

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

Marie Cariou

Janvier 2021

Contents

1	Data	2
2	Comparison of dataset	2
2.1	Data	2
2.2	Mondrian	2
2.3	subsetR	4
3	Which are these genes?	7
3.1	Gene under positive selection in both bats and primates . . .	7
3.2	Gene under positive selection only in primates	8
3.3	Gene under positive selection only in bats	10
3.4	Figure tableau	11

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] 333 161
```

2 Comparison of dataset

2.1 Data

```
tmp<-na.omit(tab[,c("Gene.name", "bats_BUSTED", "bats_BppM1M2", "bats_BppM7M8",
  "bats_codemlM1M2", "bats_codemlM7M8", "dginn.primate_codemlM1M2",
  "dginn.primate_codemlM7M8", "dginn.primate_BppM1M2",
  "dginn.primate_BppM7M8", "dginn.primate_BUSTED")])
dim(tmp)

## [1] 206 11
```

2.2 Mondrian

```
library(Mondrian)

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

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

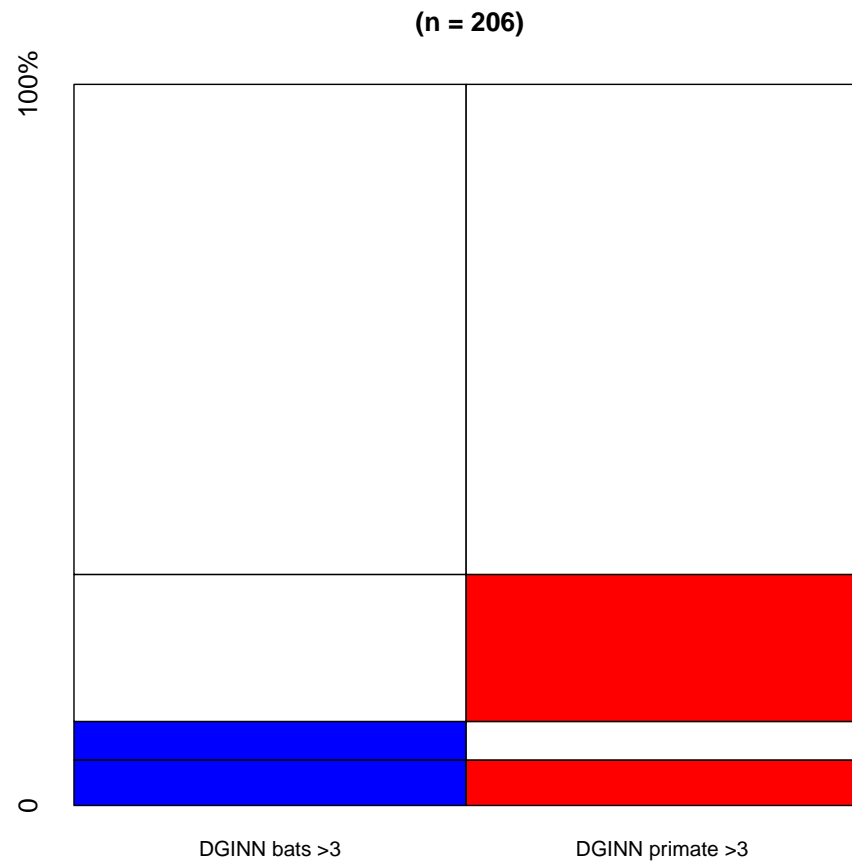
primatetmp<-rowSums(cbind(tmp$"dginn.primate_codemlM1M2"=="Y",
  tmp$"dginn.primate_codemlM7M8"=="Y", tmp$"dginn.primate_BppM1M2"=="Y",
  tmp$"dginn.primate_BppM7M8"=="Y", tmp$"dginn.primate_BUSTED"=="Y"))
```

```

monddata$bats_dginn3<-ifelse(batstmp>=3, 1,0)
monddata$primate_dginn3<-ifelse(primatetmp>=3, 1,0)
monddata$bats_dginn4<-ifelse(batstmp>=4, 1,0)
monddata$primate_dginn4<-ifelse(primatetmp>=4, 1,0)

mondrian(monddata[,2:3], labels=c("DGINN bats >3", "DGINN primate >3"))

