Genes with COMMONN SNVs from gnomad and clinvar

mfpfox

2023-06-09

```r
source("1_import_annotated_variants.R")
```

```
## Warning: `funs()` was deprecated in dplyr 0.8.0.
## Please use a list of either functions or lambdas:
##
##   # Simple named list:
##   list(mean = mean, median = median)
##
##   # Auto named with `tibble::lst()`:
##   tibble::lst(mean, median)
##
##   # Using lambdas
##   list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was generated.
## 'data.frame':    9110589 obs. of  30 variables:
##  $ keyID37aa                 : chr  "10_000093000_G/A_A" "10_000093003_C/T_V" "10_00009
3004_A/G_V/A" "10_000093007_T/A_E/V" ...
##  $ CDS.position              : chr  "1332" "1329" "1328" "1325" ...
##  $ Protein.position          : chr  "444" "443" "443" "442" ...
##  $ Amino.acids               : chr  "A" "V" "V/A" "E/V" ...
##  $ Codons                    : chr  "gcC/gcT" "gtG/gtA" "gTg/gCg" "gAg/gTg" ...
##  $ SYMBOL                    : chr  "TUBB8" "TUBB8" "TUBB8" "TUBB8" ...
##  $ SYMBOL.SOURCE             : chr  "HGNC" "HGNC" "HGNC" "HGNC" ...
##  $ SIFT                      : chr  "-" "-" "tolerated_low_confidence" "deleterious_low
_confidence" ...
##  $ SIFT.score                : num  NA NA 0.62 0 0.15 0.05 NA NA 0.69 0.6 ...
##  $ PolyPhen                  : chr  "-" "-" "benign" "benign" ...
##  $ PolyPhen.score            : num  NA NA 0 0.013 0.557 0.305 NA NA 0.001 0.001 ...
##  $ DOMAINS                   : chr  "-" "Coiled-coils_(Ncoils):Coil" "Coiled-coils_(Nco
ils):Coil" "Coiled-coils_(Ncoils):Coil,Low_complexity_(Seg):seg" ...
##  $ Source                    : chr  "['WGS', 'WES']" "['WES']" "['WES']" "['WES']" ...
##  $ AC                        : num  10 1 2 1 1 1 2 1 1 11 ...
##  $ AN                        : num  194416 171722 174636 180624 194710 ...
##  $ nhomalt                   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ AF                        : num  5.14e-05 5.82e-06 1.15e-05 5.54e-06 5.14e-06 ...
##  $ nhomalt.x2                : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ nhetalt                   : num  10 1 2 1 1 1 2 1 1 11 ...
##  $ ratio.nhomalt.over.nhetalt: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ keyAA                     : chr  "A" "V" "V/A" "E/V" ...
##  $ CONSEQ                    : chr  "synonymous_variant" "synonymous_variant" "missense
_variant" "missense_variant" ...
##  $ clinvarAA                 : chr  NA NA NA NA ...
##  $ clinvarGeneSymbol         : chr  NA NA NA NA ...
##  $ clinvarCONSEQ             : chr  NA NA NA NA ...
##  $ HGVSp.VEP                 : chr  NA NA NA NA ...
##  $ HGVSc.VEP                 : chr  NA NA NA NA ...
##  $ StarReviewStatus          : chr  NA NA NA NA ...
##  $ myClinVarLabels           : chr  NA NA NA NA ...
##  $ LABEL                     : chr  NA NA NA NA ...
```

```
# used colors
unusedcolors = c("skyblue", "#FD6467","#F4B5BD","#3B9AB2"  ,
                     "#DD8D29", "#E2D200", "#46ACC8",
                    "#7294D4", "#C6CDF7","#FD6467", "#5B1A18",
                    "#F2AD00","#90D4CC","#FD6467","#00A08A",
                    "#FF0000", "#08519c",
                 "red2", "orange2", "pink1")



maf_colors = c("#C6CDF7", "plum", "purple3")

maf2_colors = c("#C6CDF7", "purple3")

var_colors = c( "blue", "#85D4E3", "green4")

class_colors =  c( "#DD8D29", "#E2D200", "#46ACC8")

goflof_colors = c( "#FF0000", "#08519c" ,"#D9D0D3")
```

```
SUM.cv =  "1,550,594"
SUM.gnomad =  "8,390,678"
SUB1 = paste0("ClinVar v202304 n = ", SUM.cv)
SUB2 = paste0("gnomAD v2.1.1 n = ", SUM.gnomad)
SUB.total = paste(SUB1, SUB2, sep = "\n")
DBname <- "Exclusive ClinVar SNV n = 719,911\nExclusive gnomAD SNV n = 7,559,995\nOverlap gnomAD & ClinVar SNV n = 830,683"
SUB0 = "Overlap gnomAD & ClinVar SNV n = 830,683"
```

