

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

Marie Cariou

October 2020

## Contents

<b>1</b>	<b>Files manipulations</b>	<b>2</b>
1.1	Read Janet Young's table . . . . .	2
1.2	Read DGINN Young table . . . . .	2
1.3	Joining Young and DGINN Young table . . . . .	2
1.4	Read DGINN Table . . . . .	3
1.5	Join Table and DGINN table . . . . .	5
1.6	Write new table . . . . .	5
<b>2</b>	<b>Comparisons Primates</b>	<b>5</b>
2.1	DGINN results on Janet Young's alignments (DGINN-Young-primate) VS Janet Young's results . . . . .	5
2.2	DGINN results on Janet Young's alignments (DGINN-Young-primate) VS DGINN-full's results . . . . .	6
2.3	Janet Young's results (Young-primate) VS DGINN-full's results	8
<b>3</b>	<b>Overlap</b>	<b>9</b>
3.1	Mondrian . . . . .	9
3.2	subsetR . . . . .	13
<b>4</b>	<b>Gene List</b>	<b>15</b>
<b>5</b>	<b>Shiny like</b>	<b>16</b>

# 1 Files manipulations

## 1.1 Read Janet Young's table

```
workdir<-"/home/adminmarie/Documents/CIRI_BIBS_projects/2020_05_Etienne_covid/"

tab<-read.delim(paste0(workdir,
  "data/COVID_PAMLresults_332hits_plusBatScreens_2020_Apr14.csv"),
  fill=T, h=T, dec=",")
dim(tab)

## [1] 332 84

#names(tab)
```

## 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" "PSS"
```

## 1.3 Joining Young and DGINN Young table

*I hide some code corresponding to verifications of gene names coherence between tables*

```

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")

```

## 1.4 Read DGINN Table

```

dginnT<-read.delim(paste0(workdir,
"/data/DGINN_202005281649summary_cleaned.csv"),
                  fill=T, h=T, sep=",")

dim(dginnT)

```

```
## [1] 412 27

names(dginnT)

## [1] "File" "Name" "Gene" "GeneSize"
## [7] "omegaM0codeml" "BUSTED" "BUSTED.p.value" "MEME.NbSites"
## [13] "BppM1M2.p.value" "BppM1M2.NbSites" "BppM1M2.PSS" "BppM7M8"
## [19] "BppM7M8.PSS" "codemlM1M2" "codemlM1M2.p.value" "codemlM1M2.NbSites"
## [25] "codemlM7M8.p.value" "codemlM7M8.NbSites" "codemlM7M8.PSS"

# Number of genes in dginn-primate output not present in the original table
dginnT[(dginnT$Gene %in% tab$Gene.name)==F, "Gene"]

