

istotype_ext_analysis.Rmd

```
library(data.table)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:data.table':
##
##   between, first, last
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(stringr)
library(ggplot2)
library(NMF)
```

```
## Loading required package: pkgmaker
## Loading required package: registry
##
## Attaching package: 'pkgmaker'
## The following object is masked from 'package:base':
##
##   isNamespaceLoaded
## Loading required package: rngtools
## Loading required package: cluster
## NMF - BioConductor layer [NO: missing Biobase] | Shared memory capabilities [NO: bigmemory] | Cores !
##   To enable the Bioconductor layer, try: install.extras('
## NMF
## ') [with Bioconductor repository enabled]
##   To enable shared memory capabilities, try: install.extras('
## NMF
## ')
```

```
library(reshape2)
```

```
##
## Attaching package: 'reshape2'
## The following objects are masked from 'package:data.table':
##
##   dcast, melt
```

```
library(parallel)
library(RColorBrewer)
library(scales)
```

```

summarise = dplyr::summarise

load("shm_rep1_reseq.rda")

df = shm %>%
  mutate(replacement = ifelse(as.character(from.aa) != as.character(to.aa), "replacement", "silent"),
         i = isotype) %>%
  group_by(clone, sample, proj, i, replacement) %>%
  summarise(count = n())

df.s = df %>%
  group_by(proj, i) %>%
  summarise(count = n())

print(df.s)

## # A tibble: 15 x 3
## # Groups:   proj [?]
##   proj i      count
##   <chr> <chr> <int>
## 1 old  IGHA1    968
## 2 old  IGHE     45
## 3 old  IGHG1   1094
## 4 old  IGHG2     2
## 5 old  IGHG3    55
## 6 old  IGHGP    13
## 7 old  IGHM   1413
## 8 young IGHA1   1876
## 9 young IGHD     34
## 10 young IGHE    41
## 11 young IGHG1  1355
## 12 young IGHG2     2
## 13 young IGHG3    82
## 14 young IGHGP    11
## 15 young IGHM   1622

df = df %>% filter(i != "IGHD", !is.na(i)) %>%
  mutate(isotype.full = i, isotype = str_sub(i, 4, 4))

df$proj = factor(df$proj, levels = c('young', 'old'))