what genes have common stop_gained?

```
common_nonsense <- wgs %>%  filter(gnomadCONSEQ == "stop_gained")
print(length(unique(common_nonsense$SYMBOL)))
```

```
## [1] 18224
```

```
common_nonsense <- common_nonsense[order(-common_nonsense$AF), ]
common_nonsense <- common_nonsense[1:50, ]
print(length(unique(common_nonsense$SYMBOL)))
```

```
## [1] 50
```

```
datatable(common_nonsense, options = list(pageLength=5, scrollX='400px'), filter = 'top')
```

Show 5 **entries**

Search:

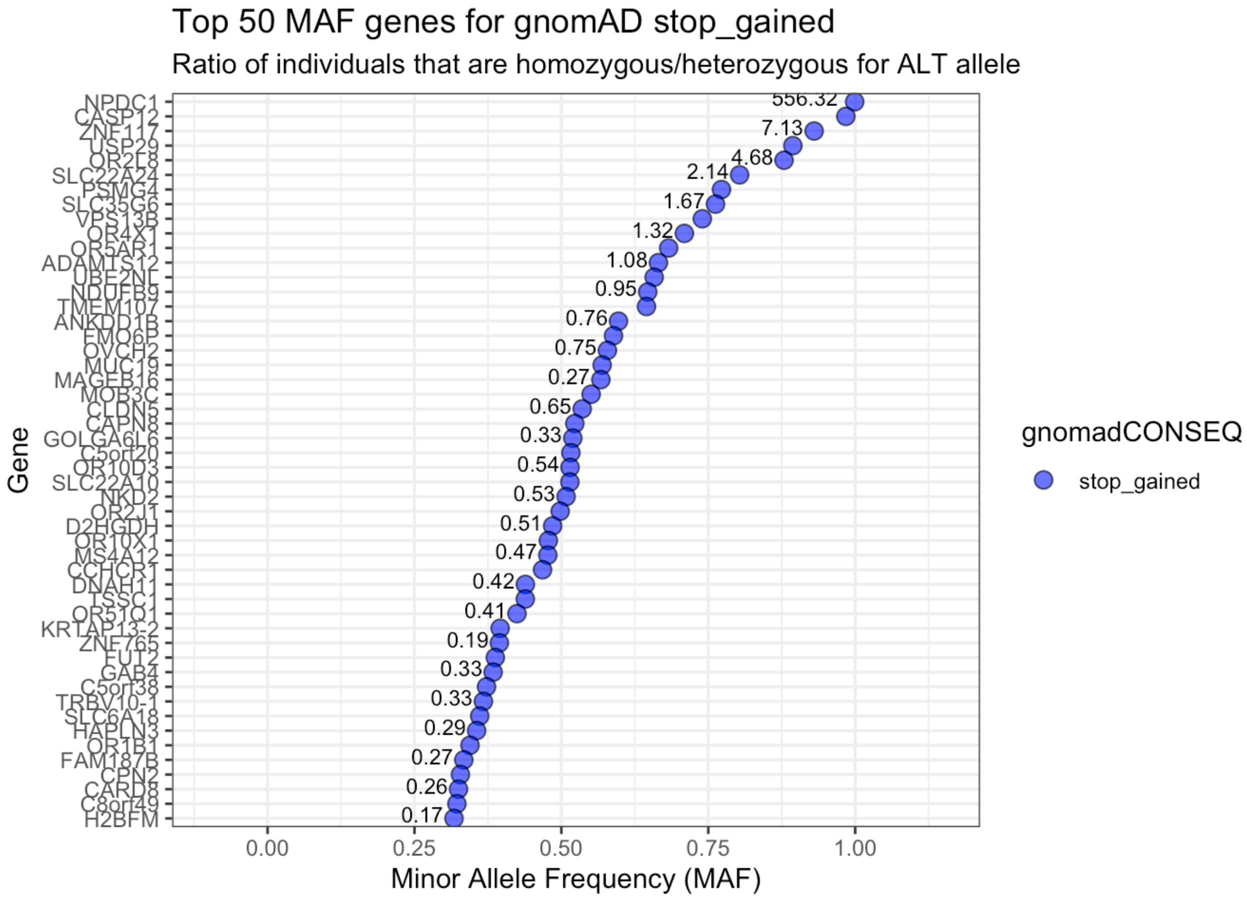
	keyID37aa	CDS.position	Protein.position	Amino.acids	Codons	SYMI