## [1] ACE2 ADAM9[0-3120] ADAM9[3119-3927] ATP5MGL
## [7] CEP135[3263-3678] CEP43 COQ8B COQ8A
## [13] CSNK2B[608-2568] CYB5R1 DDX21[0-717] DDX21[716-2538]
## [19] DPH5[0-702] DPH5[701-1326] DPY19L2 ELOC
## [25] EXOSC3[1445-1980] FBN3 GNB4 GNB2
## [31] GOLGA7[311-549] GPX1[0-1218] GPX1[1217-2946] HDAC1
## [37] ITGB1[0-2328] ITGB1[2327-2844] LMAN2L MRPS5[0-1569]
## [43] MGRN1 NDFIP2[0-768] NDFIP2[767-1314] NDUFAF2[0-258]
## [49] NUP58 NUP58[0-1824] NUP58[1823-2367] PABPC3
## [55] PABPC5 PCSK5 PRIM2[0-1071] PRIM2[1070-1902]
## [61] PTGES2[0-1587] PTGES2[1586-2202] RAB8B RAB13
## [67] RAB2B RAB5A RAB5B RAB15
## [73] EZR[0-1458] EZR[1457-3771] MSN RETREG3
## [79] SLC44A2[0-2577] SLC44A2[2576-3657] SPART SRP72[0-2604]
## [85] STOM[1046-1800] STOML3 TIMM29 TLE4
## [91] TLE2[1301-3987] TMPRSS2 TOMM70 TOR1B
## [97] WFS1[2345-3216] YIF1B
## 411 Levels: AAR2 AASS AATF ABCC1 ACAD9 ACADM ACE2 ACSL3 ADAM9 ADAM9[0-3120] ADAM9[3119-3927] APOA1 APOA1[0-243] APOA1[243-486] APOA2 APOA2[0-243] APOA2[243-486] APOA3 APOA3[0-243] APOA3[243-486] APOA4 APOA4[0-243] APOA4[243-486] APOA5 APOA5[0-243] APOA5[243-486] APOA6 APOA6[0-243] APOA6[243-486] APOA7 APOA7[0-243] APOA7[243-486] APOA8 APOA8[0-243] APOA8[243-486] APOA9 APOA9[0-243] APOA9[243-486] APOA10 APOA10[0-243] APOA10[243-486] APOA11 APOA11[0-243] APOA11[243-486] APOA12 APOA12[0-243] APOA12[243-486] APOA13 APOA13[0-243] APOA13[243-486] APOA14 APOA14[0-243] APOA14[243-486] APOA15 APOA15[0-243] APOA15[243-486] APOA16 APOA16[0-243] APOA16[243-486] APOA17 APOA17[0-243] APOA17[243-486] APOA18 APOA18[0-243] APOA18[243-486] APOA19 APOA19[0-243] APOA19[243-486] APOA20 APOA20[0-243] APOA20[243-486] APOA21 APOA21[0-243] APOA21[243-486] APOA22 APOA22[0-243] APOA22[243-486] APOA23 APOA23[0-243] APOA23[243-486] APOA24 APOA24[0-243] APOA24[243-486] APOA25 APOA25[0-243] APOA25[243-486] APOA26 APOA26[0-243] APOA26[243-486] APOA27 APOA27[0-243] APOA27[243-486] APOA28 APOA28[0-243] APOA28[243-486] APOA29 APOA29[0-243] APOA29[243-486] APOA30 APOA30[0-243] APOA30[243-486] APOA31 APOA31[0-243] APOA31[243-486] APOA32 APOA32[0-243] APOA32[243-486] APOA33 APOA33[0-243] APOA33[243-486] APOA34 APOA34[0-243] APOA34[243-486] APOA35 APOA35[0-243] APOA35[243-486] APOA36 APOA36[0-243] APOA36[243-486] APOA37 APOA37[0-243] APOA37[243-486] APOA38 APOA38[0-243] APOA38[243-486] APOA39 APOA39[0-243] APOA39[243-486] APOA40 APOA40[0-243] APOA40[243-486] APOA41 APOA41[0-243] APOA41[243-486] APOA42 APOA42[0-243] APOA42[243-486] APOA43 APOA43[0-243] APOA43[243-486] APOA44 APOA44[0-243] APOA44[243-486] APOA45 APOA45[0-243] APOA45[243-486] APOA46 APOA46[0-243] APOA46[243-486] APOA47 APOA47[0-243] APOA47[243-486] APOA48 APOA48[0-243] APOA48[243-486] APOA49 APOA49[0-243] APOA49[243-486] APOA50 APOA50[0-243] APOA50[243-486] APOA51 APOA51[0-243] APOA51[243-486] APOA52 APOA52[0-243] APOA52[243-486] APOA53 APOA53[0-243] APOA53[243-486] APOA54 APOA54[0-243] APOA54[243-486] APOA55 APOA55[0-243] APOA55[243-486] APOA56 APOA56[0-243] APOA56[243-486] APOA57 APOA57[0-243] APOA57[243-486] APOA58 APOA58[0-243] APOA58[243-486] APOA59 APOA59[0-243] APOA59[243-486] APOA60 APOA60[0-243] APOA60[243-486] APOA61 APOA61[0-243] APOA61[243-486] APOA62 APOA62[0-243] APOA62[243-486] APOA63 APOA63[0-243] APOA63[243-486] APOA64 APOA64[0-243] APOA64[243-486] APOA65 APOA65[0-243] APOA65[243-486] APOA66 APOA66[0-243] APOA66[243-486] APOA67 APOA67[0-243] APOA67[243-486] APOA68 APOA68[0-243] APOA68[243-486] APOA69 APOA69[0-243] APOA69[243-486] APOA70 APOA70[0-243] APOA70[243-486] APOA71 APOA71[0-243] APOA71[243-486] APOA72 APOA72[0-243] APOA72[243-486] APOA73 APOA73[0-243] APOA73[243-486] APOA74 APOA74[0-243] APOA74[243-486] APOA75 APOA75[0-243] APOA75[243-486] APOA76 APOA76[0-243] APOA76[243-486] APOA77 APOA77[0-243] APOA77[243-486] APOA78 APOA78[0-243] APOA78[243-486] APOA79 APOA79[0-243] APOA79[243-486] APOA80 APOA80[0-243] APOA80[243-486] APOA81 APOA81[0-243] APOA81[243-486] APOA82 APOA82[0-243] APOA82[243-486] APOA83 APOA83[0-243] APOA83[243-486] APOA84 APOA84[0-243] APOA84[243-486] APOA85 APOA85[0-243] APOA85[243-486] APOA86 APOA86[0-243] APOA86[243-486] APOA87 APOA87[0-243] APOA87[243-486] APOA88 APOA88[0-243] APOA88[243-486] APOA89 APOA89[0-243] APOA89[243-486] APOA90 APOA90[0-243] APOA90[243-486] APOA91 APOA91[0-243] APOA91[243-486] APOA92 APOA92[0-243] APOA92[243-486] APOA93 APOA93[0-243] APOA93[243-486] APOA94 APOA94[0-243] APOA94[243-486] APOA95 APOA95[0-243] APOA95[243-486] APOA96 APOA96[0-243] APOA96[243-486] APOA97 APOA97[0-243] APOA97[243-486] APOA98 APOA98[0-243] APOA98[243-486] APOA99 APOA99[0-243] APOA99[243-486] APOA100 APOA100[0-243] APOA100[243-486] APOA101 APOA101[0-243] APOA101[243-486] APOA102 APOA102[0-243] APOA102[243-486] APOA103 APOA103[0-243] APOA103[243-486] APOA104 APOA104[0-243] APOA104[243-486] APOA105 APOA105[0-243] APOA105[243-486] APOA106 APOA106[0-243] APOA106[243-486] APOA107 APOA107[0-243] APOA107[243-486] APOA108 APOA108[0-243] APOA108[243-486] APOA109 APOA109[0-243] APOA109[243-486] APOA110 APOA110[0-243] APOA110[243-486] APOA111 APOA111[0-243] APOA111[243-486] APOA112 APOA112[0-243] APOA112[243-486] APOA113 APOA113[0-243] APOA113[243-486] APOA114 APOA114[0-243] APOA114[243-486] APOA115 APOA115[0-243] APOA115[243-486] APOA116 APOA116[0-243] APOA116[243-486] APOA117 APOA117[0-243] APOA117[243-486] APOA118 APOA118[0-243] APOA118[243-486] APOA119 APOA119[0-243] APOA119[243-486] APOA120 APOA120[0-243] APOA120[243-486] APOA121 APOA121[0-243] APOA121[243-486] APOA122 APOA122[0-243] APOA122[243-486] APOA123 APOA123[0-243] APOA123[243-486] APOA124 APOA124[0-243] APOA124[243-486] APOA125 APOA125[0-243] APOA125[243-486] APOA126 APOA126[0-243] APOA126[243-486] APOA127 APOA127[0-243] APOA127[243-486] APOA128 APOA128[0-243] APOA128[243-486] APOA129 APOA129[0-243] APOA129[243-486] APOA130 APOA130[0-243] APOA130[243-486] APOA131 APOA131[0-243] APOA131[243-486] APOA132 APOA132[0-243] APOA132[243-486] APOA133 APOA133[0-243] APOA133[243-486] APOA134 APOA134[0-243] APOA134[243-486] APOA135 APOA135[0-243] APOA135[243-486] APOA136 APOA136[0-243] APOA136[243-486] APOA137 APOA137[0-243] APOA137[243-486] APOA138 APOA138[0-243] APOA138[243-486] APOA139 APOA139[0-243] APOA139[243-486] APOA140 APOA140[0-243] APOA140[243-486] APOA141 APOA141[0-243] APOA141[243-486] APOA142 APOA142[0-243] APOA142[243-486] APOA143 APOA143[0-243] APOA143[243-486] APOA144 APOA144[0-243] APOA144[243-486] APOA145 APOA145[0-243] APOA145[243-486] APOA146 APOA146[0-243] APOA146[243-486] APOA147 APOA147[0-243] APOA147[243-486] APOA148 APOA148[0-243] APOA148[243-486] APOA149 APOA149[0-243] APOA149[243-486] APOA150 APOA150[0-243] APOA150[243-486] APOA151 APOA151[0-243] APOA151[243-486] APOA152 APOA152[0-243] APOA152[243-486] APOA153 APOA153[0-243] APOA153[243-486] APOA154 APOA154[0-243] APOA154[243-486] APOA155 APOA155[0-243] APOA155[243-486] APOA156 APOA156[0-243] APOA156[243-486] APOA157 APOA157[0-243] APOA157[243-486] APOA158 APOA158[0-243] APOA158[243-486] APOA159 APOA159[0-243] APOA159[243-486] APOA160 APOA160[0-243] APOA160[243-486] APOA161 APOA161[0-243] APOA161[243-486] APOA162 APOA162[0-243] APOA162[243-486] APOA163 APOA163[0-243] APOA163[243-486] APOA164 APOA164[0-243] APOA164[243-486] APOA165 APOA165[0-243] APOA165[243-486] APOA166 