dt.p = data.table()

for (iso in unique(df$isotype)) {
  tmp = df %>% filter(isotype == iso)
  x = (tmp %>% filter(proj == "old"))$count
  y = (tmp %>% filter(proj != "old"))$count
  kk = ks.test(x, y)
  p = kk$p.value
  dt.p = rbind(dt.p,
               data.table(isotype = iso, p=p))
}

## Warning in ks.test(x, y): p-value will be approximate in the presence of
## ties

```

```
## Warning in ks.test(x, y): p-value will be approximate in the presence of
## ties
```

```
## Warning in ks.test(x, y): cannot compute exact p-value with ties
```

```
## Warning in ks.test(x, y): p-value will be approximate in the presence of
## ties
```

```
dt.p$p.adj = p.adjust(dt.p$p, method = "BH")
print(dt.p %>% arrange(p.adj))
```

```
##      isotype      p      p.adj
## 1      M 0.0000365121 0.0001460484
## 2      G 0.0002211444 0.0004422888
## 3      A 0.0929452388 0.1239269851
## 4      E 0.2791116837 0.2791116837
```

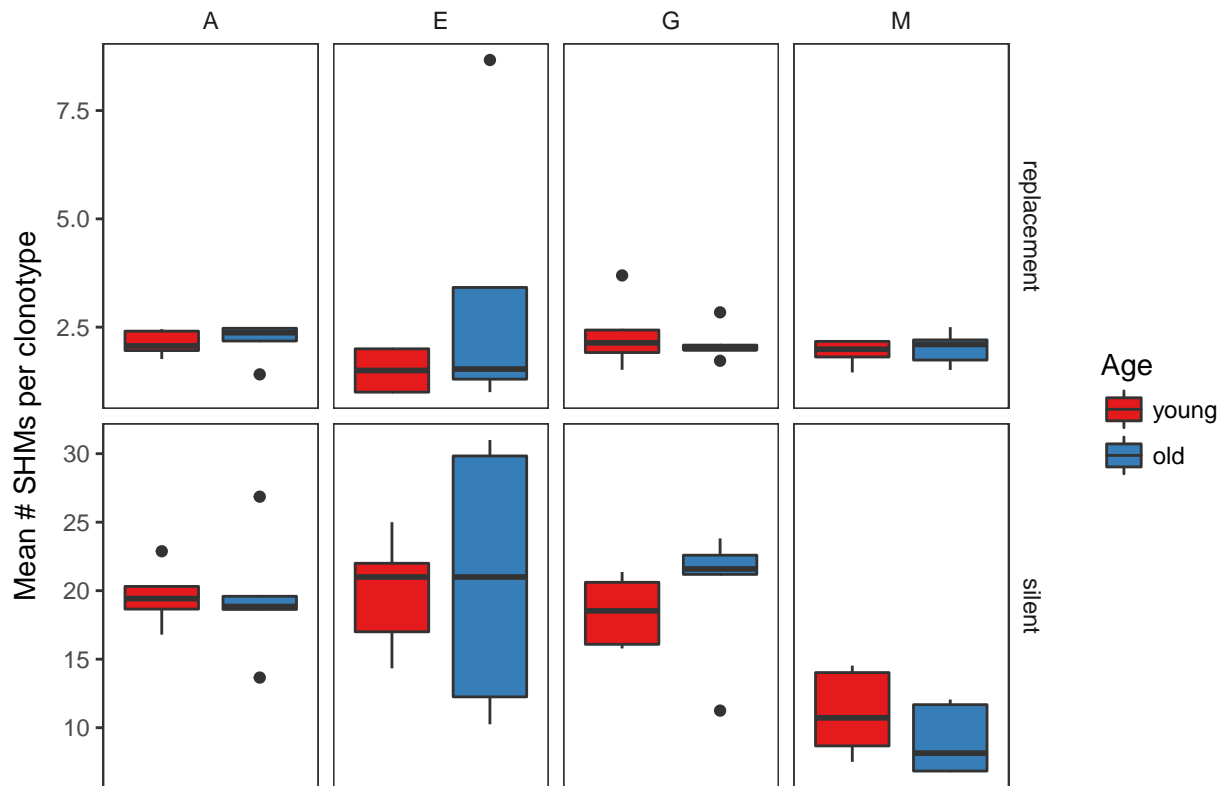
```
p13=ggplot(df, aes(x = count, fill = proj)) +
  geom_density(alpha = 0.9, color = NA) +
  facet_wrap(~isotype) +
  scale_fill_brewer("Age", palette = "Set1") +
  xlab("SHMs per clonotype") + ylab("") +
  theme_bw() +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        strip.background = element_blank())
ggsave("figures/p13.pdf", p13)
```

```
## Saving 6.5 x 4.5 in image
```

```
df.1 = df %>%
  group_by(sample, proj, replacement, isotype) %>%
  summarise(shms = mean(count))
```

```
p10=ggplot(df.1, aes(x=proj, fill = proj, y = shms)) +
  geom_boxplot() +
  facet_grid(replacement~isotype, scales = "free") +
  scale_fill_brewer("Age", palette = "Set1") +
  xlab("") + ylab("Mean # SHMs per clonotype") +
  theme_bw() +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        axis.text.x = element_blank(), axis.ticks.x = element_blank(),
        strip.background = element_blank())
```

```
p10
```



```
ggsave("figures/p10.pdf", p10)
```

```
## Saving 6.5 x 4.5 in image
```

```