	All	All	All	All		
164396	9_139937799_G/A_R/*	73	25	R/*	Cga/Tga	NPDC1
14924	11_104763117_G/A_R/*	373	125	R/*	Cga/Tga	CASP1
146743	7_064438667_G/A_R/*	1282	428	R/*	Cga/Tga	ZNF11
72208	19_057642782_C/A_Y/*	2739	913	Y/*	taC/taA	USP29
90901	1_248113026_T/A_Y/*	867	289	Y/*	taT/taA	OR2L8

Showing 1 to 5 of 50 entries

Previous 1 2 3 4 5 … 10 Next

```
plot.common_nonsense = ggplot(common_nonsense, aes(x= reorder(SYMBOL, AF),  y=AF, color=gnomadCONSEQ)) +
    geom_point(shape = 21,
              colour = "black",
              aes(fill = gnomadCONSEQ),
              size = 3,
              stroke = 0.5,
             alpha=0.7) +
   geom_text(aes(label=paste0("", as.character(round(ratio.nhomalt.over.nhetalt, 2)))),
               size=3,
             color="black",
               hjust=1.25,
               vjust=0.25,
               show.legend = FALSE,
             check_overlap = TRUE) +
  scale_y_continuous(limits = c(-0.1, 1.15), n.breaks =6) +
  scale_color_manual(values=var_colors) +
   scale_fill_manual(values=var_colors) +
 labs(title= "Top 50 MAF genes for gnomAD stop_gained",
       subtitle = "Ratio of individuals that are homozygous/heterozygous for ALT allele",
       y="Minor Allele Frequency (MAF)",
       x="Gene") +
  theme_bw() +
  coord_flip()
plot.common_nonsense
```



```
ggsave("Top50_genes_w_common_gnomad_stopgain.png", width=8.5, height=9)
#facet_grid(ProteinConsequence ~ .) +
```

what genes have common missense?

```
common_missense <- wgs %>%  filter(gnomadCONSEQ == "missense_variant")
print(length(unique(common_missense$SYMBOL)))
```

```
## [1] 20212
```

```
common_missense <- common_missense[order(-common_missense$AF), ]
common_missense <- common_missense[1:50, ]
print(length(unique(common_missense$SYMBOL)))
```

```
## [1] 46
```

```
datatable(common_missense, options = list(pageLength=5, scrollX='400px'), filter = 'top')
```

Show 5 **entries** **Search:**

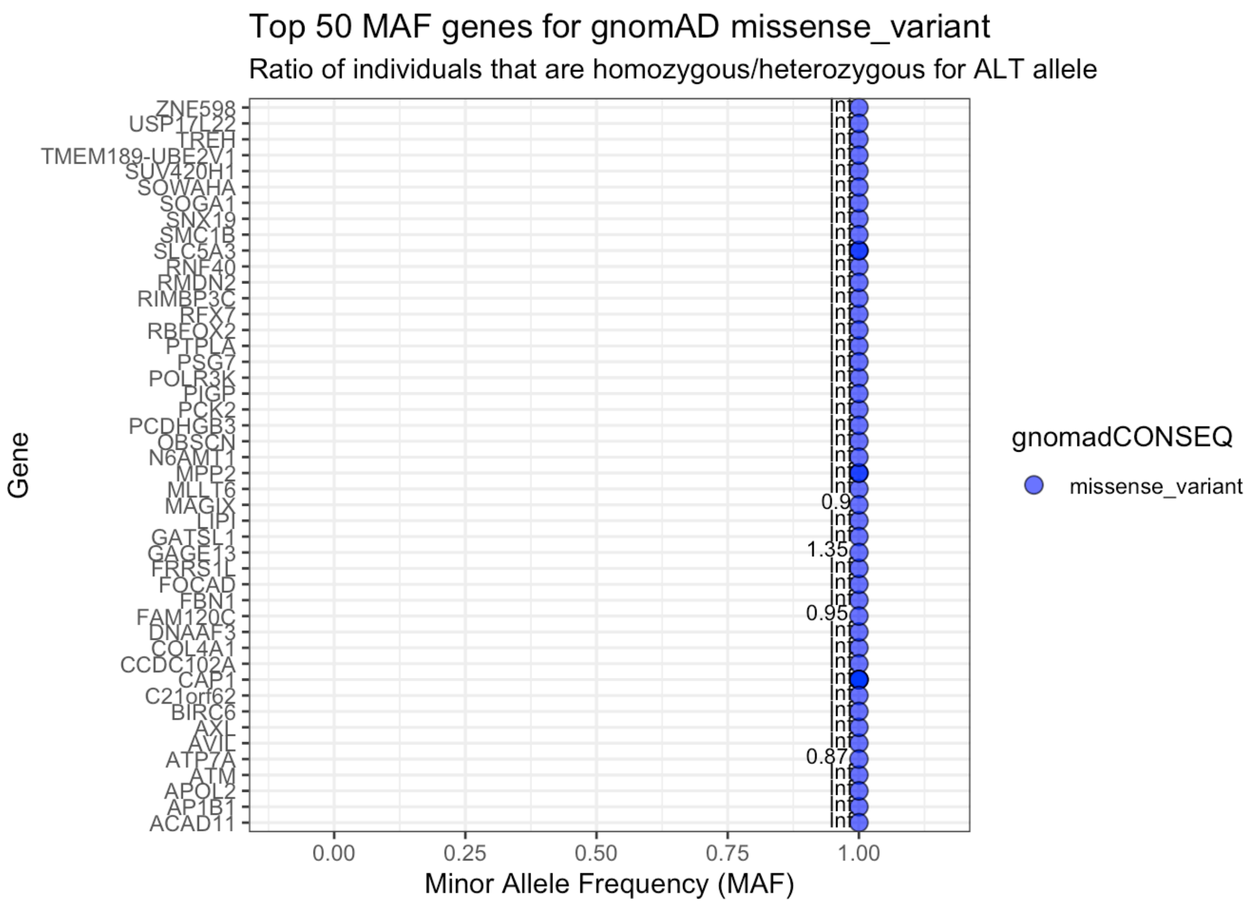
	keyID37aa	CDS.position	Protein.position	Amino.acids	Codons	SYM