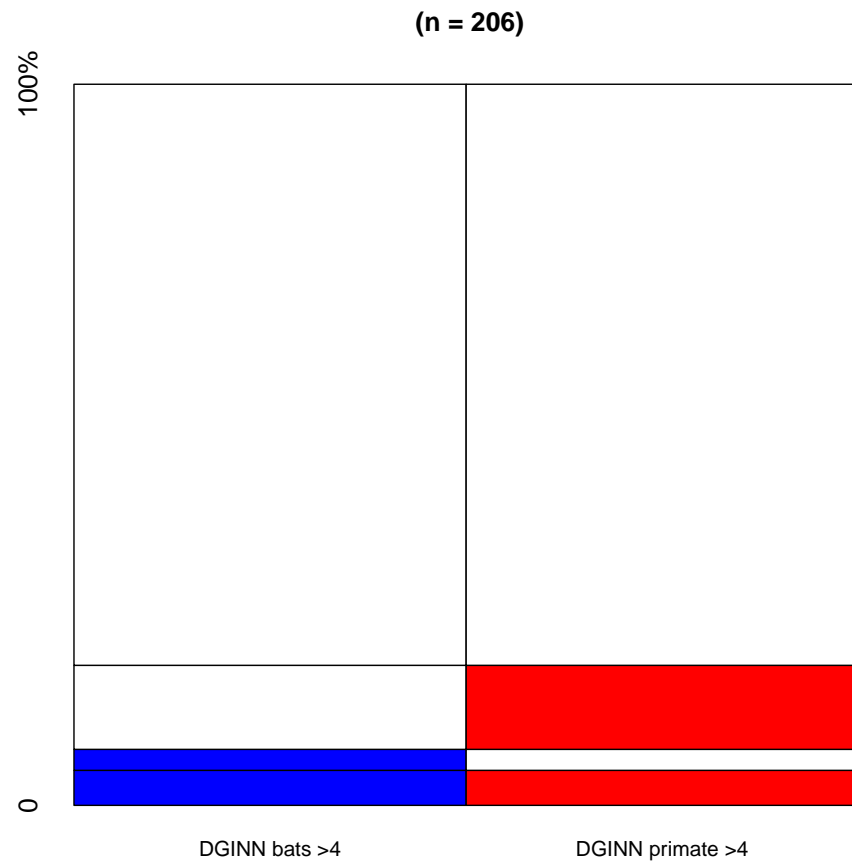
```



```

mondrian(monddata[,4:5], labels=c("DGINN bats >4", "DGINN primate >4"))