APOA166[0-243] APOA166[243-486] APOA167 APOA167[0-243] APOA167[243-486] APOA168 APOA168[0-243] APOA168[243-486] APOA169 APOA169[0-243] APOA169[243-486] APOA170 APOA170[0-243] APOA170[243-486] APOA171 APOA171[0-243] APOA171[243-486] APOA172 APOA172[0-243] APOA172[243-486] APOA173 APOA173[0-243] APOA173[243-486] APOA174 APOA174[0-243] APOA174[243-486] APOA175 APOA175[0-243] APOA175[243-486] APOA176 APOA176[0-243] APOA176[243-486] APOA177 APOA177[0-243] APOA177[243-486] APOA178 APOA178[0-243] APOA178[243-486] APOA179 APOA179[0-243] APOA179[243-486] APOA180 APOA180[0-243] APOA180[243-486] APOA181 APOA181[0-243] APOA181[243-486] APOA182 APOA182[0-243] APOA182[243-486] APOA183 APOA183[0-243] APOA183[243-486] APOA184 APOA184[0-243] APOA184[243-486] APOA185 APOA185[0-243] APOA185[243-486] APOA186 APOA186[0-243] APOA186[243-486] APOA187 APOA187[0-243] APOA187[243-486] APOA188 APOA188[0-243] APOA188[243-486] APOA189 APOA189[0-243] APOA189[243-486] APOA190 APOA190[0-243] APOA190[243-486] APOA191 APOA191[0-243] APOA191[243-486] APOA192 APOA192[0-243] APOA192[243-486] APOA193 APOA193[0-243] APOA193[243-486] APOA194 APOA194[0-243] APOA194[243-486] APOA195 APOA195[0-243] APOA195[243-486] APOA196 APOA196[0-243] APOA196[243-486] APOA197 APOA197[0-243] APOA197[243-486] APOA198 APOA198[0-243] APOA198[243-486] APOA199 APOA199[0-243] APOA199[243-486] APOA200 APOA200[0-243] APOA200[243-486] APOA201 APOA201[0-243] APOA201[243-486] APOA202 APOA202[0-243] APOA202[243-486] APOA203 APOA203[0-243] APOA203[243-486] APOA204 APOA204[0-243] APOA204[243-486] APOA205 APOA205[0-243] APOA205[243-486] APOA206 APOA206[0-243] APOA206[243-486] APOA207 APOA207[0-243] APOA207[243-486] APOA208 APOA208[0-243] APOA208[243-486] APOA209 APOA209[0-243] APOA209[243-486] APOA210 APOA210[0-243] APOA210[243-486] APOA211 APOA211[0-243] APOA211[243-486] APOA212 APOA212[0-243] APOA212[243-486] APOA213 APOA213[0-243] APOA213[243-486] APOA214 APOA214[0-243] APOA214[243-486] APOA215 APOA215[0-243] APOA215[243-486] APOA216 APOA216[0-243] APOA216[243-486] APOA217 APOA217[0-243] APOA217[243-486] APOA218 APOA218[0-243] APOA218[243-486] APOA219 APOA219[0-243] APOA219[243-486] APOA220 APOA220[0-243] APOA220[243-486] APOA221 APOA221[0-243] APOA221[243-486] APOA222 APOA222[0-243] APOA222[243-486] APOA223 APOA223[0-243] APOA223[243-486] APOA224 APOA224[0-243] APOA224[243-486] APOA225 APOA225[0-243] APOA225[243-486] APOA226 APOA226[0-243] APOA226[243-486] APOA227 APOA227[0-243] APOA227[243-486] APOA228 APOA228[0-243] APOA228[243-486] APOA229 APOA229[0-243] APOA229[243-486] APOA230 APOA230[0-243] APOA230[243-486] APOA231 APOA231[0-243] APOA231[243-486] APOA232 APOA232[0-243] APOA232[243-486] APOA233 APOA233[0-243] APOA233[243-486] APOA234 APOA234[0-243] APOA234[243-486] APOA235 APOA235[0-243] APOA235[243-486] APOA236 APOA236[0-243] APOA236[243-486] APOA237 APOA237[0-243] APOA237[243-486] APOA238 APOA238[0-243] APOA238[243-486] APOA239 APOA239[0-243] APOA239[243-486] APOA240 APOA240[0-243] APOA240[243-486] APOA241 APOA241[0-243] APOA241[243-486] APOA242 APOA242[0-243] APOA242[243-486] APOA243 APOA243[0-243] APOA243[243-486] APOA244 APOA244[0-243] APOA244[243-486] APOA245 APOA245[0-243] APOA245[243-486] APOA246 APOA246[0-243] APOA246[243-486] APOA247 APOA247[0-243] APOA247[243-486] APOA248 APOA248[0-243] APOA248[243-486] APOA249 APOA249[0-243] APOA249[243-486] APOA250 APOA250[0-243] APOA250[243-486] APOA251 APOA251[0-243] APOA251[243-486] APOA252 APOA252[0-243] APOA252[243-486] APOA253 APOA253[0-243] APOA253[243-486] APOA254 APOA254[0-243] APOA254[243-486] APOA255 APOA255[0-243] APOA255[243-486] APOA256 APOA256[0-243] APOA256[243-486] APOA257 APOA257[0-243] APOA257[243-486] APOA258 APOA258[0-243] APOA258[243-486] APOA259 APOA259[0-243] APOA259[243-486] APOA260 APOA260[0-243] APOA260[243-486] APOA261 APOA261[0-243] APOA261[243-486] APOA262 APOA262[0-243] APOA262[243-486] APOA263 APOA263[0-243] APOA263[243-486] APOA264 APOA264[0-243] APOA264[243-486] APOA265 APOA265[0-243] APOA265[243-486] APOA266 APOA266[0-243] APOA266[243-486] APOA267 APOA267[0-243] APOA267[243-486] APOA268 APOA268[0-243] APOA268[243-486] APOA269 APOA269[0-243] APOA269[243-486] APOA270 APOA270[0-243] APOA270[243-486] APOA271 APOA271[0-243] APOA271[243-486] APOA272 APOA272[0-243] APOA272[243-486] APOA273 APOA273[0-243] APOA273[243-486] APOA274 APOA274[0-243] APOA274[243-486] APOA275 APOA275[0-243] APOA275[243-486] APOA276 APOA276[0-243] APOA276[243-486] APOA277 APOA277[0-243] APOA277[243-486] APOA278 APOA278[0-243] APOA278[243-486] APOA279 APOA279[0-243] APOA279[243-486] APOA280 APOA280[0-243] APOA280[243-486] APOA281 APOA281[0-243] APOA281[243-486] APOA282 APOA282[0-243] APOA282[243-486] APOA283 APOA283[0-243] APOA283[243-486] APOA284 APOA284[0-243] APOA284[243-486] APOA285 APOA285[0-243] APOA285[243-486] APOA286 APOA286[0-243] APOA286[243-486] APOA287 APOA287[0-243] APOA287[243-486] APOA288 APOA288[0-243] APOA288[243-486] APOA289 APOA289[0-243] APOA289[243-486] APOA290 APOA290[0-243] APOA290[243-486] APOA291 APOA291[0-243] APOA291[243-486] APOA292 APOA292[0-243] APOA292[243-486] APOA293 APOA293[0-243] APOA293[243-486] APOA294 APOA294[0-243] APOA294[243-486] APOA295 APOA295[0-243] APOA295[243-486] APOA296 APOA296[0-243] APOA296[243-486] APOA297 APOA297[0-243] APOA297[243-486] APOA298 APOA298[0-243] APOA298[243-486] APOA299 APOA299[0-243] APOA299[243-486] APOA300 APOA300[0-243] APOA300[243-486] APOA301 APOA301[0-243] APOA301[243-486] APOA302 APOA302[0-243] APOA302[243-486] APOA303 APOA303[0-243] APOA303[243-486] APOA304 APOA304[0-243] APOA304[243-486] APOA305 APOA305[0-243] APOA305[243-486] APOA306 APOA306[0-243] APOA306[243-486] APOA307 APOA307[0-243] APOA307[243-486] APOA308 APOA308[0-243] APOA308[243-486] APOA309 APOA309[0-243] APOA309[243-486] APOA310 APOA310[0-243] APOA310[243-486] APOA311 APOA311[0-243] APOA311[243-486] APOA312 APOA312[0-243] APOA312[243-486] APOA313 APOA313[0-243] APOA313[243-486] APOA314 APOA314[0-243] APOA314[243-486] APOA315 APOA315[0-243] APOA315[243-486] APOA316 APOA316[0-243] APOA316[243-486] APOA317 APOA317[0-243] APOA317[243-486] APOA318 APOA318[0-243] APOA318[243-486] APOA319 APOA319[0-243] APOA319[243-486] APOA320 APOA320[0-243] APOA320[243-486] APOA321 APOA321[0-243] APOA321[243-486] APOA322 APOA322[0-243] APOA322[243-486] APOA323 APOA323[0-243] APOA323[243-486] APOA324 APOA324[0-243] APOA324[243-486] APOA325 APOA325[0-243] APOA325[243-486] APOA326 APOA326[0-243] APOA326[243-486] APOA327 APOA327[0-243] APOA327[243-486] APOA328 APOA328[0-243] APOA328[243-486] APOA329 APOA329[0-243] APOA329[243-486] APOA330 APOA330[0-243] APOA330[243-486] APOA331 APOA331[0-243] APOA331[243-486] APOA332 APOA332[0-243] APOA332[243-486] APOA333 APOA333[0-243] APOA333[243-486] APOA334 APOA334[0-243] APOA334[243-486] APOA335 APOA335[0-243] APOA335[243-486] APOA336 APOA336[0-243] APOA336[243-486] APOA337 APOA337[0-243] APOA337[243-486] APOA338 APOA338[0-243] APOA338[243-486] APOA339 APOA339[0-243] APOA339[243-486] APOA340 APOA340[0-243] APOA340[243-486] APOA341 APOA341[0-243] APOA341[243-486] APOA342 APOA342[0-243] APOA342[243-486] APOA343 APOA343[0-243] APOA343[243-486] APOA344 APOA344[0-243] APOA344[243-486] APOA345 APOA345[0-243] APOA345[243-486] APOA346 APOA346[0-243] APO
```