	All	All	All	All	.	/
24063	10_017659131_C/A_V/F	208	70	V/F	Gtc/Ttc	PTPLA
441799	11_067957518_A/T_I/N	26	9	I/N	aTc/aAc	SUV42
511601	11_108183167_A/G_N/S	5948	1983	N/S	aAt/aGt	ATM
536270	11_118529069_G/C_P/A	1681	561	P/A	Cct/Gct	TREH
571078	11_130780225_C/A_L/F	1854	618	L/F	ttG/ttT	SNX1

Showing 1 to 5 of 50 entries

Previous 1 2 3 4 5 … 10 Next

```
plot.common_missense = ggplot(common_missense, aes(x= reorder(SYMBOL, AF),  y=AF, color=gnomadCONSEQ)) +
    geom_point(shape = 21,
              colour = "black",
              aes(fill = gnomadCONSEQ),
              size = 3,
              stroke = 0.5,
             alpha=0.7) +
   geom_text(aes(label=paste0("", as.character(round(ratio.nhomalt.over.nhetalt, 2)))),
                size=3,
             color="black",
               hjust=1.25,
               vjust=0.25,
                show.legend = FALSE,
             check_overlap = TRUE) +
  scale_y_continuous(limits = c(-0.1, 1.15), n.breaks =6) +
  scale_color_manual(values=var_colors) +
   scale_fill_manual(values=var_colors) +
 labs(title= "Top 50 MAF genes for gnomAD missense_variant",
       subtitle = "Ratio of individuals that are homozygous/heterozygous for ALT allele",
       y="Minor Allele Frequency (MAF)",
       x="Gene") +
  theme_bw() +
  coord_flip()
plot.common_missense
```



```
ggsave("Top50_genes_w_common_gnomad_missense.png", width=8.5, height=9)
```

what genes have common silent?

```
common_silent <- wgs %>%  filter(gnomadCONSEQ == "synonymous_variant")
print(length(unique(common_silent$SYMBOL)))
```

```
## [1] 20177
```

```
common_silent <- common_silent[order(-common_silent$AF), ]
common_silent <- common_silent[1:50, ]
print(length(unique(common_silent$SYMBOL)))
```

```
## [1] 48
```

```
datatable(common_silent, options = list(pageLength=5, scrollX='400px'), filter = 'top')
```

Show 5 entries

Search:

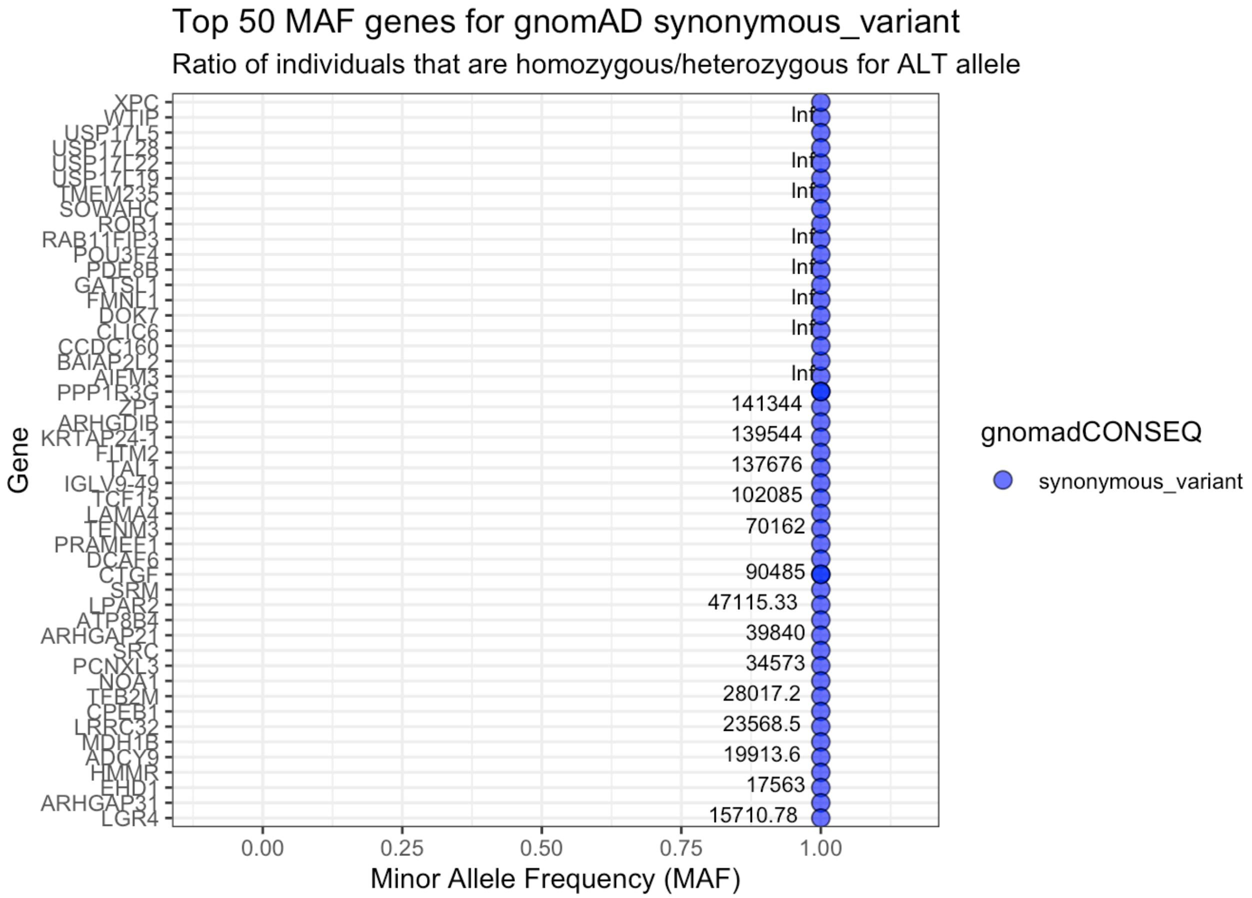
	keyID37aa	CDS.position	Protein.position	Amino.acids	Codons	SYM

	All	All	All	All		A
654854	16_000476096_A/G_A	90	30	A	gcA/gcG	RAB1
890449	17_043318778_G/C_G	1362	454	G	ggG/ggC	FMNI
939529	17_076228088_T/A_P	534	178	P	ccT/ccA	TMEI
1104565	19_034973413_T/C_A	534	178	A	gcT/gcC	WTIP
1318893	1_064608329_G/T_A	1170	390	A	gcG/gcT	ROR1

Showing 1 to 5 of 50 entries

Previous 1 2 3 4 5 … 10 Next

```
plot.common_silent = ggplot(common_silent, aes(x= reorder(SYMBOL, AF),  y=AF, color=gnomadCONSEQ)) +
    geom_point(shape = 21,
              colour = "black",
              aes(fill = gnomadCONSEQ),
              size = 3,
              stroke = 0.5,
             alpha=0.7) +
   geom_text(aes(label=paste0("", as.character(round(ratio.nhomalt.over.nhetalt, 2)))),
                size=3,
                #fontface = "bold",
             color="black",
               hjust=1.25,
               vjust=0.25,
                show.legend = FALSE,
             check_overlap = TRUE) +
  scale_y_continuous(limits = c(-0.1, 1.15), n.breaks =6) +
  scale_color_manual(values=var_colors) +
   scale_fill_manual(values=var_colors) +
 labs(title= "Top 50 MAF genes for gnomAD synonymous_variant",
       subtitle = "Ratio of individuals that are homozygous/heterozygous for ALT allele",
       y="Minor Allele Frequency (MAF)",
       x="Gene") +
  theme_bw() +
  coord_flip()
plot.common_silent
```



```
ggsave("Top50_genes_w_common_gnomad_synonymous.png", width=8.5, height=9)
```