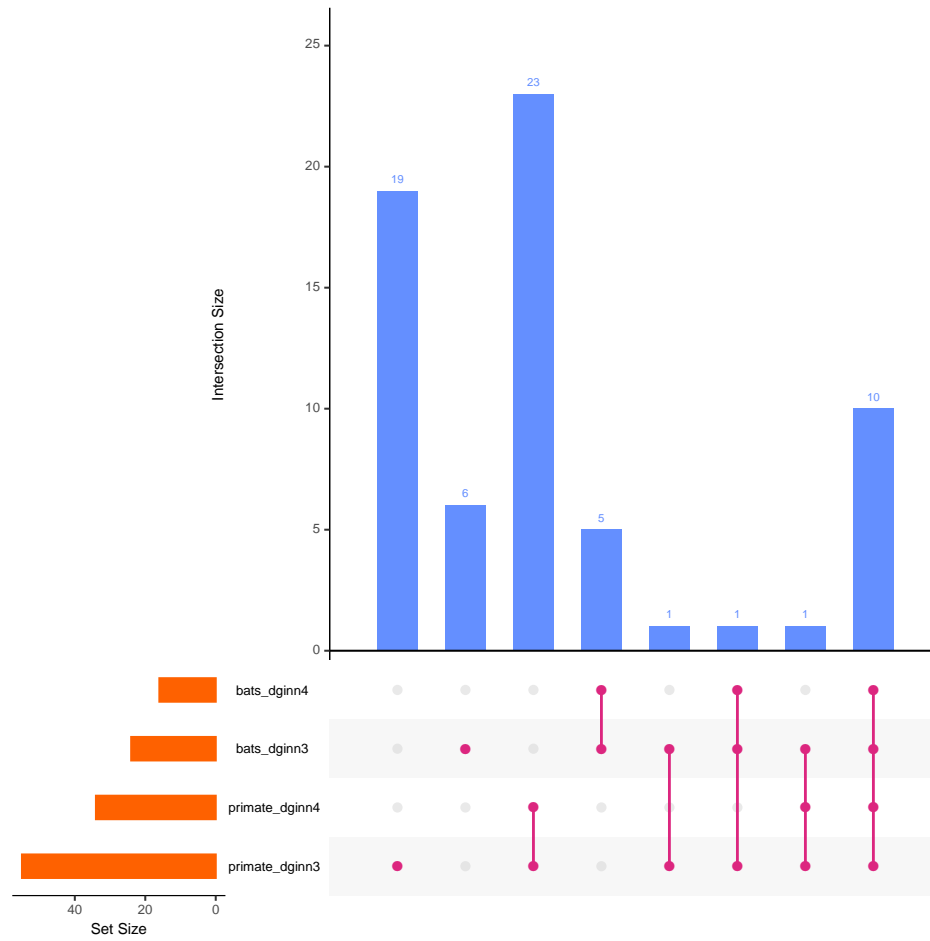
```



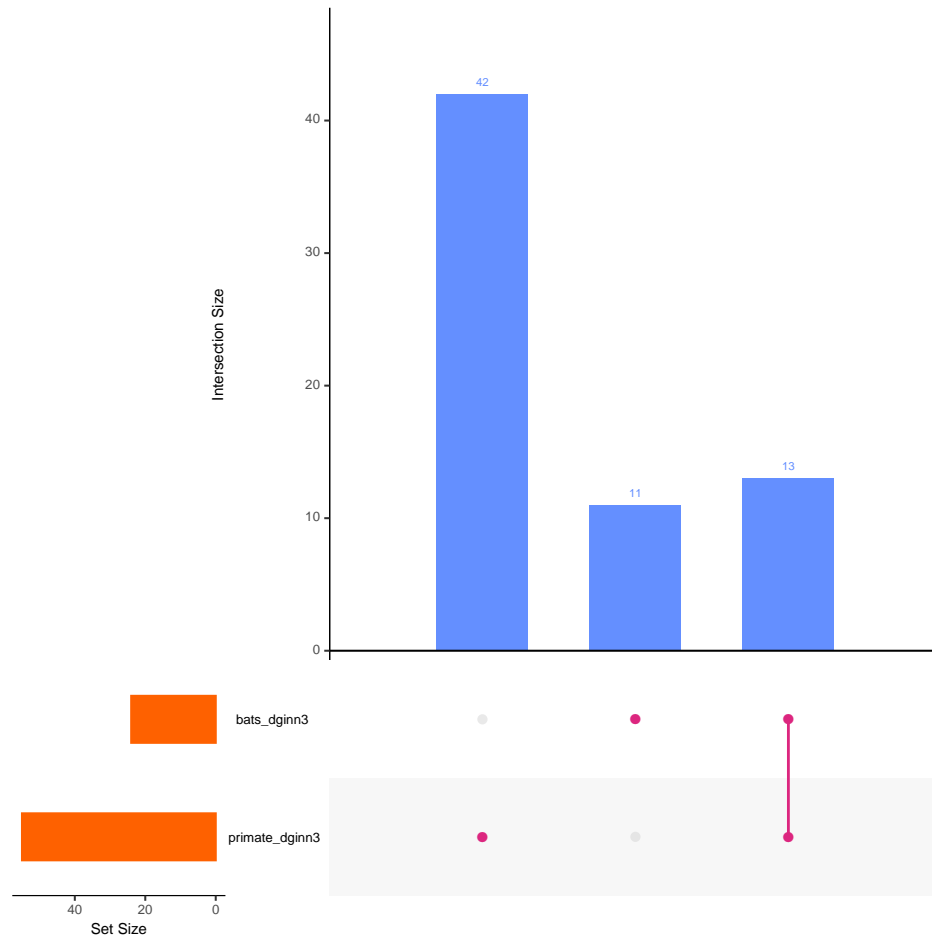
2.3 subsetR