```
names(dginnT)<-c("File", "Name", "Gene.name", "GeneSize", "dginn-primate_NbSpecies",
  "dginn-primate_omegaM0codeml", "dginn-primate_BUSTED", "dginn-primate_BUSTED.p.valu
  "dginn-primate_MEME.NbSites", "dginn-primate_MEME.PSS", "dginn-prim
  "dginn-primate_BppM1M2.p.value", "dginn-primate_BppM1M2.NbSites", "dginn-prima
  "dginn-primate_BppM7M8", "dginn-primate_BppM7M8.p.value", "dginn-prim
  "dginn-primate_BppM7M8.PSS", "dginn-primate_codemlM1M2", "dginn-prim
  "dginn-primate_codemlM1M2.NbSites", "dginn-primate_codemlM1M2.PSS", "dginn-prim
  "dginn-primate_codemlM7M8.p.value", "dginn-primate_codemlM7M8.NbSites", "dginn-prim
```

## 1.5 Join Table and DGINN table

```
tab<-merge(tab,dginnT, by="Gene.name", all.x=T)
```

## 1.6 Write new table

```
write.table(tab,
  "COVID_PAMLresults_332hits_plusBatScreens_plusDGINN_20201014.txt",
  row.names=F, quote=F, sep="\t")
```

# 2 Comparisons Primates

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

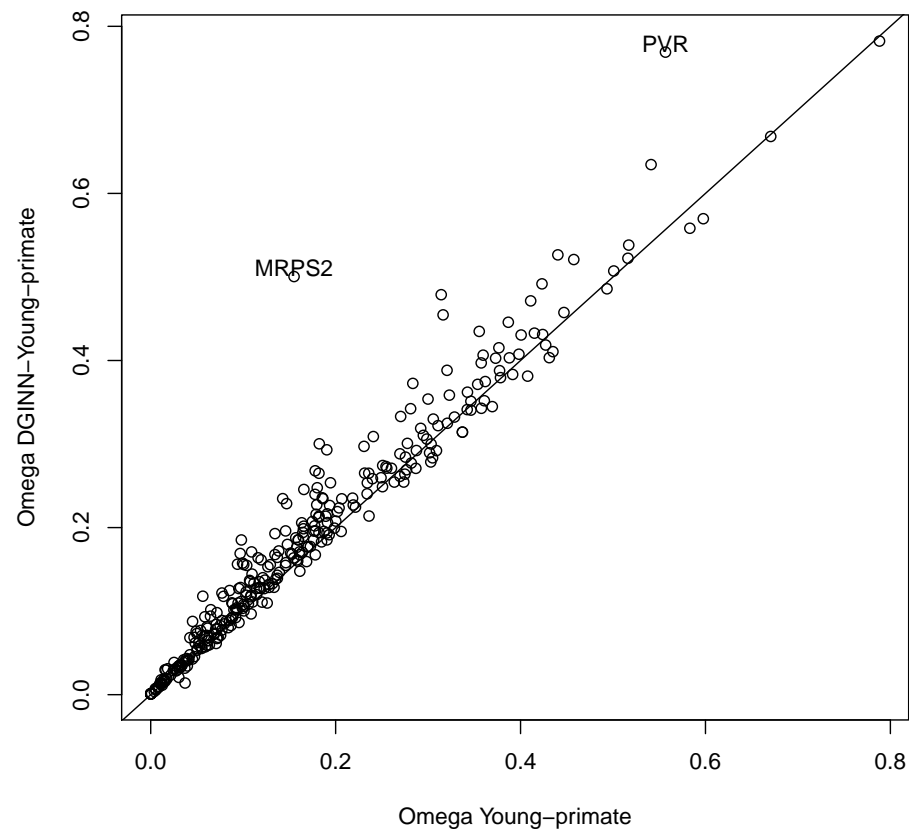
Comparaison des Omega: colonne L "whole.gene.dN.dS.model.0" VS colonne "omega" dans la sortie de dginn.

