

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

Marie Cariou

March 2021

Contents

1	Data	2
2	MAIC	2
2.1	Boxplot	2
2.2	Dotchart	4
3	Pan Corona	8

1 Data

output from covid_comp_dataset.

```
tablo<-read.table("primatesVbats.csv",  
                  h=T, sep=",")
```

Output MAIC formatted by Léa. This table includes the DGINN "score".

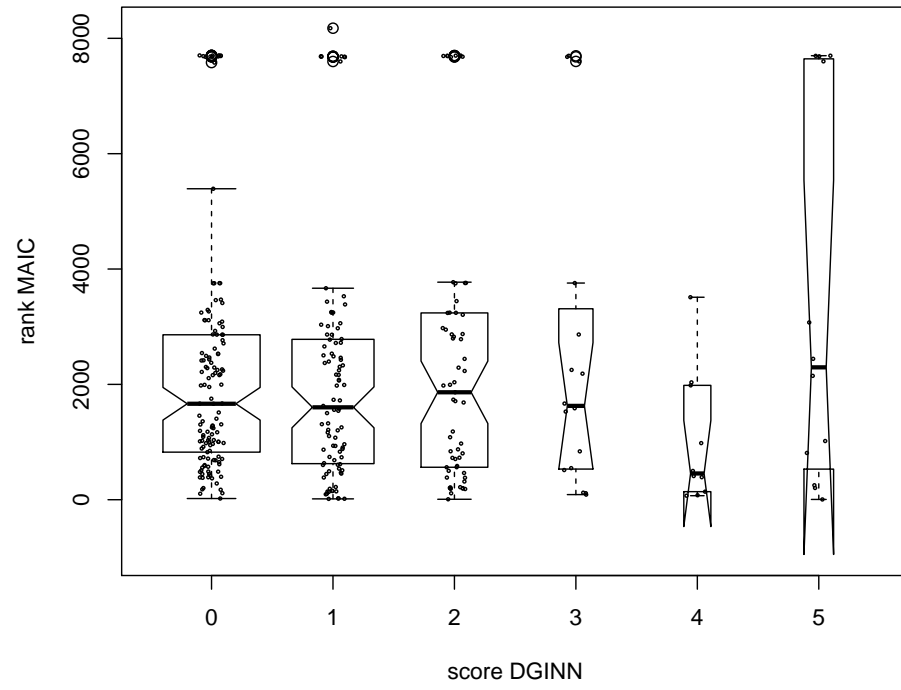
```
home<-"/home/adminmarie/Documents/"  
workdir<-paste0(home, "CIRI_BIBS_projects/2020_05_Etienne_covid/")  
maic<-read.table(paste0(workdir, "data/covid_comp_maic.txt"),  
                 h=T)
```

2 MAIC

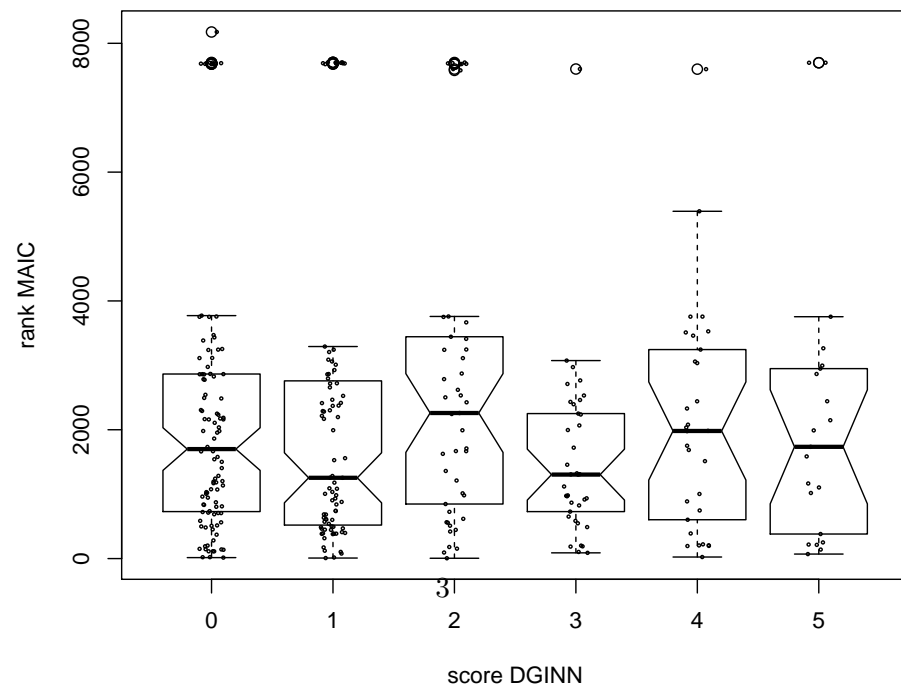
2.1 Boxplot

```
par(mfrow=c(2,1))  
boxplot(maic$rank~maic$nbats, notch=TRUE, varwidth=TRUE, xlab="score DGINN", ylab="r  
  
## Warning in bxp(list(stats = structure(c(21, 825, 1664, 2860,  
5392, 15, 625.5, : some notches went outside hinges ('box')): maybe  
set notch=FALSE  
  
stripchart(maic$rank~maic$nbats, method="jitter", vertical=TRUE, pch=1, cex=0.3, add=  
  
boxplot(maic$rank~maic$nprimates, notch=TRUE, xlab="score DGINN", ylab="rank MAIC",  
stripchart(maic$rank~maic$nprimates, method="jitter", vertical=TRUE, pch=1, cex=0.3,
```

