

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

Marie Cariou

March 2021

Contents

1	Files manipulations	2
1.1	Complete table	2
1.2	Read DGINN Young table	2
2	Comparisons Primates	3
2.1	DGINN results on Janet Young's alignments (DGINN-Young-primate) VS Janet Young's results	3
2.2	DGINN results on Janet Young's alignments (DGINN-Young-primate) VS DGINN-full's results	5
2.3	Janet Young's results (Young-primate) VS DGINN-full's results	7
3	Overlap	9
3.1	Mondrian	9
3.2	subsetR	13
4	Gene List	14

1 Files manipulations

1.1 Complete table

```
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=="MARC1"]<-"MARC1"
```

1.2 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 Comparisons Primates

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

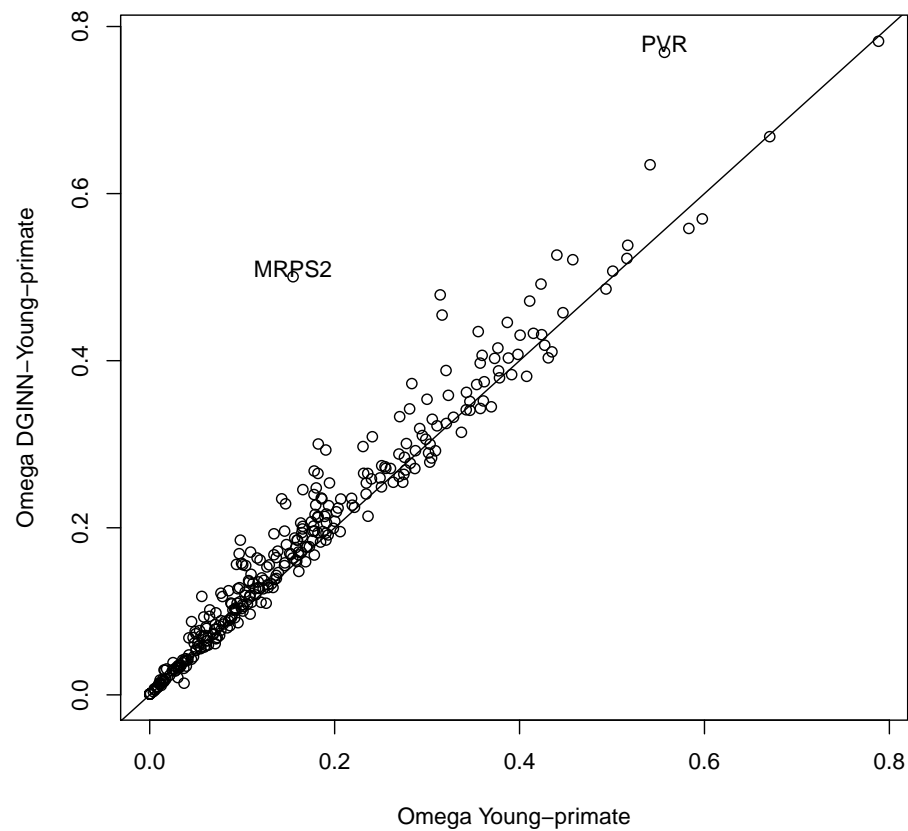
Comparaison des Omega: colonne L "whole.gene.dN.dS.model.0" VS colonne "omega" dans la sortie de dginn.

```

tab$whole.gene.dN.dS.model.0<-as.numeric(
  as.character(tab$whole.gene.dN.dS.model.0))
plot(tab$whole.gene.dN.dS.model.0, tab$Omega_PamlM7M8,
      xlab="Omega Young-primate", ylab="Omega DGINN-Young-primate")
abline(0,1)
outlier<-tab[tab$whole.gene.dN.dS.model.0<0.2 & tab$Omega_PamlM7M8>0.4,]
text(x=outlier$whole.gene.dN.dS.model.0,
     y=(outlier$Omega_PamlM7M8+0.01),
     outlier$Gene.name)

outlier<-tab[tab$whole.gene.dN.dS.model.0<0.6 & tab$Omega_PamlM7M8>0.7,]
text(x=outlier$whole.gene.dN.dS.model.0,
     y=(outlier$Omega_PamlM7M8+0.01),
     outlier$Gene.name)