```
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$'dginn-primate_omegaMOBpp'<-as.numeric(as.character(tab$'dginn-primate_omegaMOBpp'))

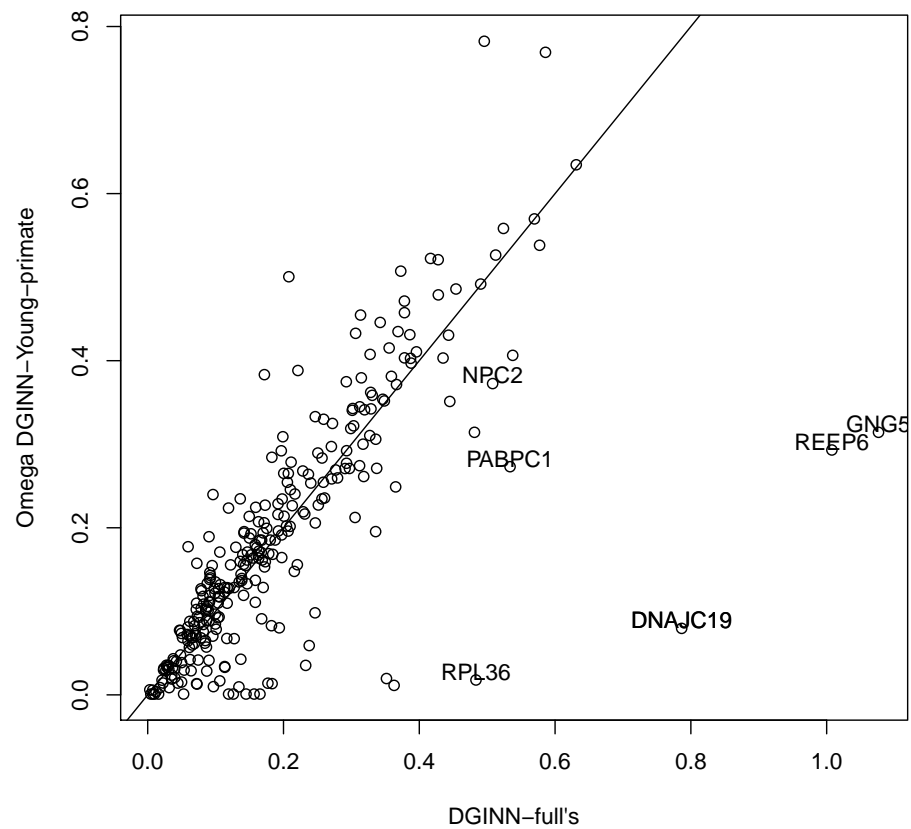
## Warning:  NAs introduits lors de la conversion automatique

plot(tab$'dginn-primate_omegaMOBpp', tab$Omega_PamlM7M8,
      xlab="DGINN-full's", ylab="Omega DGINN-Young-primate")
abline(0,1)

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

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

```



### 2.3 Janet Young's results (Young-primate) VS DGINN-full's results

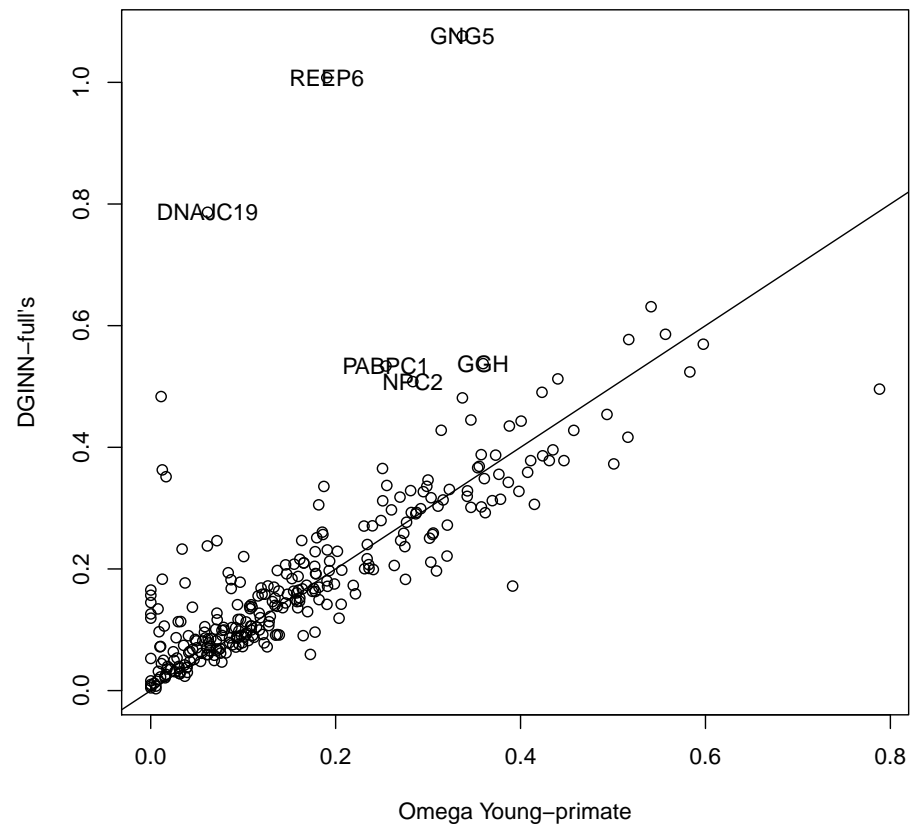
Comparaison des Omega: colonne L "whole.gene.dN.dS.model.0" VS colonne "omega" dans la sortie de dginn.