Bats



Primates

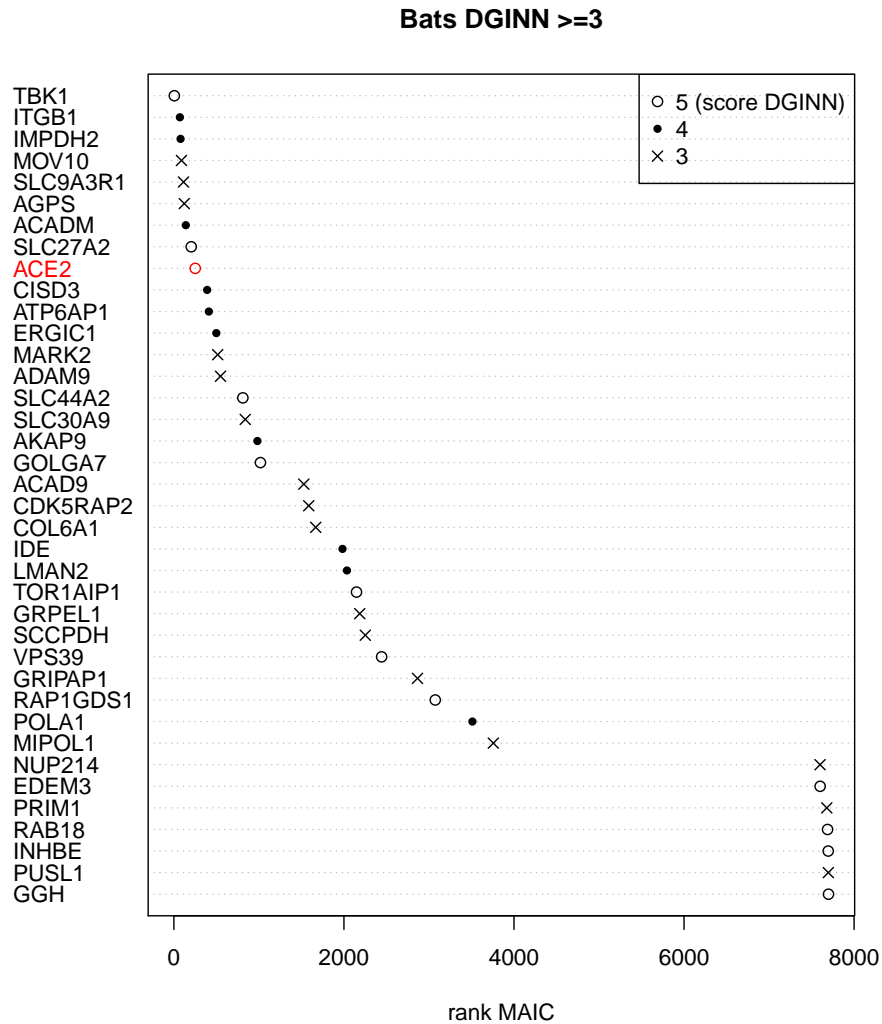


2.2 Dotchart

```
tmp<-maic[maic$nbats>=3, c("gene", "rank", "nbats")]
tmp<-tmp[order(tmp$rank, decreasing = TRUE),]
tmp$col<-"black"
tmp$col[tmp$gene=="ACE2"]<-"red"
tmp$col[tmp$gene=="TPRSS2"]<-"red"

tmp$pch[tmp$nbats==5]<-1
tmp$pch[tmp$nbats==4]<-20
tmp$pch[tmp$nbats==3]<-4

dotchart(tmp$rank, main="Bats DGINN >=3", xlab="rank MAIC", labels=tmp$gene, pch=tmp$pch)
legend("topright", c("5 (score DGINN)", "4", "3"), pch=c(1,20,4))
```