a = aov(shms ~ replacement + isotype + proj, df.1)
summary(a)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## replacement  1   4479    4479 282.392 < 2e-16 ***
## isotype      3    378     126   7.944 0.00012 ***
## proj         1      1      1    0.066 0.79869
## Residuals   72   1142      16
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
TukeyHSD(a, "proj")
```

```
##      Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = shms ~ replacement + isotype + proj, data = df.1)
##
## $proj
##           diff          lwr          upr        p adj
## old-young 0.2308599 -1.566933 2.028652 0.7986911
```

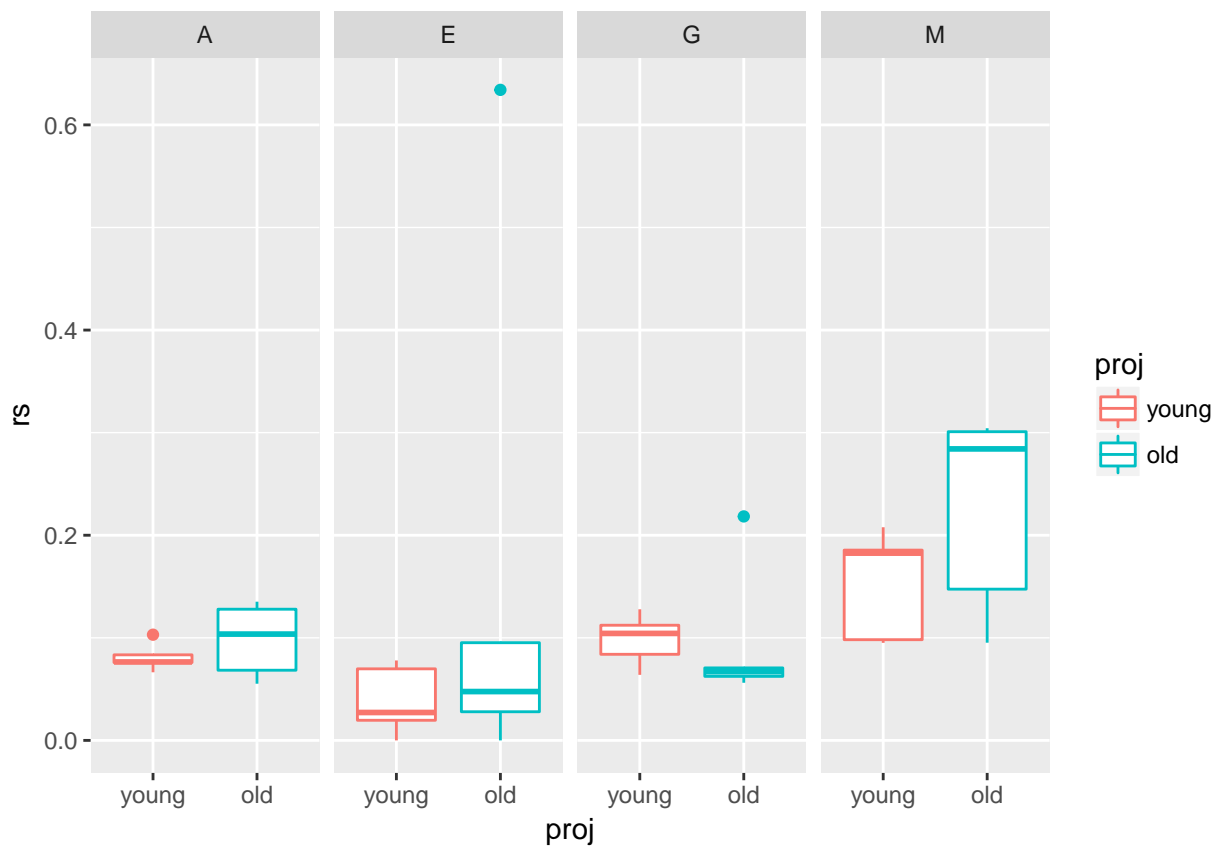
```
TukeyHSD(a, "isotype")
```

```
##      Tukey multiple comparisons of means
##      95% family-wise confidence level
##
```

```
## Fit: aov(formula = shms ~ replacement + isotype + proj, data = df.1)
##
## $isotype
##          diff          lwr          upr          p adj
## E-A  0.6516117 -2.751334  4.054557  0.9579676
## G-A -0.1053994 -3.417583  3.206785  0.9997878
## M-A -4.8323319 -8.144516 -1.520148  0.0014808
## G-E -0.7570110 -4.159957  2.645935  0.9362752
## M-E -5.4839436 -8.886889 -2.080998  0.0003764
## M-G -4.7269325 -8.039117 -1.414749  0.0019491
```

```
df.2 = df %>%
  group_by(sample, proj, isotype) %>%
  summarise(rs = sum(count[which(replacement == "replacement")]) / sum(count[which(replacement != "replacement")]))

ggplot(df.2, aes(x=proj, color = proj, y = rs)) +
  geom_boxplot() +
  facet_grid(.~isotype, scales = "free")
```



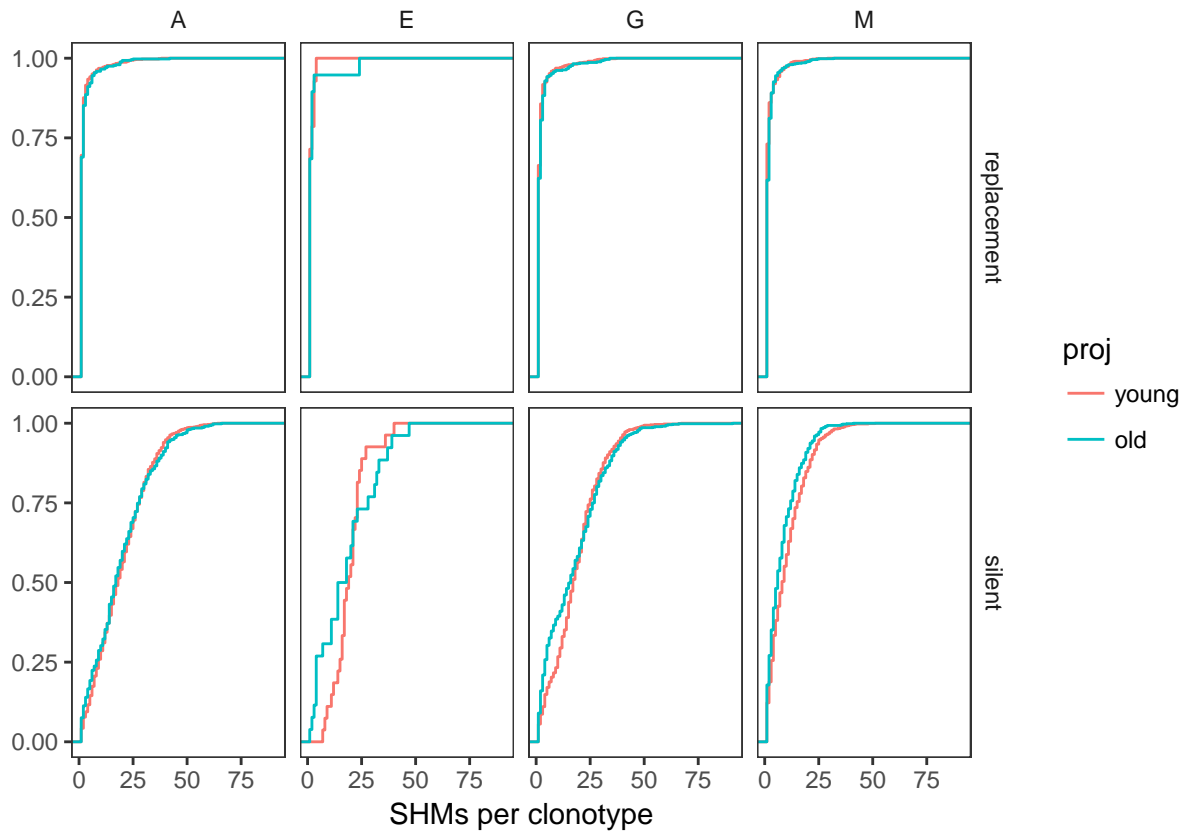
```
a = aov(rs ~ isotype + proj, df.2)
summary(a)
```

```
##          Df Sum Sq Mean Sq F value Pr(>F)
## isotype    3  0.0679  0.02262    2.028  0.128
## proj        1  0.0271  0.02709    2.429  0.128
## Residuals   35  0.3904  0.01115
```

```
TukeyHSD(a, "proj")
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = rs ~ isotype + proj, data = df.2)