what genes have common PATHO?

```
common_patho <- wgs %>%  filter(myClinVarLabels == "PATHO") %>% filter(MAF2 == "Common (>=
5%)")
print(length(unique(common_patho$clinvarGeneSymbol)))
```

```
## [1] 23
```

```
print(length(unique(common_patho$keyID37aa)))
```

```
## [1] 26
```

```
common_patho <- common_patho[order(-common_patho$AF), ]
# common_patho <- common_patho[1:50, ]
# print(length(unique(common_patho$clinvarGeneSymbol)))

datatable(common_patho, options = list(pageLength=5, scrollX='400px'), filter = 'top')
```

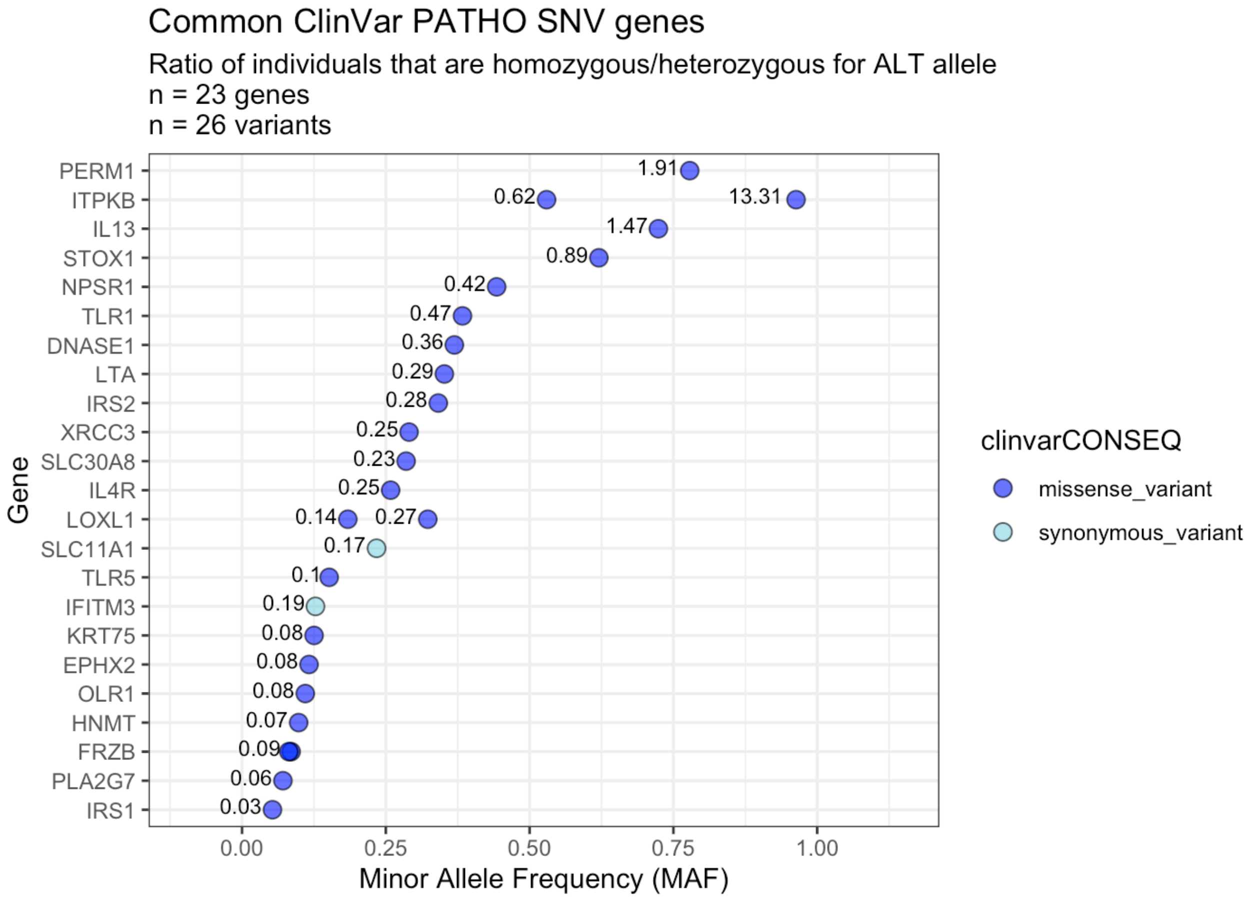
Show 5 **entries** **Search:**

	keyID37aa	CDS.position	Protein.position	Amino.acids	Codons	SYMBOL
	All	All	All	All	.	A
13	1_226923505_G/T_P/Q	1655	552	P/Q	cCg/cAg	ITPKB
11	1_000911595_A/G_V/A	2048	683	V/A	gTg/gCg	C1orf170
21	5_131995964_A/G_Q/R	431	144	Q/R	cAg/cGg	IL13
1	10_070641860_T/C_Y/H	457	153	Y/H	Tac/Cac	STOX1
14	1_226923938_A/C_S/A	1222	408	S/A	Tcc/Gcc	ITPKB

Showing 1 to 5 of 26 entries

Previous 1 2 3 4 5 6 Next

```
plot.common.patho = ggplot(common_patho,
                           aes(x= reorder(clinvarGeneSymbol, AF),  y=AF)) +
    geom_point(shape = 21,
              colour = "black",
              aes(fill = clinvarCONSEQ),
              size = 3,
              stroke = 0.5,
             alpha=0.7) +
      # geom_text(aes(label= HGVSp.VEP, color=clinvarCONSEQ),
      #           size=3,
      #           fontface = "bold",
      #          hjust= -0.8,