```



2.2 DGINN results on Janet Young's alignments (DGINN-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$'dginn.primite_omegaMOBpp'<-as.numeric(
  as.character(tab$'dginn.primite_omegaMOBpp'))

## Warning:  NAs introduits lors de la conversion automatique

plot(tab$'dginn.primite_omegaMOBpp', tab$Omega_Pam1M7M8,
      xlab="DGINN-full's", ylab="Omega DGINN-Young-primate")
```

```

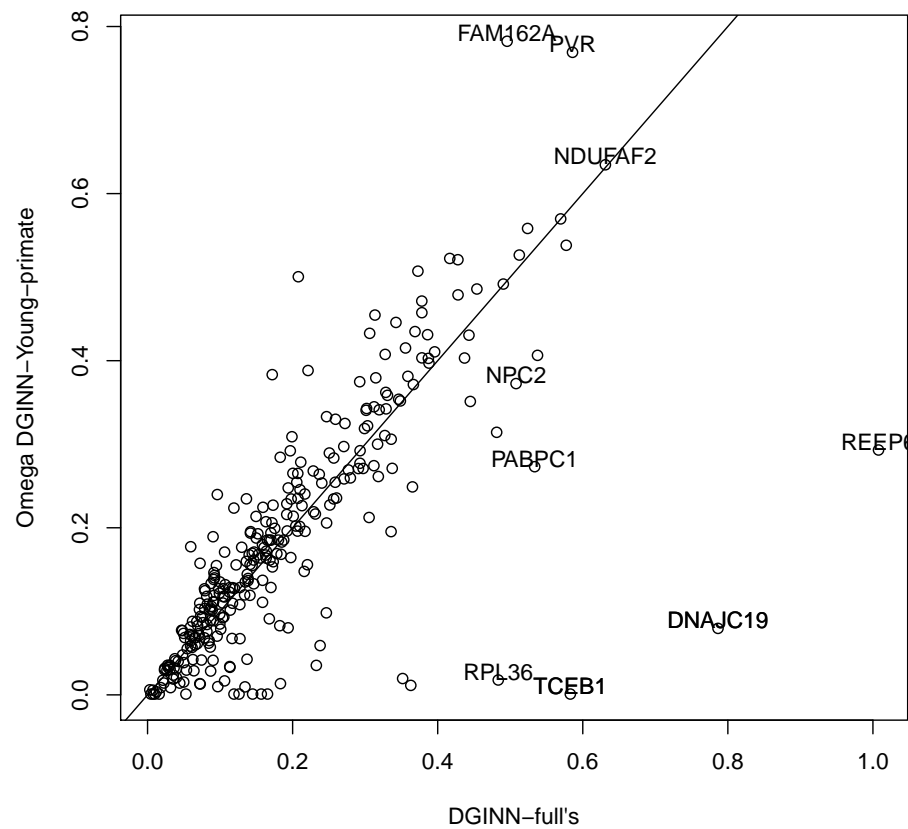
abline(0,1)

outlier<-tab[tab$'dginn.primate_omegaM0Bpp'>0.4 & tab$Omega_PamlM7M8<0.2,]
text(x=outlier$'dginn.primate_omegaM0Bpp',
y=(outlier$Omega_PamlM7M8+0.01),
outlier$Gene.name)

outlier<-tab[tab$'dginn.primate_omegaM0Bpp'>0.5 & tab$Omega_PamlM7M8<0.4,]
text(x=outlier$'dginn.primate_omegaM0Bpp',
y=(outlier$Omega_PamlM7M8+0.01),
outlier$Gene.name)

outlier<-tab[tab$'dginn.primate_omegaM0Bpp'>0.2 & tab$Omega_PamlM7M8>0.6,]
text(x=outlier$'dginn.primate_omegaM0Bpp',
y=(outlier$Omega_PamlM7M8+0.01),
outlier$Gene.name)

```



2.3 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.primite_omegaMOBpp')),
      xlab="Omega Young-primate", ylab="DGINN-full's")
abline(0,1)
```