```
plot(tab$whole.gene.dN.dS.model.0, as.numeric(as.character(tab$'dginn-primate_omegaMOBpp')),
      xlab="Omega Young-primate", ylab="DGINN-full's")
abline(0,1)
```

```
outlier<-tab[tab$whole.gene.dN.dS.model.0<0.4 & tab$'dginn-primate_omegaMOBpp'>0.5,]
```



```
text(x=outlier$whole.gene.dN.dS.model.0,
y=outlier$'dginn-primate_omegaMOBpp',
outlier$Gene.name)
```



### 3 Overlap

#### 3.1 Mondrian

```
library(Mondrian)

#####
```

```

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

## [1] 333 1

dginnyoungtmp<-rowSums(cbind(tab$PosSel_PamlM1M2=="Y", tab$PosSel_PamlM7M8=="Y",
tab$PosSel_BppM1M2=="Y", tab$PosSel_BppM7M8=="Y", tab$PosSel_BUSTED=="Y"))

#monddata$primates_dginn_young<-ifelse(tmp$PosSel_PamlM7M8=="Y", 1,0)

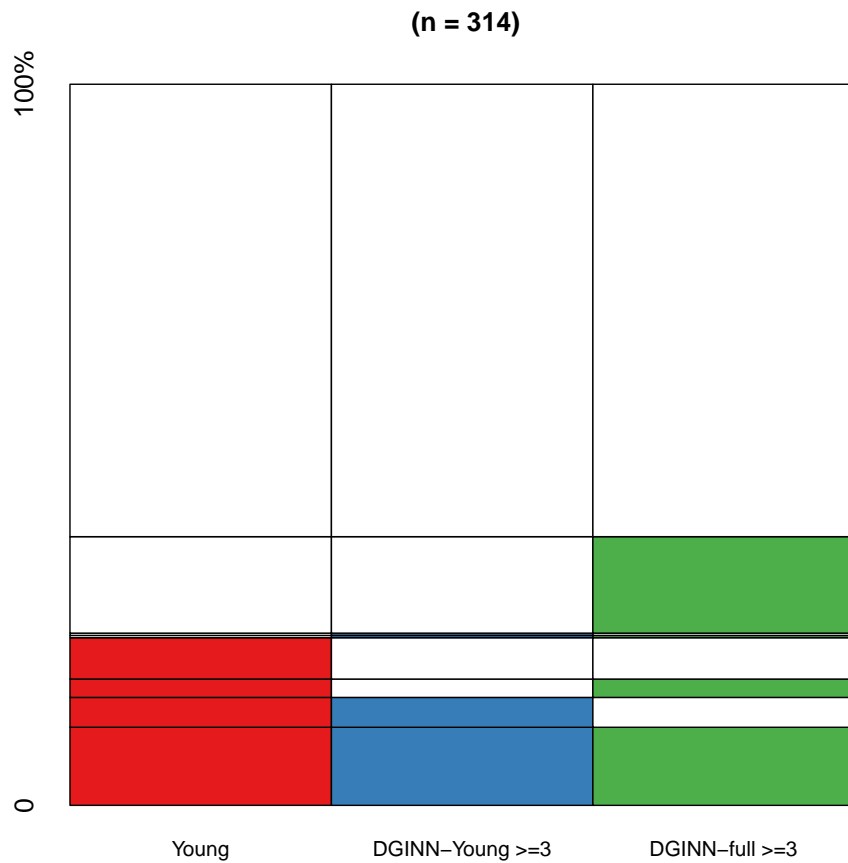
dginfulltmp<-rowSums(cbind(tab$'dginn-primate_BUSTED'=="Y", tab$'dginn-primate_BppM1
tab$'dginn-primate_BppM7M8'=="Y", tab$'dginn-primate_codemlM1M2'=="Y", tab$'dginn-pri

monddata$primates_young<-ifelse(tab$pVal.M8vsM7<0.05, 1, 0)
#monddata$primates_cooper<-ifelse(tab$cooper.primates.M7.M8_p_val<0.05, 1, 0)

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

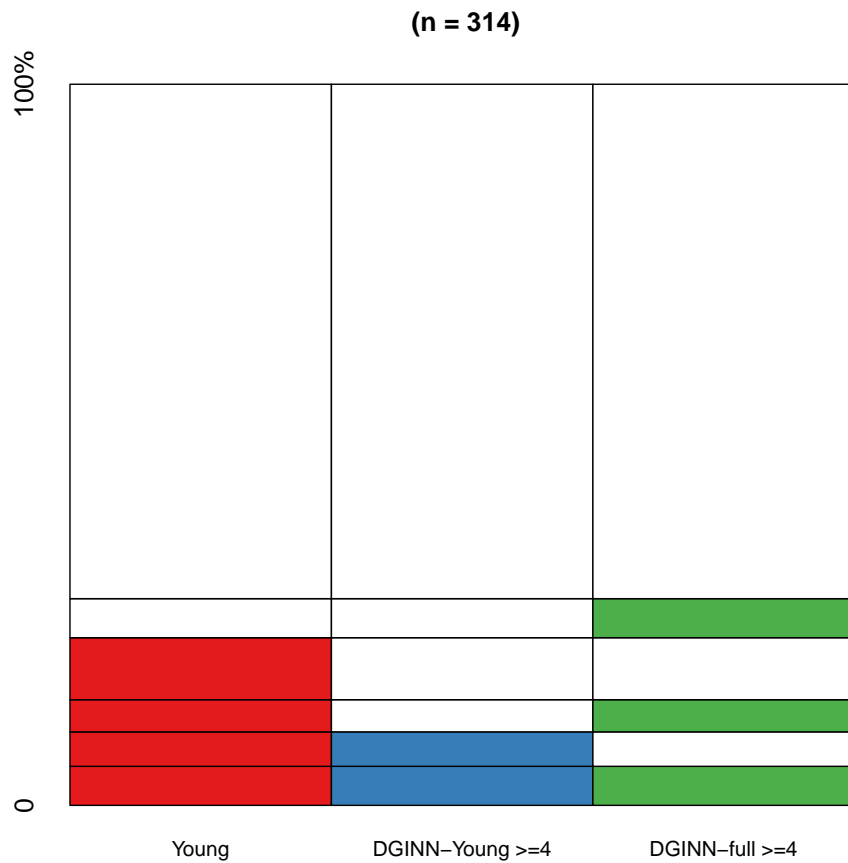
mondrian(na.omit(monddata[,2:4]), labels=c("Young", "DGINN-Young >=3", "DGINN-full >=

```



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

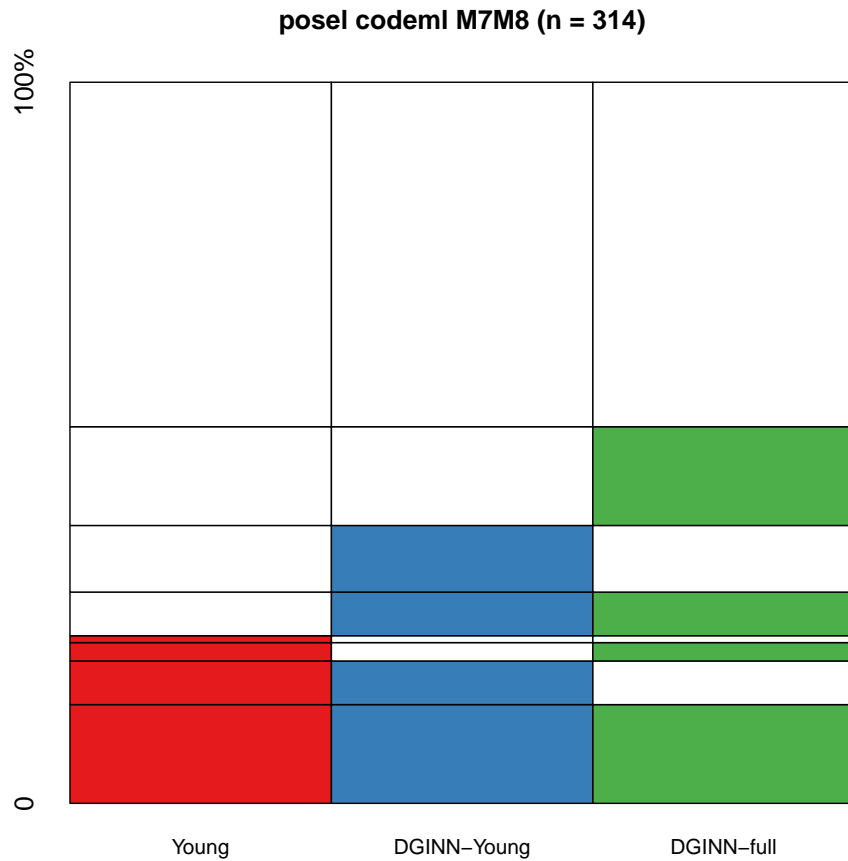
mondrian(na.omit(monddata[,2:4]), labels=c("Young", "DGINN-Young >=4", "DGINN-full >=4"))
```