      #          vjust=0.25,
      #           show.legend = FALSE,
      #        check_overlap = TRUE) +
   geom_text(aes(label=paste0("", as.character(round(ratio.nhomalt.over.nhetalt, 2)))),
                size=3,
               color="black",
               hjust= 1.25,
               vjust=0.25,
                show.legend = FALSE,
              check_overlap = TRUE) +
  scale_y_continuous(limits = c(-0.1, 1.15), n.breaks =6) +
   scale_fill_manual(values=var_colors) +
     scale_color_manual(values=var_colors) +
labs(title= "Common ClinVar PATHO SNV genes",
       subtitle = "Ratio of individuals that are homozygous/heterozygous for ALT allele\nn
= 23 genes\nn = 26 variants",
       y="Minor Allele Frequency (MAF)",
       x="Gene") +
  theme_bw() +
  coord_flip()
plot.common.patho
```



```
ggsave("Common_clinvar_patho_genes.png", width=8, height=7)
```

what genes have common VUS?

```
common_vus <- wgs %>%  filter(myClinVarLabels == "VUS") %>% filter(MAF2 == "Common (>=
5%)")
print(length(unique(common_vus$clinvarGeneSymbol)))
```

```
## [1] 85
```

```
print(length(unique(common_vus$keyID37aa)))
```

```
## [1] 103
```

```
common_vus <- common_vus[order(-common_vus$AF), ]
common_vus <- common_vus[1:50, ]
print(length(unique(common_vus$clinvarGeneSymbol)))
```

```
## [1] 43
```

```
print(length(unique(common_vus$keyID37aa)))
```

```
## [1] 50
```

```
datatable(common_vus, options = list(pageLength=5, scrollX='400px'), filter = 'top')
```

