M8_PSS")]

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