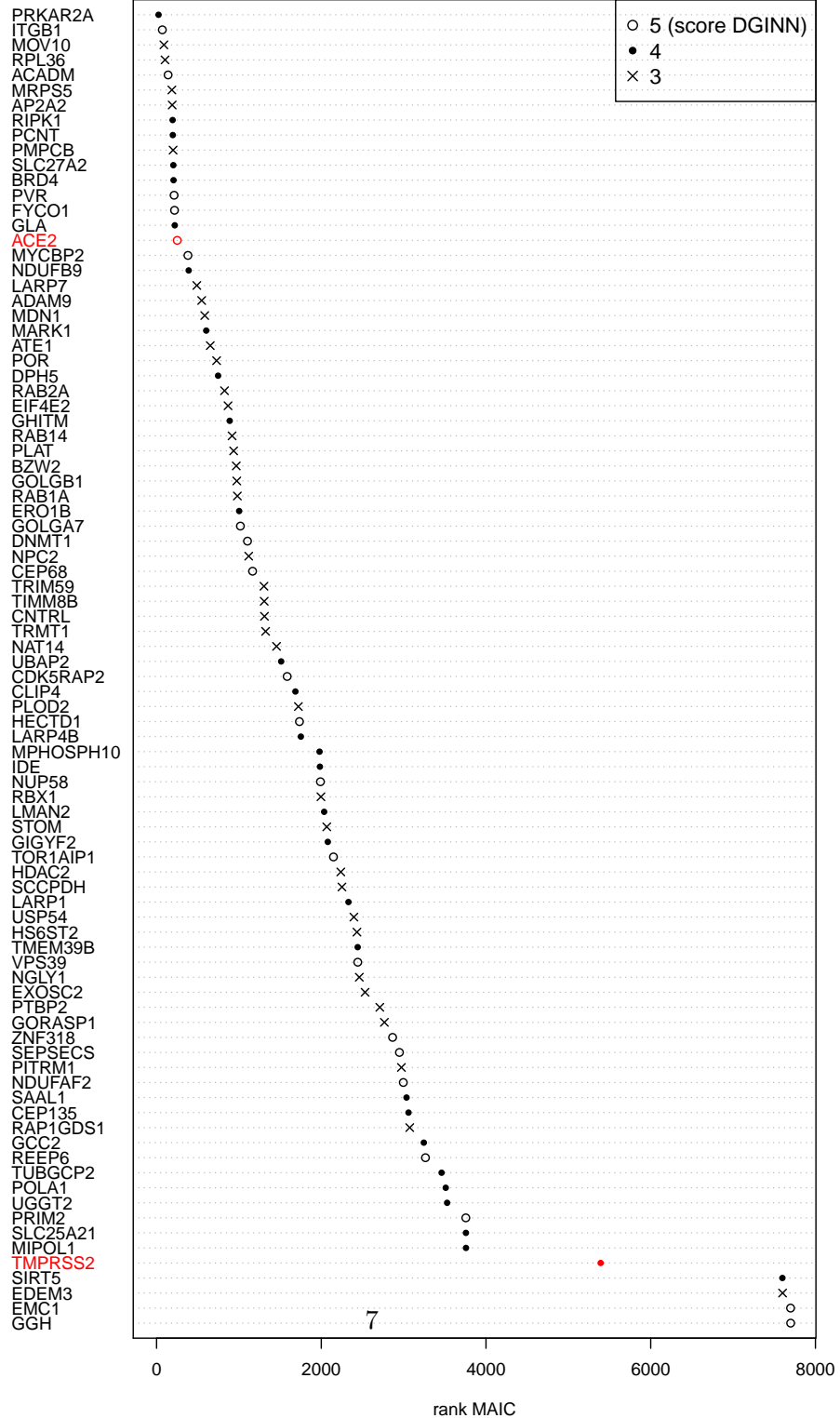
```
tmp<-maic[maic$nprimates>=3, c("gene", "rank", "nprimates")]
tmp<-tmp[order(tmp$rank, decreasing = TRUE),]

tmp$pch[tmp$nprimates==5]<-1
tmp$pch[tmp$nprimates==4]<-20
tmp$pch[tmp$nprimates==3]<-4
tmp$col<-"black"
tmp$col[tmp$gene=="ACE2"]<-"red"
```

```
tmp$col[tmp$gene=="TPRSS2"]<-"red"

dotchart(tmp$rank, main="Primates DGINN >=3", xlab="rank MAIC", labels=tmp$gene, pch=
legend("topright", c("5 (score DGINN)", "4", "3"), pch=c(1,20,4))
```