```
library(UpSetR)

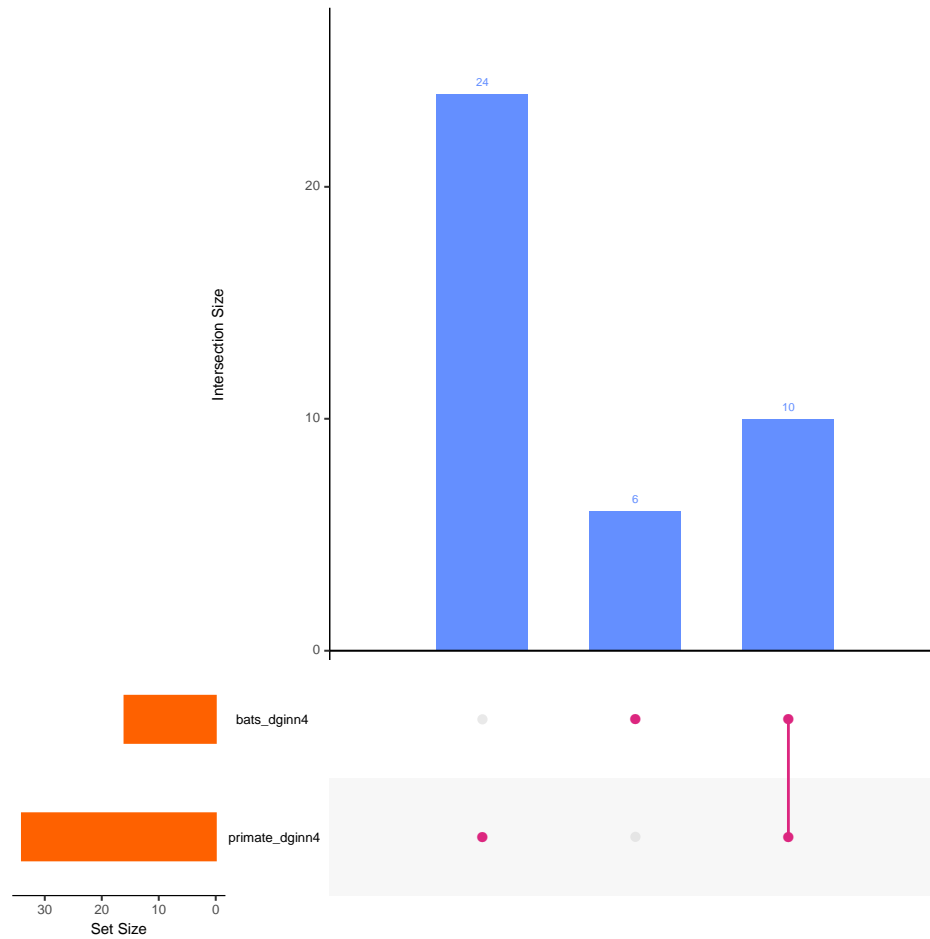
upset(monddata, nsets = 4, matrix.color = "#DC267F",
      main.bar.color = "#648FFF", sets.bar.color = "#FE6100")
```



```
upset(monddata[,1:3], nsets = 2, matrix.color = "#DC267F",
      main.bar.color = "#648FFF", sets.bar.color = "#FE6100")
```



```
upset(monddata[,c(1,4,5)], nsets = 2, matrix.color = "#DC267F",
      main.bar.color = "#648FFF", sets.bar.color = "#FE6100")
```



3 Which are these genes?

3.1 Gene under positive selection in both bats and primates

4 methods:

```
monddata[monddata$bats_dginn4==1 & monddata$primate_dginn4==1,]