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"), mai
```



### 3.2 subsetR

Just another representation of the same result.

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

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

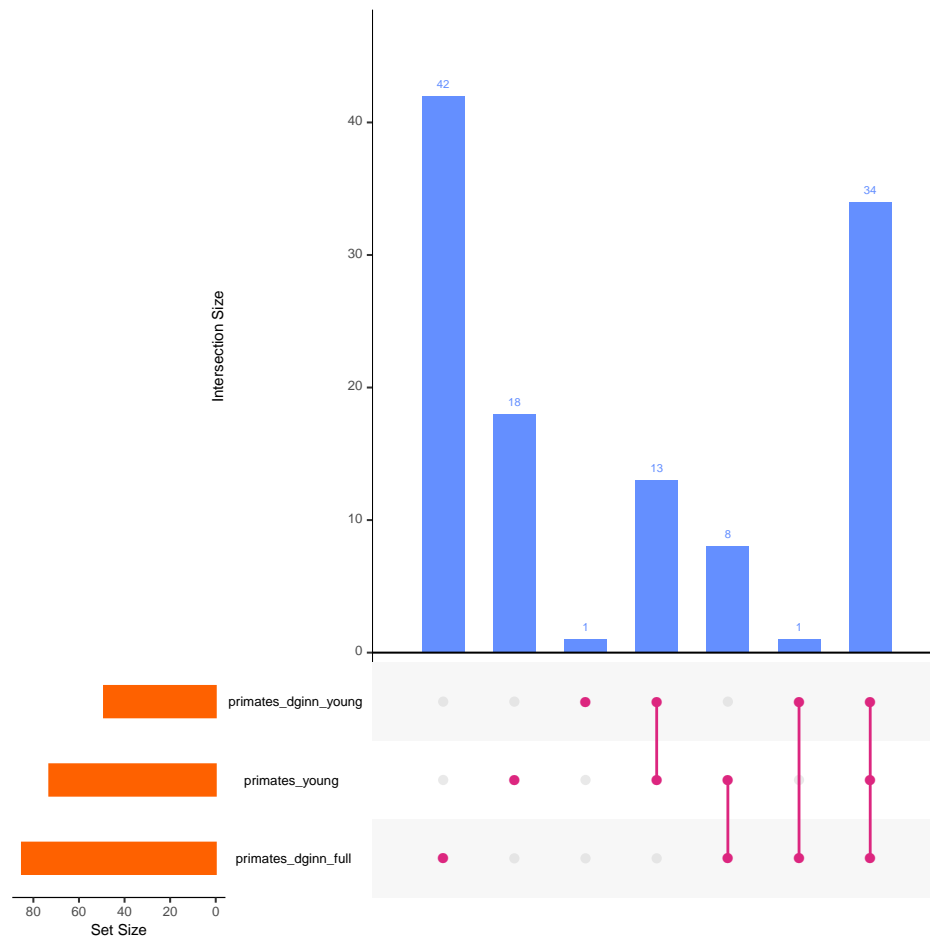
###
```

```

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")