Primates DGINN >=3



3 Pan Corona

```

pancorona<-read.table(paste0(workdir, "data/pancorona_S5.csv"),
                      h=T, fill = TRUE, sep="\t")
names(pancorona)<-c("tmp.Gene.name", names(pancorona)[-1])

# Genes en commun
pancorona$tmp.Gene.name[pancorona$tmp.Gene.name %in% tablo$tmp.Gene.name]

## [1] TBK1      MARK3      GIGYF2     MARK2      G3BP1      LARP1      ACE2       PABPC1
## [9] TMPRSS2    AP3B1      CLCC1      CSDE1      HECTD1     MARK1      MEPCE      PDE4DIP
## [17] POR        PRKAR2B    RAB5C      RTN4       SRP54      UBAP2      UBAP2L     UBXN8
## [25] SPART      BZW2       EIF4E2     SMOC1      STOML2     DDX21      FAM98A     G3BP2
## [33] MOV10      PABPC4     UPF1
## 105 Levels: ACE2 ANPEP AP3B1 ATXN2L BTF3 BZW2 CKAP5 CLCC1 ... YTHDF2

# Uniquement dans le tableau pancorona
sort(pancorona$tmp.Gene.name[(pancorona$tmp.Gene.name %in% tablo$tmp.Gene.name)==FALSE])

## [1] ANPEP      ATXN2L     BTF3       CKAP5      CTSB       CTSL
## [7] CYB5R3     DDX1       DDX5       DDX58      DHX9       DNM1L
## [13] EEF1A1     EIF2A      EIF3F      EIF4B      EZR        FLNA
## [19] FURIN      FUS        GSK3A      GSK3B      HDLBP      HNRNPA1
## [25] HNRNPD     HNRNPF     HNRNPU     IFIH1      IGF2BP1    IKBKB
## [31] IKBKE      IRF3       ISG15      KPNA3      KPNB1      MYH9
## [37] NCL        POLD1      POLR2B     PRKRA      RBM14      RCHY1
## [43] RPL13A     RPL26      RPS13      RPS17      RPS19      RPS9
## [49] SDCBP      SERBP1     SGTA       SLC1A5     SNAP47     SSB
## [55] STING1     SYNCRIP    TANC1      TBCB       TMPRSS11D  TRAF3
## [61] TUBA4A     TUBB2A     TUBB4A     TUBB6      USP10      VPS36
## [67] XRCC5      XRCC6      YBX1       YTHDF2
## 105 Levels: ACE2 ANPEP AP3B1 ATXN2L BTF3 BZW2 CKAP5 CLCC1 ... YTHDF2

## Uniquement dans tableau
sort(tablo$tmp.Gene.name[(tablo$tmp.Gene.name %in% pancorona$tmp.Gene.name)==FALSE])

## [1] AAR2      AASS       AATF       ABCC1      ACAD9      ACADM
## [7] ACSL3     ADAM9      ADAMTS1    AGPS       AKAP8      AKAP8L
## [13] AKAP9     ALG11      ALG5       ALG8       ANO6       AP2A2
## [19] AP2M1     ARF6       ATE1       ATP13A3    ATP1B1     ATP6AP1

