Clinvar mutations in clients of hsp

Load the data

```
clinvar_path <- read.delim('../../body/1raw/clinvar_patho_missense.tsv', sep = ' ')
clinvar_ben <- read.delim('../../body/1raw/clinvar_benign_missense.tsv', sep = ' ')
clients_hsp90 <- read.delim('../../body/2derived/clients.ids.list.txt', header = F)
colnames(clients_hsp90) <- 'Gene'
nonclients_hsp90 <- read.delim('../../body/2derived/nonclients.ids.list.txt', header = F)
colnames(nonclients_hsp90) <- 'Gene'
```

Merge clinvar and clients information

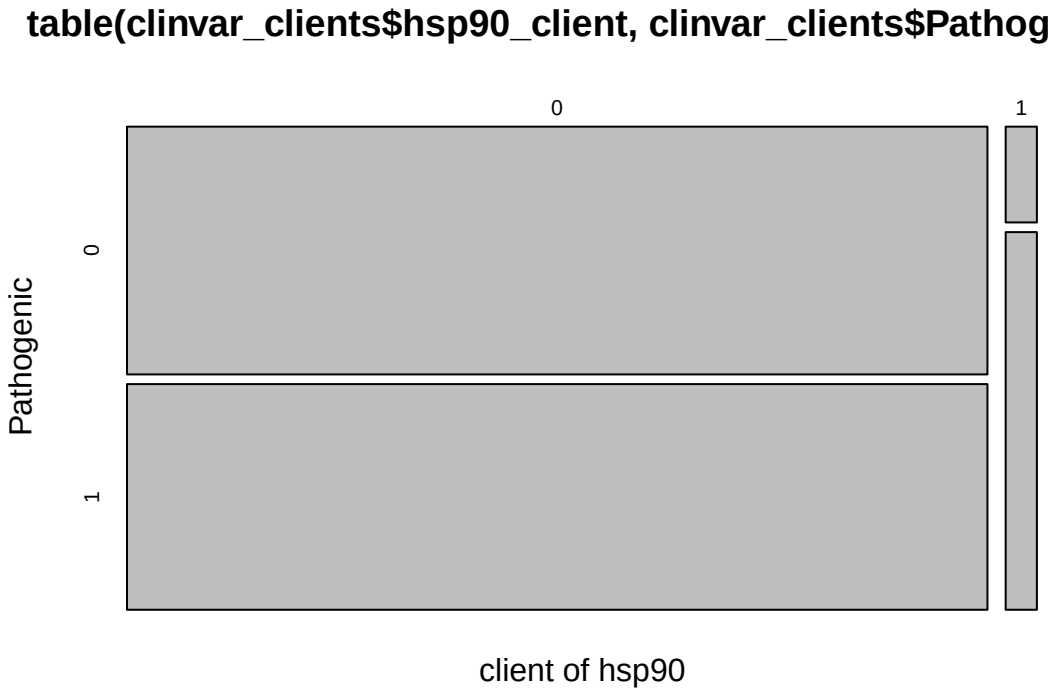
```
clinvar_ben$hsp90_client <- ifelse(clinvar_ben$Gene %in% clients_hsp90$Gene, 1,0)
clinvar_path$hsp90_client <- ifelse(clinvar_path$Gene %in% clients_hsp90$Gene, 1,0)

clinvar_ben$Pathogenic <- 0
clinvar_path$Pathogenic <- 1

clinvar_clients <- rbind(clinvar_ben, clinvar_path)
clinvar_clients <- clinvar_clients[,c(8,12,16,17)]
```

Compare clients with all others genes - Fisher

```
mosaicplot(table(clinvar_clients$hsp90_client, clinvar_clients$Pathogenic), xlab = 'client of hsp90', yl
```



```
ft <- fisher.test(table(clinvar_clients$hsp90_client, clinvar_clients$Pathogenic))
print(ft)
```