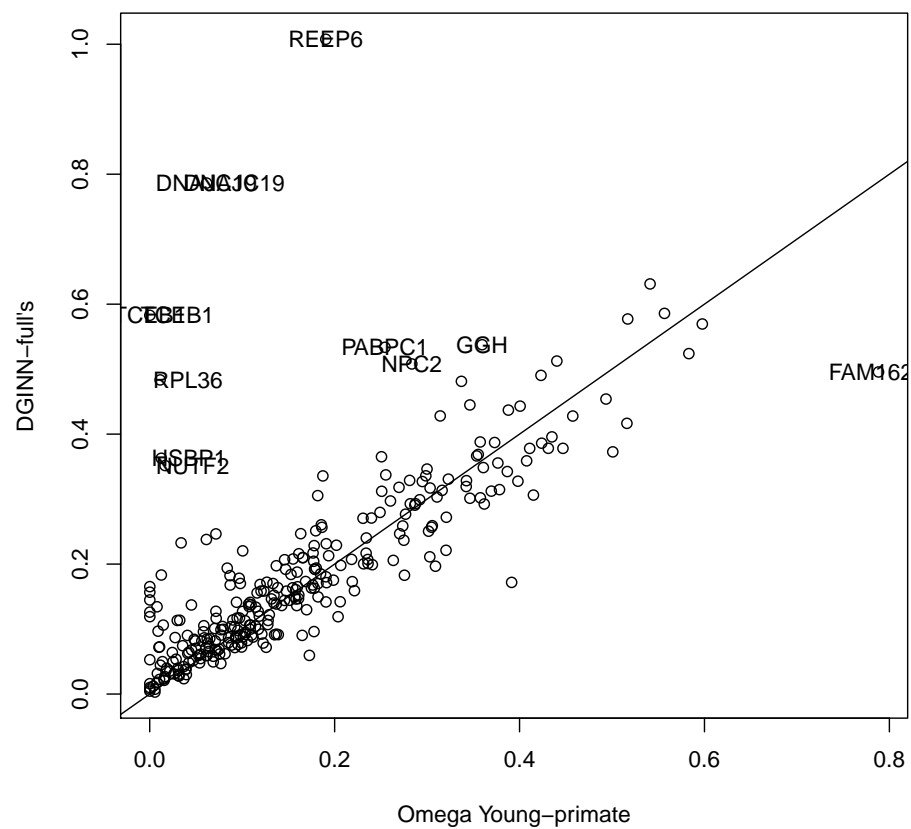
```

outlier<-tab[tab$whole.gene.dN.dS.model.0<0.4 &
  as.numeric(as.character(tab$'dginn.primate_omegaMOBpp'))>0.5,]
text(x=outlier$whole.gene.dN.dS.model.0,
y=outlier$'dginn.primate_omegaMOBpp',
outlier$Gene.name)

outlier<-tab[tab$whole.gene.dN.dS.model.0>0.7 &
  as.numeric(as.character(tab$'dginn.primate_omegaMOBpp'))>0,]
text(x=outlier$whole.gene.dN.dS.model.0,
y=outlier$'dginn.primate_omegaMOBpp',
outlier$Gene.name)

outlier<-tab[tab$whole.gene.dN.dS.model.0<0.1 &
  as.numeric(as.character(tab$'dginn.primate_omegaMOBpp'))>0.3,]
text(x=outlier$whole.gene.dN.dS.model.0+0.03,
y=outlier$'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

dginnyoungtmp<-rowSums(cbind(tab$PosSel_PamlM1M2=="Y",