##      tmp$Gene.name bats_dginn3 primate_dginn3 bats_dginn4 primate_dginn4
## 4          ACADM          1          1          1          1
## 73           GGH          1          1          1          1
## 78        GOLGA7          1          1          1          1
```

## 89	IDE	1	1	1	1
## 94	ITGB1	1	1	1	1
## 97	LMAN2	1	1	1	1
## 132	POLA1	1	1	1	1
## 164	SLC27A2	1	1	1	1
## 188	TOR1AIP1	1	1	1	1
## 201	VPS39	1	1	1	1

3 methods:

```
monddata[monddata$bats_dginn3==1 & monddata$primate_dginn3==1,]
```

##	tmp\$Gene.name	bats_dginn3	primate_dginn3	bats_dginn4	primate_dginn4
## 4	ACADM	1	1	1	1
## 20	CDK5RAP2	1	1	0	1
## 48	EDEM3	1	1	1	0
## 73	GGH	1	1	1	1
## 78	GOLGA7	1	1	1	1
## 89	IDE	1	1	1	1
## 94	ITGB1	1	1	1	1
## 97	LMAN2	1	1	1	1
## 104	MOV10	1	1	0	0
## 132	POLA1	1	1	1	1
## 164	SLC27A2	1	1	1	1
## 188	TOR1AIP1	1	1	1	1
## 201	VPS39	1	1	1	1

3.2 Gene under positive selection only in primates

4 methods:

```
monddata[monddata$bats_dginn4==0 & monddata$primate_dginn4==1,]
```

##	tmp\$Gene.name	bats_dginn3	primate_dginn3	bats_dginn4	primate_dginn4
## 20	CDK5RAP2	1	1	0	1
## 23	CEP135	0	1	0	1
## 26	CEP68	0	1	0	1
## 32	CLIP4	0	1	0	1
## 45	DPH5	0	1	0	1

## 50	EMC1	0	1	0	1
## 68	FYC01	0	1	0	1
## 74	GIGYF2	0	1	0	1
## 75	GLA	0	1	0	1
## 83	HECTD1	0	1	0	1
## 95	LARP1	0	1	0	1
## 108	MYCBP2	0	1	0	1
## 111	NDUFB9	0	1	0	1
## 138	PRKAR2A	0	1	0	1
## 153	RIPK1	0	1	0	1
## 156	SAAL1	0	1	0	1
## 160	SEPSECS	0	1	0	1
## 162	SIRT5	0	1	0	1
## 163	SLC25A21	0	1	0	1
## 185	TMEM39B	0	1	0	1
## 191	TUBGCP2	0	1	0	1
## 194	UBAP2	0	1	0	1
## 197	UGGT2	0	1	0	1
## 204	ZNF318	0	1	0	1

3 methods:

```
monddata[monddata$bats_dginn3==0 & monddata$primate_dginn3==1,]
```

##	tmp\$Gene.name	bats_dginn3	primate_dginn3	bats_dginn4	primate_dginn4
## 23	CEP135	0	1	0	1
## 26	CEP68	0	1	0	1
## 32	CLIP4	0	1	0	1
## 33	CNTRL	0	1	0	0
## 45	DPH5	0	1	0	1
## 50	EMC1	0	1	0	1
## 56	EXOSC2	0	1	0	0
## 68	FYC01	0	1	0	1
## 74	GIGYF2	0	1	0	1
## 75	GLA	0	1	0	1
## 79	GOLGB1	0	1	0	0
## 80	GORASP1	0	1	0	0
## 83	HECTD1	0	1	0	1
## 95	LARP1	0	1	0	1
## 96	LARP7	0	1	0	0

## 100	MDN1	0	1	0	0
## 107	MRPS5	0	1	0	0
## 108	MYCBP2	0	1	0	1
## 111	NDUFB9	0	1	0	1
## 115	NGLY1	0	1	0	0
## 118	NPC2	0	1	0	0
## 127	PITRM1	0	1	0	0
## 128	PLAT	0	1	0	0
## 134	POR	0	1	0	0
## 138	PRKAR2A	0	1	0	1
## 144	RAB14	0	1	0	0
## 145	RAB2A	0	1	0	0
## 153	RIPK1	0	1	0	1
## 156	SAAL1	0	1	0	1
## 160	SEPSECS	0	1	0	1
## 162	SIRT5	0	1	0	1
## 163	SLC25A21	0	1	0	1
## 172	STOM	0	1	0	0
## 181	TIMM8B	0	1	0	0
## 185	TMEM39B	0	1	0	1
## 189	TRIM59	0	1	0	0
## 190	TRMT1	0	1	0	0
## 191	TUBGCP2	0	1	0	1
## 194	UBAP2	0	1	0	1
## 197	UGGT2	0	1	0	1
## 199	USP54	0	1	0	0
## 204	ZNF318	0	1	0	1

3.3 Gene under positive selection only in bats

4 methods:

```
monddata[monddata$bats_dginn4==1 & monddata$primate_dginn4==0,]
```

##	tmp\$Gene.name	bats_dginn3	primate_dginn3	bats_dginn4	primate_dginn4
## 48	EDEM3	1	1	1	0
## 52	ERGIC1	1	0	1	0
## 91	IMPDH2	1	0	1	0
## 92	INHBE	1	0	1	0

## 166	SLC44A2	1	0	1	0
## 177	TBK1	1	0	1	0

3 methods:

```
monddata[monddata$bats_dginn3==1 & monddata$primate_dginn3==0,]

##      tmp$Gene.name bats_dginn3 primate_dginn3 bats_dginn4 primate_dginn4
## 7          AGPS          1          0          0          0
## 52         ERGIC1          1          0          1          0
## 81        GRIPAP1          1          0          0          0
## 82        GRPEL1          1          0          0          0
## 91        IMPDH2          1          0          1          0
## 92         INHBE          1          0          1          0
## 120       NUP214          1          0          0          0
## 136        PRIM1          1          0          0          0
## 165       SLC30A9          1          0          0          0
## 166       SLC44A2          1          0          1          0
## 177        TBK1          1          0          1          0
```

3.4 Figure tableau

```
tablo<-as.data.frame(tmp$Gene.name)
tablo$nbats<-batstmp
tablo$nprimates<-primatetmp

plot(NULL, xlim=c(-0.5,5.5), ylim=c(-3,5.5), xlab="bats", ylab="primates", main="Gene")

text(x=rep(-0.6, 6), y=0:5, 0:5)
text(y=rep(-0.65, 6), x=0:5, 0:5)
sapply(seq(from=-0.5, to=5.5, by=1), function(x){
  segments(x0=x, x1=x, y0=-0.5, y1=5.5)
})

## [[1]]
## NULL
##
## [[2]]
```

```

## NULL
##
## [[3]]
## NULL
##
## [[4]]
## NULL
##
## [[5]]
## NULL
##
## [[6]]
## NULL
##
## [[7]]
## NULL

sapply(seq(from=-0.5, to=5.5, by=1), function(x){
  segments(x0=-0.5, x1=5.5, y0=x, y1=x)
})

## [[1]]
## NULL
##
## [[2]]
## NULL
##
## [[3]]
## NULL
##
## [[4]]
## NULL
##
## [[5]]
## NULL
##
## [[6]]
## NULL
##
## [[7]]

```

```

## NULL

for (p in 0:5){
  for (b in 0:5){
    tmp<-tablo$tmp$Gene.name`[tablo$nbats==b & tablo$nprimates==p]
    if(length(tmp)>0 & length(tmp)<=8){
      text(b,seq(from=(p-0.4), to=(p+0.4), length.out = length(tmp)), tmp, cex=0.4)
    }else if (length(tmp)>8 & length(tmp)<=16){
      print(c(p, b))
      text((b-0.3),seq(from=(p-0.4), to=(p+0.4), length.out = 8), tmp[1:8], cex=0.4)
      text((b+0.3),seq(from=(p-0.4), to=(p+0.4), length.out = (length(tmp)-8)), tmp[9:16], cex=0.4)
    }else if (length(tmp)>16){
      text(b,p, paste0(length(tmp), " values"))
    }
  }
}