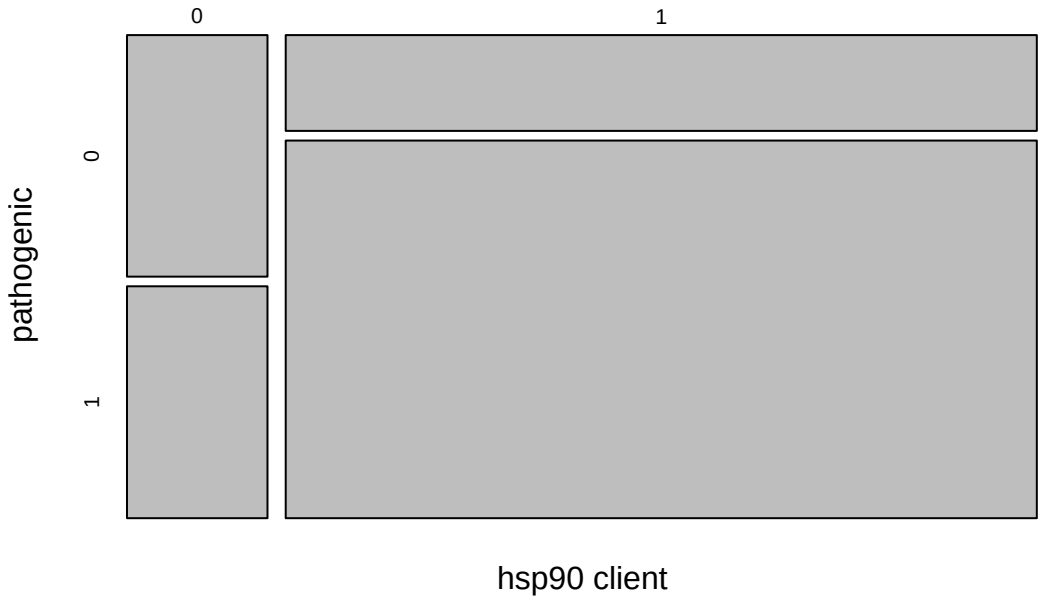
```
##
##  Fisher's Exact Test for Count Data
##
## data:  table(clinvar_clients$hsp90_client, clinvar_clients$Pathogenic)
## p-value < 2.2e-16
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  3.961509 4.734276
## sample estimates:
## odds ratio
##   4.328559
```

Compare clients with control nonclients genes - Fisher

```
clinvar_control_noncl <- clinvar_clients[(clinvar_clients$Gene %in% clients_hsp90$Gene) | (clinvar_clie


mosaicplot(table(clinvar_control_noncl$hsp90_client, clinvar_control_noncl$Pathogenic), xlab = 'hsp90 c
```

able(clinvar_control_noncl$hsp90_client, clinvar_control_noncl$Pathog




```
fisher.test(table(clinvar_control_noncl$hsp90_client, clinvar_control_noncl$Pathogenic))
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  table(clinvar_control_noncl$hsp90_client, clinvar_control_noncl$Pathogenic)
## p-value < 2.2e-16
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  3.401513 4.953585
```

```
## sample estimates:
## odds ratio
##   4.104137
```

How many htere are motations in one gene?