```

```

        tab$PosSel_PamlM7M8=="Y",
        tab$PosSel_BppM1M2=="Y",
        tab$PosSel_BppM7M8=="Y",
        tab$PosSel_BUSTED=="Y"))

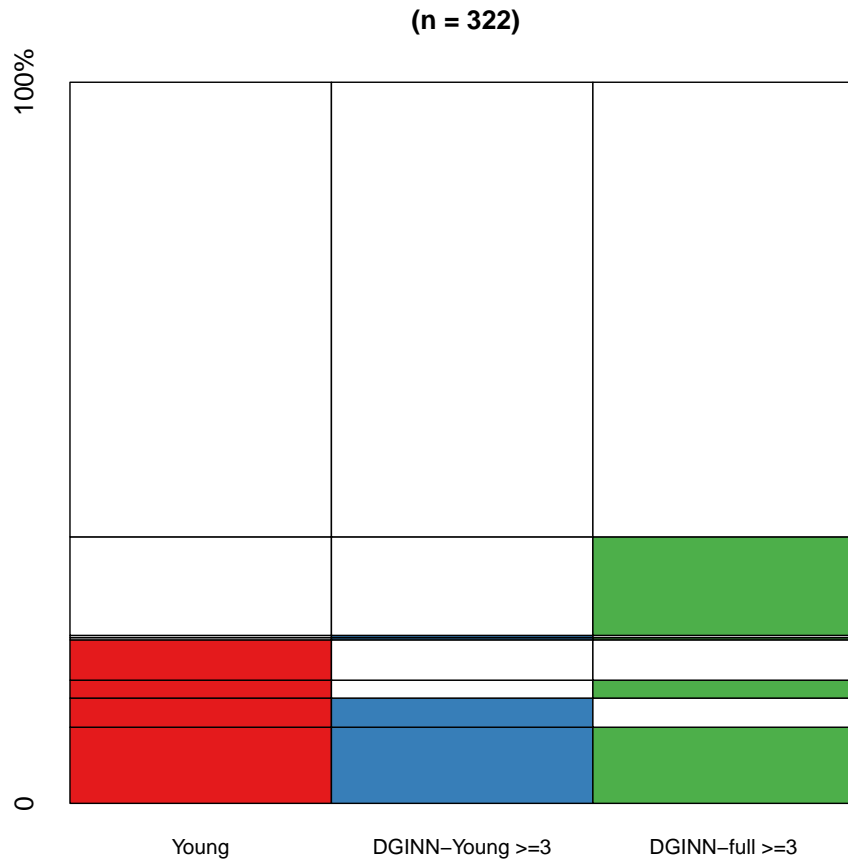
dginnfulltmp<-rowSums(cbind(tab$'dginn.primate_BUSTED'=="Y",
        tab$'dginn.primate_BppM1M2'=="Y",
        tab$'dginn.primate_BppM7M8'=="Y",
        tab$'dginn.primate_codemlM1M2'=="Y",
        tab$'dginn.primate_codemlM7M8'=="Y"))

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

monddata$primates_dginn_young<-ifelse(dginnyoungtmp>=3, 1,0)
monddata$primates_dginn_full<-ifelse(dginnfulltmp>=3, 1,0)

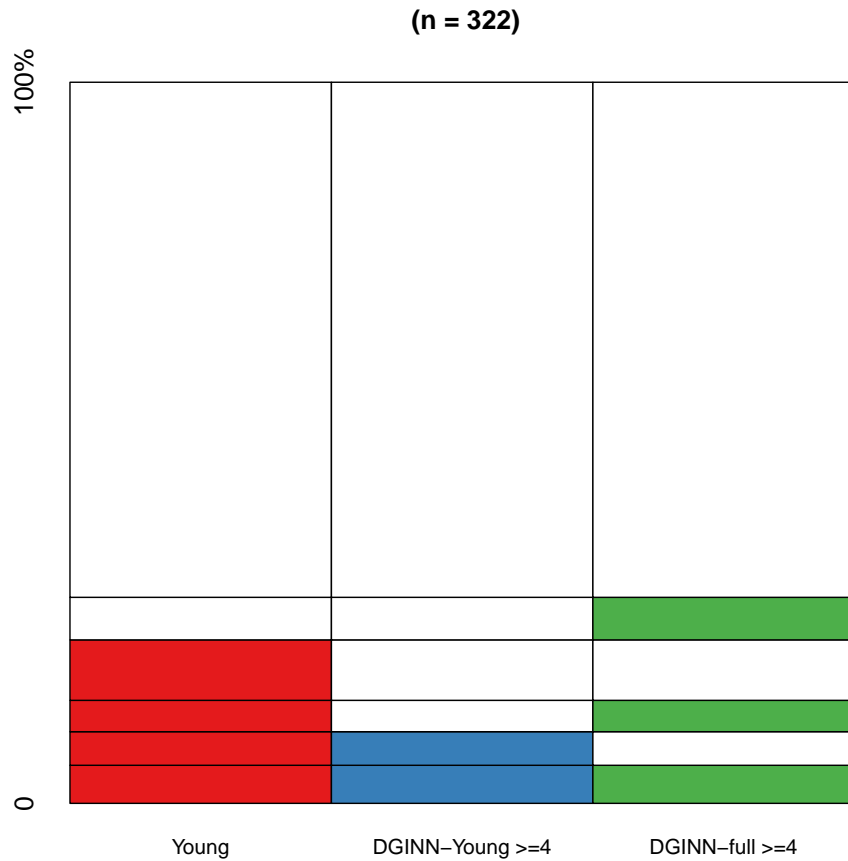
mondrian(na.omit(monddata[,2:4]),
        labels=c("Young", "DGINN-Young >=3", "DGINN-full >=3" ))