```



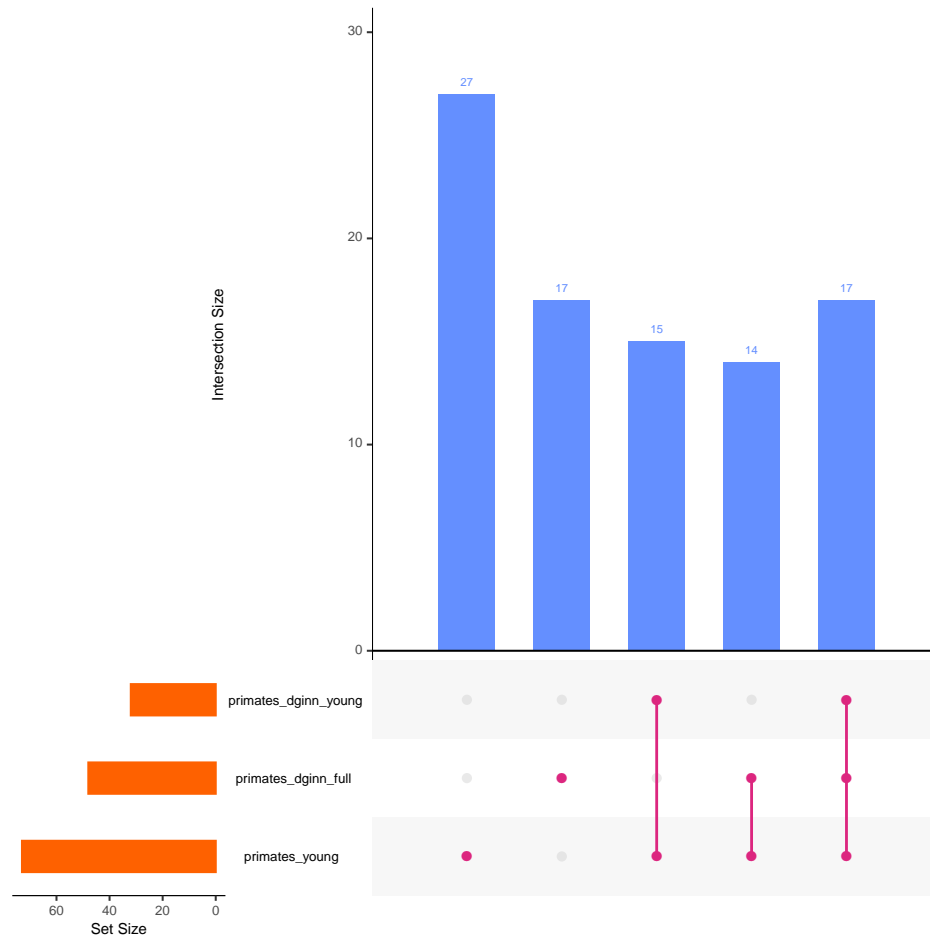
```

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

upset(na.omit(upsetdata), nsets = 3, matrix.color = "#DC267F",

```

```
main.bar.color = "#648FFF", sets.bar.color = "#FE6100")
```



## 4 Gene List

Genes under positive selection for at least 4 methods.

```
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"))
```

```

tab$Gene.name[dginnfulltmp>=4 & is.na(dginnfulltmp)==F]

## [1] "ACADM"      "BCS1L"      "BRD4"      "CDK5RAP2"   "CEP135"     "CEP68"      "CLIP
## [11] "FYC01"      "GCC2"       "GGH"       "GHITM"      "GIGYF2"     "GLA"        "GOLG
## [21] "LARP1"      "LARP4B"     "LMAN2"     "MARK1"      "MIPOL1"     "MPHOSPH10"  "MYCB
## [31] "POLA1"      "PRIM2"      "PRKAR2A"   "PVR"        "REEP6"      "RIPK1"      "SAAL
## [41] "SLC27A2"    "TMEM39B"    "TOR1AIP1"  "TUBGCP2"    "UBAP2"      "UGGT2"      "VPS3

tab$Gene.name[dginnfulltmp>=3 & is.na(dginnfulltmp)==F]

## [1] "ACADM"      "ADAM9"      "AP2A2"     "ATE1"       "BCS1L"      "BRD4"      "BZW2
## [11] "CLIP4"      "CNTRL"      "DNMT1"     "DPH5"       "EDEM3"      "EIF4E2"     "EMC1
## [21] "GGH"        "GHITM"      "GIGYF2"    "GLA"        "GOLGA7"     "GOLGB1"     "GORA
## [31] "IDE"        "ITGB1"      "LARP1"     "LARP4B"     "LARP7"      "LMAN2"      "MARK
## [41] "MPHOSPH10"  "MRPS5"      "MYCBP2"    "NAT14"      "NDUFAF2"    "NDUFB9"     "NGLY
## [51] "PLAT"       "PLOD2"      "PMPCB"     "POLA1"      "POR"        "PRIM2"      "PRKA
## [61] "RAB1A"      "RAB2A"      "RAP1GDS1"  "RBX1"       "REEP6"      "RIPK1"      "RPL3
## [71] "SIRT5"      "SLC25A21"   "SLC27A2"   "STOM"       "TIMM8B"     "TMEM39B"    "TOR1
## [81] "UBAP2"      "UGGT2"      "USP54"     "VPS39"      "ZNF318"

tmp<-tab[dginnfulltmp>=4 & is.na(dginnfulltmp)==F,
c("Gene.name","dginn-primate_BUSTED", "dginn-primate_BppM1M2",
  "dginn-primate_BppM7M8","dginn-primate_codemlM1M2","dginn-primate_codemlM7M8")]

write.table(tmp, "geneList_DGINN_full_primate_pos4.txt", row.names=F, quote=F)

```

## 5 Shiny like

```

makeFig1 <- function(df){

  # prepare data for colors etc
  colMethods <- c("deepskyblue4", "darkorange", "deepskyblue3", "mediumseagreen",
nameMethods <- c("BUSTED", "BppM1M2", "BppM7M8", "codemlM1M2", "codemlM7M8", "MEME"
metColor <- data.frame(Name = nameMethods , Col = colMethods , stringsAsFactors = F

  # subset for this specific figure
  #df <- df[df$nbY >= 1, ] # to drop genes found by 0 methods (big datasets)

```



```

xt <- df[, c("BUSTED", "BppM1M2", "BppM7M8", "codemlM1M2", "codemlM7M8")]
xt$Gene <- df$Gene
nbrMeth <- 5
# reverse order of dataframe so that genes with the most Y are at the bottom (to be
xt[,1:5] <- ifelse(xt[,1:5] == "Y", 1, 0)
# sort and Filter the 0 lines
xt<-xt[order(rowSums(xt[,1:5])),]
xt<-xt[rowSums(xt[,1:5])>2,]

row.names(xt)<-xt$Gene
xt<-xt[,1:5]

colFig1 <- metColor[which(metColor$Name %in% colnames(xt)) , ]

##### PART 1 : NUMBER OF METHODS
par(xpd = NA , mar=c(2,7,4,0) , oma = c(0,0,0,0) , mgp = c(3,0.3,0))

h = barplot(
  t(xt),
  border = NA ,
  axes = F ,
  col = adjustcolor(colFig1$Col, alpha.f = 1),
  horiz = T ,
  las = 2 ,
  main = "Methods detecting positive selection" ,
  cex.main = 0.85,
  cex.names = min(50/nrow(xt), 1.5)
)

axis(3, line = 0, at = c(0:nbrMeth), label = c("0", rep("", nbrMeth -1), nbrMeth),

legend("bottomleft",
  horiz = T,
  border = colFig1$Col,
  legend = colFig1$Name,
  fill = colFig1$Col,
  cex = 0.8,
  bty = "n",
  xpd = NA

```

```
)  
}  
  
df<-read.delim(paste0(workdir,  
"/data/DGINN_202005281649summary_cleaned.csv"),  
fill=T, h=T, sep=",")  
  
makeFig1(df)
```