##
## $proj
##              diff              lwr              upr              p adj
## old-young 0.05204729 -0.0157523 0.1198469 0.1281264
```

```
ggplot(df, aes(x=count, color = proj)) +
  stat_ecdf() +
  facet_grid(replacement~isotype) + #, scales = "free") +
  #scale_fill_brewer("Age", palette = "Set1") +
  xlab("SHMs per clonotype") + ylab("") +
  theme_bw() +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        strip.background = element_blank())
```



```
dt.p = data.table()

for (iso in unique(df$isotype)) {
  for (rr in unique(df$replacement)) {
    tmp = df %>% filter(isotype == iso, replacement == rr)
    x = (tmp %>% filter(proj == "old"))$count
    y = (tmp %>% filter(proj != "old"))$count
    kk = ks.test(x, y)
```

```

p = kk$p.value
dt.p = rbind(dt.p,
             data.table(isotype = iso, replacement = rr, p=p))
}
}

## Warning in ks.test(x, y): p-value will be approximate in the presence of
## ties

## Warning in ks.test(x, y): p-value will be approximate in the presence of
## ties

## Warning in ks.test(x, y): p-value will be approximate in the presence of
## ties

## Warning in ks.test(x, y): p-value will be approximate in the presence of
## ties

## Warning in ks.test(x, y): cannot compute exact p-value with ties

## Warning in ks.test(x, y): cannot compute exact p-value with ties

## Warning in ks.test(x, y): p-value will be approximate in the presence of
## ties

## Warning in ks.test(x, y): p-value will be approximate in the presence of
## ties

dt.p$p.adj = p.adjust(dt.p$p, method = "BH")
print(dt.p %>% arrange(p.adj))

##   isotype replacement          p          p.adj
## 1      G      silent 1.173632e-07 9.389059e-07
## 2      M      silent 4.082755e-06 1.633102e-05
## 3      M replacement 2.310786e-04 6.162096e-04
## 4      E      silent 2.584502e-01 5.169004e-01
## 5      A      silent 3.277368e-01 5.243789e-01
## 6      G replacement 4.454430e-01 5.939240e-01
## 7      A replacement 9.673522e-01 9.999793e-01
## 8      E replacement 9.999793e-01 9.999793e-01

```