## [1] 0 1
## [1] 1 2
## [1] 2 0
## [1] 2 1

tmp<-tablo$tmp$Gene.name`[tablo$nbats==0 & tablo$nprimates==1]
text(-0.4,-1.2, "p=1/n=0", cex=0.6)
text(seq(from=0.1, to=5.5, length.out = 18),-1.1, tmp[1:18], cex=0.4)
text(seq(from=0.1, to=3, length.out = length(tmp)-18),-1.3, tmp[19:length(tmp)], cex=0.4)

tmp<-tablo$tmp$Gene.name`[tablo$nbats==1 & tablo$nprimates==1]
text(-0.4,-1.7, "p=1/n=1", cex=0.6)
text(seq(from=0.1, to=5.5, length.out = 18),-1.6, tmp[1:18], cex=0.4)
text(seq(from=0.1, to=2, length.out = length(tmp)-18),-1.8, tmp[19:length(tmp)], cex=0.4)

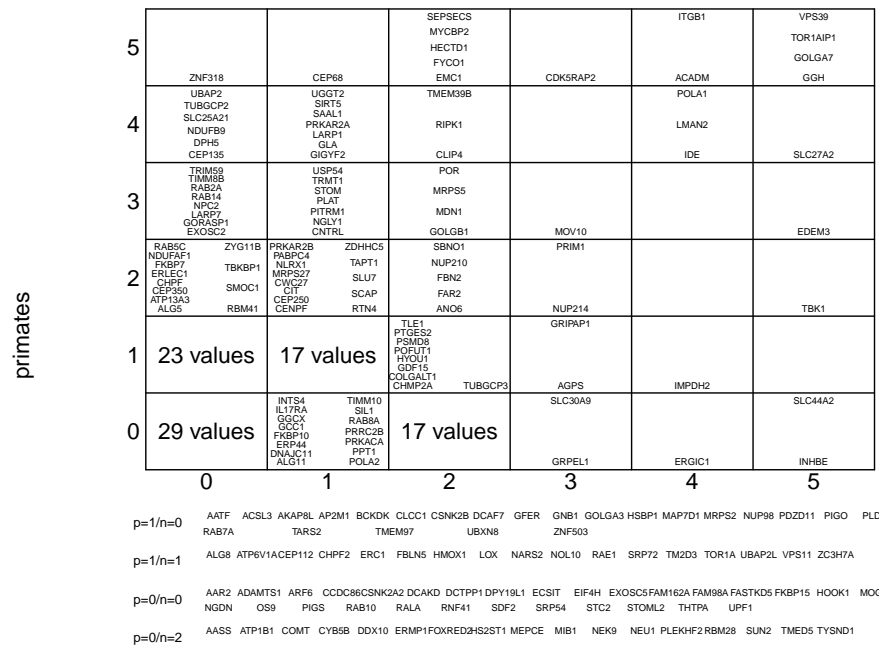
## Error in seq.default(from = 0.1, to = 2, length.out = length(tmp)
- 18): 'length.out' must be a non-negative number

tmp<-tablo$tmp$Gene.name`[tablo$nbats==0 & tablo$nprimates==0]
text(-0.4,-2.2, "p=0/n=0", cex=0.6)
text(seq(from=0.1, to=5.5, length.out = 17),-2.1, tmp[1:17], cex=0.4)
text(seq(from=0.1, to=4.4, length.out = length(tmp)-17),-2.3, tmp[18:length(tmp)], cex=0.4)

```

```
tmp<-tablo$`tmp$Gene.name`[tablo$nbats==2 & tablo$nprimates==0]
text(-0.4,-2.7, "p=0/n=2", cex=0.6)
text(seq(from=0.1, to=5.5, length.out = 18),-2.6, tmp[1:18], cex=0.4)
```

Genes supported by x,y methods in bats and primates



bats

```
text(seq(from=0.1, to=1, length.out = length(tmp)-18),-2.8, tmp[19:length(tmp)], cex=
## Error in seq.default(from = 0.1, to = 1, length.out = length(tmp)
- 18): 'length.out' must be a non-negative number
```