```



```
monddata$primates_dginn_young<-ifelse(dginnyoungtmp>=4, 1,0)
monddata$primates_dginn_full<-ifelse(dginnfulltmp>=4, 1,0)

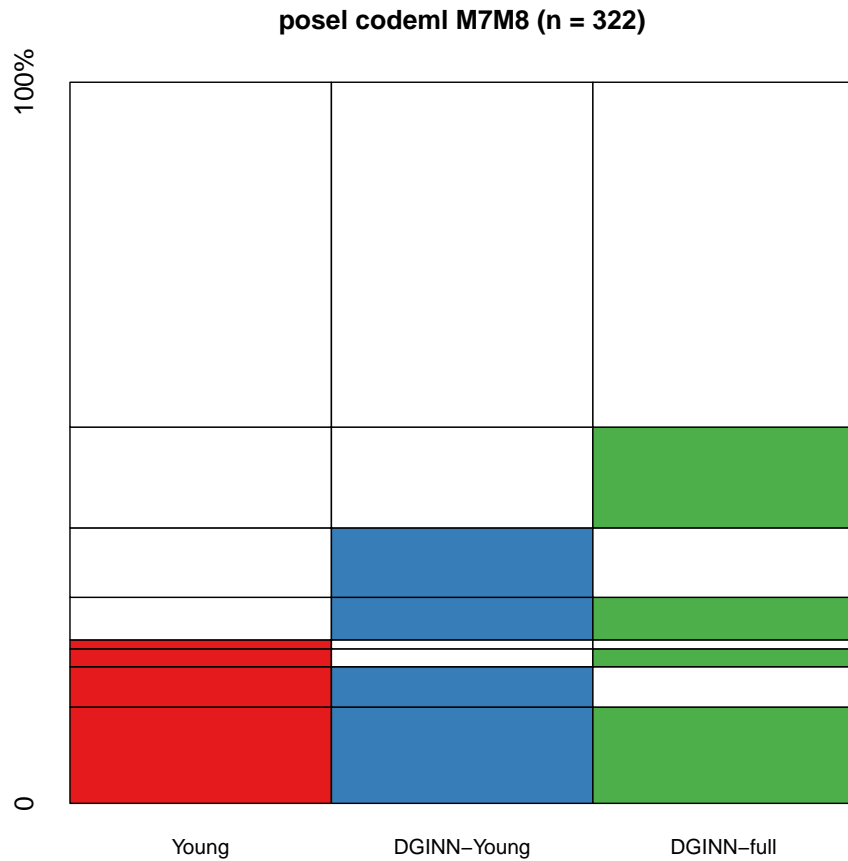
mondrian(na.omit(monddata[,2:4]),
         labels=c("Young", "DGINN-Young >=4", "DGINN-full >=4"))
```