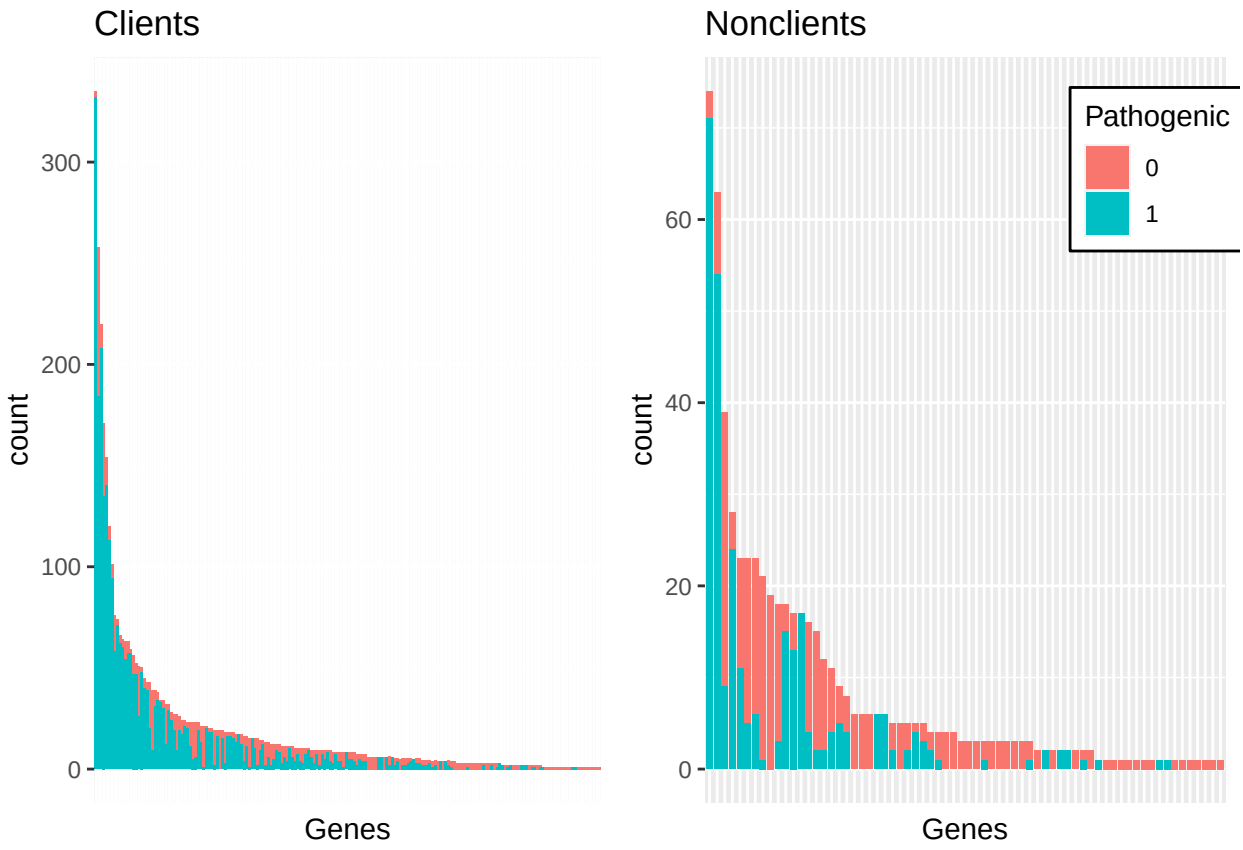
```
require(gridExtra)
```

```
## Loading required package: gridExtra
```

```
p1 <- ggplot(clinvar_control_noncl[,], aes(reorder(Gene, Gene, function(x)-length(x))))+
  geom_bar(aes(fill = as.factor(Pathogenic)))+
  #theme_bw()+
  xlab('Genes')+
  theme(axis.text.x=element_blank(), axis.ticks.x = element_blank(),
        legend.position = 'None')+
  ggtitle('Clients')

p2 <- ggplot(clinvar_control_noncl[clinvar_control_noncl$hsp90_client == 0,], aes(reorder(Gene, Gene, fu
  geom_bar(aes(fill = as.factor(Pathogenic)))+
  #theme_bw()+
  xlab('Genes')+
  theme(axis.text.x=element_blank(), axis.ticks.x = element_blank(), legend.position = c(0.87,0.85), leg
  ggtitle('Nonclients')+
  scale_fill_discrete(name = "Pathogenic")


grid.arrange(p1, p2, ncol=2)
```


Clients
Nonclients
count
Genes
Pathogenic
0
1
300
200
100
0
60
40
20
0