```


##	[25]	ATP6V1A	BAG5	BCKDK	BRD2	BRD4	CCDC86
##	[31]	CDK5RAP2	CENPF	CEP112	CEP135	CEP250	CEP350
##	[37]	CEP68	CHMP2A	CHPF	CHPF2	CISD3	CIT
##	[43]	CLIP4	CNTRL	COL6A1	COLGALT1	COMT	COQ8B
##	[49]	CRTC3	CSNK2A2	CSNK2B	CUL2	CWC27	CYB5B
##	[55]	DCAF7	DCAKD	DCTPP1	DDX10	DNAJC11	DNAJC19
##	[61]	DNMT1	DPH5	DPY19L1	ECSIT	EDEM3	EIF4H
##	[67]	ELOC	EMC1	ERC1	ERGIC1	ERLEC1	ERMP1
##	[73]	ERO1B	ERP44	ETFA	EXOSC2	EXOSC3	EXOSC5
##	[79]	EXOSC8	F2RL1	FAM162A	FAM8A1	FAR2	FASTKD5
##	[85]	FBLN5	FBN1	FBN2	FBXL12	FKBP10	FKBP15
##	[91]	FKBP7	FOXRED2	FYCO1	GCC1	GCC2	GDF15
##	[97]	GFER	GGCX	GGH	GHITM	GLA	GNB1
##	[103]	GNG5	GOLGA2	GOLGA3	GOLGA7	GOLGB1	GORASP1
##	[109]	GPAA1	GPX1	GRIPAP1	GRPEL1	GTF2F2	HDAC2
##	[115]	HEATR3	HMOX1	HOOK1	HS2ST1	HS6ST2	HSBP1
##	[121]	HYOU1	IDE	IL17RA	IMPDH2	INHBE	INTS4
##	[127]	ITGB1	JAKMIP1	KDELC1	KDELC2	LARP4B	LARP7
##	[133]	LMAN2	LOX	MAP7D1	MARC1	MAT2B	MDN1
##	[139]	MIB1	MIPOL1	MOGS	MPHOSPH10	MRPS2	MRPS25
##	[145]	MRPS27	MRPS5	MTCH1	MYCBP2	NARS2	NAT14
##	[151]	NDFIP2	NDUFAB1	NDUFAB2	NDUFB9	NEK9	NEU1
##	[157]	NGDN	NGLY1	NIN	NINL	NLRX1	NOL10
##	[163]	NPC2	NPTX1	NSD2	NUP210	NUP214	NUP54
##	[169]	NUP58	NUP62	NUP88	NUP98	NUTF2	OS9
##	[175]	PCNT	PCSK5	PCSK6	PDZD11	PIGO	PIGS
##	[181]	PITRM1	PKP2	PLAT	PLD3	PLEKHA5	PLEKHF2
##	[187]	PLOD2	PMPCA	PMPCB	POFUT1	POLA1	POLA2
##	[193]	PPIL3	PPT1	PRIM1	PRIM2	PRKACA	PRKAR2A
##	[199]	PRRC2B	PSMD8	PTBP2	PTGES2	PUSL1	PVR
##	[205]	QSOX2	RAB10	RAB14	RAB18	RAB1A	RAB2A
##	[211]	RAB7A	RAB8A	RAE1	RALA	RAP1GDS1	RBM28
##	[217]	RBM41	RBX1	RDX	REEP5	REEP6	RETREG3
##	[223]	RHOA	RIPK1	RNF41	RPL36	RRP9	SAAL1
##	[229]	SBN01	SCAP	SCARB1	SCCPDH	SDF2	SELENOS
##	[235]	SEPSECS	SIL1	SIRT5	SLC25A21	SLC27A2	SLC30A6
##	[241]	SLC30A7	SLC30A9	SLC44A2	SLC9A3R1	SLU7	SNIP1
##	[247]	SRP19	SRP72	STC2	STOM	SUN2	TAPT1
##	[253]	TARS2	TBCA	TBKB1	TCF12	THTPA	TIMM10

```
## [259] TIMM10B    TIMM29    TIMM8B    TIMM9     TLE1      TLE3
## [265] TM2D3      TMED5     TMEM39B   TMEM97    TOMM70    TOR1A
## [271] TOR1AIP1   TRIM59    TRMT1     TUBGCP2   TUBGCP3   TYSND1
## [277] UGGT2      USP54     VPS11     VPS39     WASHC4    WFS1
## [283] YIF1A      ZC3H18    ZC3H7A    ZDHHC5    ZNF318    ZNF503
## [289] ZYG11B
## 324 Levels: AAR2 AASS AATF ABCC1 ACAD9 ACADM ACE2 ACSL3 ... ZYG11B
```

```
pancorona<-pancorona[,c("tmp.Gene.name", "TOTAL")]
```

```
pandginn<-na.omit(merge(pancorona, tablo, by="tmp.Gene.name", all.x=TRUE))
```

```
pandginn<-pandginn[order(pandginn$nprimates),]
```