Comparison of results with the same method.

```
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", "DGINN-Young", "DGINN-full"),
  main="posel codeml M7M8")
```



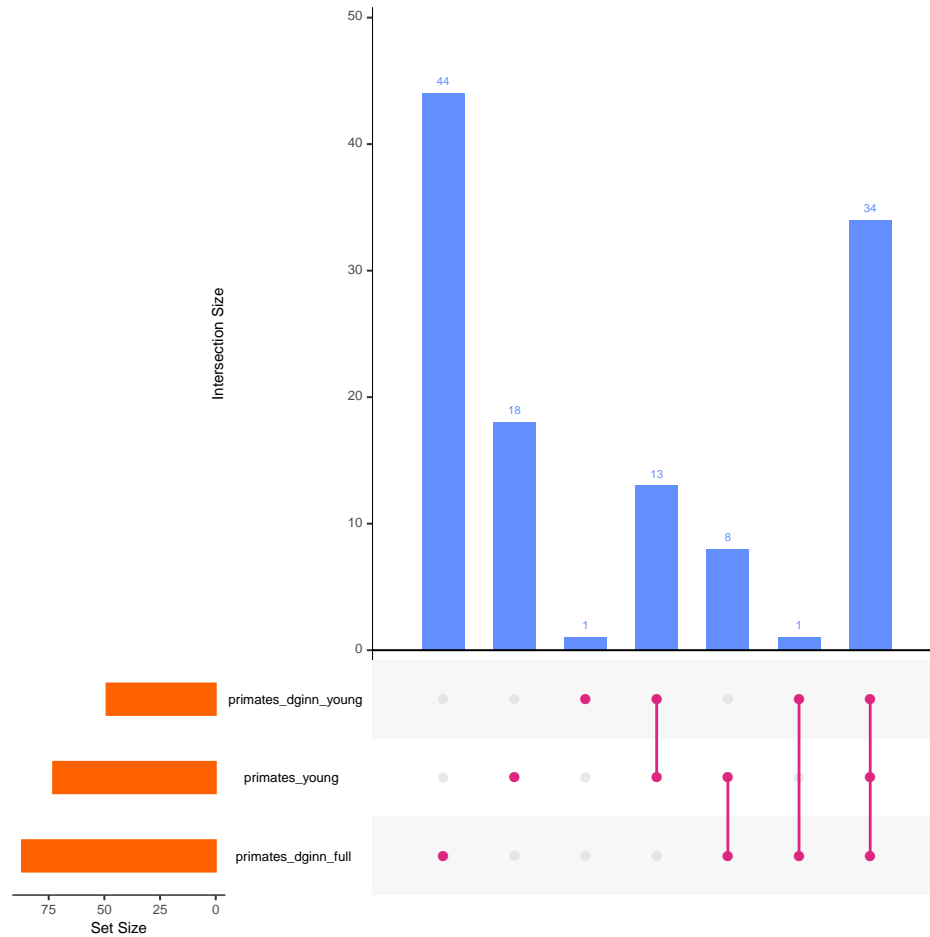
3.2 subsetR

Just another representation of the same result, for now, I focus on the gene positive in 3 methodes for DGINN analysis.

```
library(UpSetR)
upsetdata<-as.data.frame(tab$Gene.name)
upsetdata$primates_young<-ifelse(tab$Val.M8vsM7<0.05, 1, 0)

###
upsetdata$primates_dginn_young<-ifelse(dginnyoungtmp>=3, 1,0)
upsetdata$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")
```



4 Gene List

List of the 34 genes found under positive selection in all analysis.

```
upsetdata$`tab$Gene.name`[(upsetdata$primates_young==TRUE &
                             upsetdata$primates_dginn_young==TRUE &
                             upsetdata$primates_dginn_full==TRUE)]
```

##	[1]	ACADM	ATE1	BCS1L	BRD4	CDK5RAP2	CEP68

```
## [7] CNTRL      DNMT1      EDEM3      FYC01      GCC2      GHITM
## [13] GIGYF2     GOLGB1     GORASP1     ITGB1     MDN1     MPHOSPH10
## [19] MRPS5      NDUFAF2    PCNT       PLAT      POLA1     PRIM2
## [25] PVR        SAAL1      SEPSECS     SIRT5     SLC25A21  SLC27A2
## [31] TOR1AIP1   UGGT2      USP54      ZNF318
## 332 Levels: AAR2 AASS AATF ABCC1 ACAD9 ACADM ACSL3 ADAM9 ... ZYG11B
```