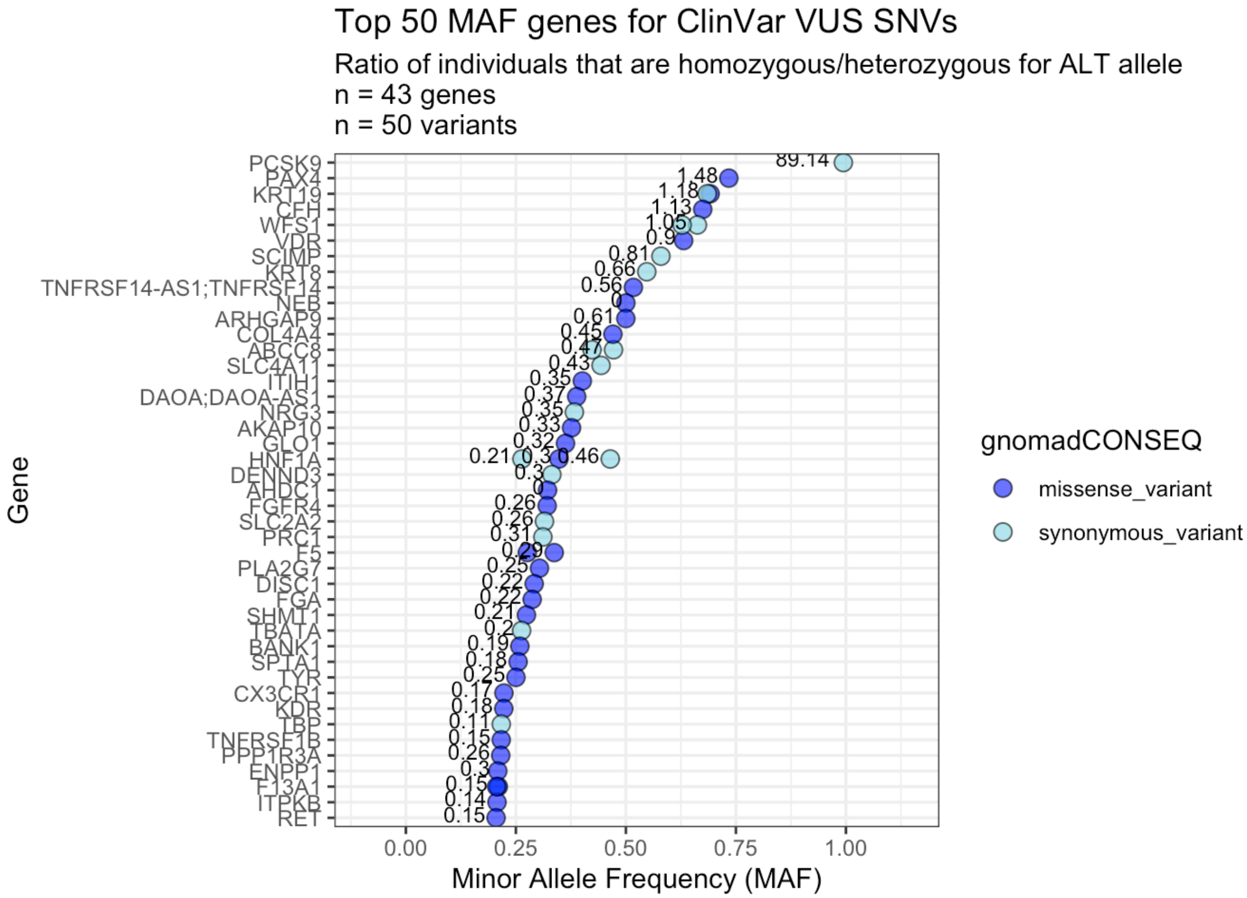
Show 5 **entries** **Search:**

	keyID37aa	**CDS.position**	**Protein.position**	**Amino.acids**	**Codons**	**SYMBOL**
	All	All	All	All	.	/
50	1_055523033_A/G_Q	1026	342	Q	caA/caG	PCSK9
97	7_127251188_T/G_H/P	962	321	H/P	cAc/cCc	PAX4
35	17_039684321_G/C_A/G	179	60	A/G	gCc/gGc	KRT19
34	17_039681475_A/G_N	471	157	N	aaT/aaC	KRT19
57	1_196659237_C/T_H/Y	1204	402	H/Y	Cat/Tat	CFH

Showing 1 to 5 of 50 entries

Previous 1 2 3 4 5 … 10 Next

```
plot.common.vus = ggplot(common_vus, aes(x= reorder(clinvarGeneSymbol, AF),  y=AF)) +
    geom_point(shape = 21,
              colour = "black",
              aes(fill = gnomadCONSEQ),
              size = 3,
              stroke = 0.5,
             alpha=0.7) +
   geom_text(aes(label=paste0("", as.character(round(ratio.nhomalt.over.nhetalt, 2)))),
                size=3,
             color="black",
               hjust=1.25,
               vjust=0.25,
                show.legend = FALSE,
             check_overlap = TRUE) +
   # geom_text(aes(label= HGVSp.VEP, color=clinvarCONSEQ),
   #              size=3,
   #              fontface = "bold",
   #             hjust= 1.25,
   #             vjust=0.25,
   #              show.legend = FALSE,
   #           check_overlap = TRUE) +
  scale_y_continuous(limits = c(-0.1, 1.15), n.breaks =6) +
   scale_fill_manual(values=var_colors) +
     scale_color_manual(values=var_colors) +
labs(title= "Top 50 MAF genes for ClinVar VUS SNVs",
        subtitle = "Ratio of individuals that are homozygous/heterozygous for ALT allele\n
n = 43 genes\nn = 50 variants",
       y="Minor Allele Frequency (MAF)",
       x="Gene") +
  theme_bw() +
  coord_flip()
plot.common.vus
```



```
ggsave("Top50_genes_w_common_clinvar_vus.png", width=8.5, height=9)
```