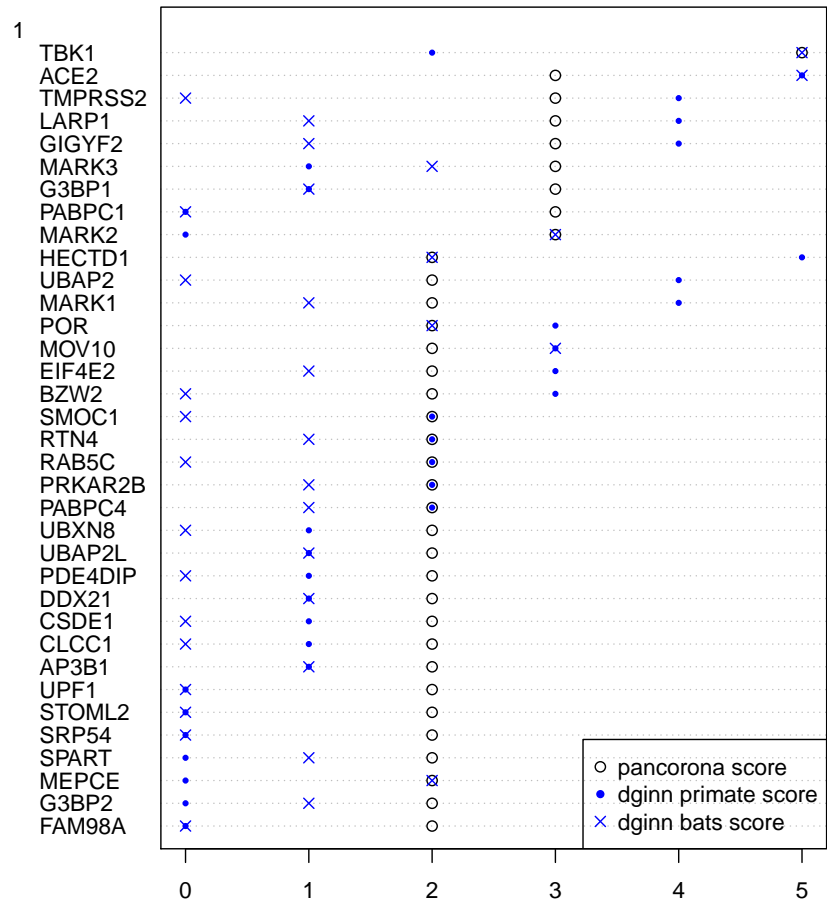
```
pandginn<-pandginn[order(pandginn$TOTAL),]
```

```
dotchart(as.matrix(pandginn[,2]), labels = pandginn$tmp.Gene.name, xlim=c(0,5))
```

```
points(pandginn[,4], 1:nrow(pandginn), col="blue", pch=20, cex=0.7)
```

```
points(pandginn[,3], 1:nrow(pandginn), col="blue", pch=4)
```

```
legend("bottomright", c("pancorona score", "dginn primate score", "dginn bats score")
```



A-t-on un enrichissement en Pan-corona dans nos gènes sous PS?

```
pandginnall<-merge(pancorona, tablo, by="tmp.Gene.name", all.x=FALSE,all.y=TRUE)

dim(pandginnall)

## [1] 324 4

# test indépendance: under PS / in the pancorona list
table(is.na(pandginnall$TOTAL)==FALSE)
```

```
##
## FALSE TRUE
## 289 35

table(pandginnall$nbats>=3)

##
## FALSE TRUE
## 286 38

chi<-table(is.na(pandginnall$TOTAL)==FALSE,pandginnall$nbats>=3)
chi

##
## FALSE TRUE
## FALSE 255 34
## TRUE 31 4

chisq.test(chi)

## Warning in chisq.test(chi): Chi-squared approximation may be
incorrect

##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data: chi
## X-squared = 7.6869e-31, df = 1, p-value = 1

table(is.na(pandginnall$TOTAL)==FALSE)

##
## FALSE TRUE
## 289 35

table(pandginnall$nprimates>=3)

##
## FALSE TRUE
## 236 88
```

```

chi<-table(is.na(pandginnall$TOTAL)==FALSE,pandginnall$nprimates>=3)
chi

##
##      FALSE TRUE
## FALSE   212   77
##  TRUE    24   11

chisq.test(chi)

##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  chi
## X-squared = 0.15992, df = 1, p-value = 0.6892

```

No enrichment in PanCORONA in our genes under PS.