List of the 13 genes found under positive selection in both Young analysis and DGINN-Young alignments (but not full-DGINN).

```
upsetdata$`tab$Gene.name`[(upsetdata$primates_young==TRUE &
                             upsetdata$primates_dginn_young==TRUE &
                             upsetdata$primates_dginn_full==FALSE)]

## [1] ABCC1     ALG8      CEP250    CEP350    ERLEC1    FASTKD5   GOLGA2    MRPS27
## [9] NINL      PDE4DIP   PRRC2B    RAB18     WFS1
## 332 Levels: AAR2 AASS AATF ABCC1 ACAD9 ACADM ACSL3 ADAM9 ... ZYG11B
```

List of the 1 gene found under positive selection in both DGINN analysis, but not Young.

```
upsetdata$`tab$Gene.name`[(upsetdata$primates_young==FALSE &
                             upsetdata$primates_dginn_young==TRUE &
                             upsetdata$primates_dginn_full==TRUE)]

## [1] MARK1
## 332 Levels: AAR2 AASS AATF ABCC1 ACAD9 ACADM ACSL3 ADAM9 ... ZYG11B
```

List of the 8 genes found under positive selection in both Young analysis and full-DGINN, but not DGINN-young.

```
upsetdata$`tab$Gene.name`[(upsetdata$primates_young==TRUE &
                             upsetdata$primates_dginn_young==FALSE &
                             upsetdata$primates_dginn_full==TRUE)]

## [1] <NA>      LMAN2     NDUFB9    RIPK1     STOM      TMEM39B   TRMT1     UBAP2
## [9] VPS39
## 332 Levels: AAR2 AASS AATF ABCC1 ACAD9 ACADM ACSL3 ADAM9 ... ZYG11B
```

List of the 18 genes found under positive selection ONLY in Young analysis.

```

upsetdata$`tab$Gene.name`[(upsetdata$primates_young==TRUE &
                             upsetdata$primates_dginn_young==FALSE &
                             upsetdata$primates_dginn_full==FALSE)]

## [1] AKAP9    ALG11    ALG5     C19orf52 CENPF    CHMP2A   COLGALT1
## [8] DCTPP1    DDX21    FBN1     FBXL12   JAKMIP1  <NA>     NLRX1
## [15] NUP210    NUP98    PCSK6     PUSL1    ZYG11B
## 332 Levels: AAR2 AASS AATF ABCC1 ACAD9 ACADM ACSL3 ADAM9 ... ZYG11B

```

List of the 1 genes found under positive selection ONLY in DGINN-Young.

```

upsetdata$`tab$Gene.name`[(upsetdata$primates_young==FALSE &
                             upsetdata$primates_dginn_young==TRUE &
                             upsetdata$primates_dginn_full==FALSE)]

## [1] FBN2
## 332 Levels: AAR2 AASS AATF ABCC1 ACAD9 ACADM ACSL3 ADAM9 ... ZYG11B

```

List of the 44 genes found under positive selection ONLY in full-DGINN.

```

upsetdata$`tab$Gene.name`[(upsetdata$primates_young==FALSE &
                             upsetdata$primates_dginn_young==FALSE &
                             upsetdata$primates_dginn_full==TRUE)]

## [1] ADAM9    <NA>     AP2A2    <NA>     <NA>     BZW2     <NA>
## [8] CEP135   CLIP4    DPH5     EIF4E2   EMC1     ER01LB   EXOSC2
## [15] GGH      GLA      GOLGA7   HDAC2    HECTD1   HS6ST2   IDE
## [22] <NA>     LARP1    LARP4B   LARP7    <NA>     MIPOL1   MOV10
## [29] MYCBP2   NAT14    NGLY1    NPC2     NUPL1    PITRM1   PLOD2
## [36] PMPCB    POR      PRKAR2A  PTBP2    RAB14    RAB1A    RAB2A
## [43] RAP1GDS1 RBX1     REEP6    RPL36    SCCPDH   <NA>     <NA>
## [50] TIMM8B   TRIM59   TUBGCP2  <NA>
## 332 Levels: AAR2 AASS AATF ABCC1 ACAD9 ACADM ACSL3 ADAM9 ... ZYG11B

```