UCB

```
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(ggplot2)
library(reshape2)
```

```
##
## Attaching package: 'reshape2'
```

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

```
library(scales)
library(parallel)
library(stringr)
library(knitr)
```

```
dt.aging.stats = fread("annotations/aging_stats.txt") %>%
  filter(age > 0, !is.na(sex)) %>%
  mutate(count_total = count, occurrences_total = diversity) %>%
  select(sample_id, sex, count_total, occurrences_total)
```

Load VDJdb annotations with 1 mismatch for aging data

```
dt.aging = rbindlist(mclapply(as.list(dt.aging.stats$sample_id),
                   function(x) fread(paste0("annotations/aging_split_1mm/", x, ".annot.txt")) %>%
                     mutate(sample_id = x), mc.cores = 40)) %>%
  group_by(sample_id, cdr3) %>%
  summarise(count = sum(count), occurrences = n())
```

VDJdb data

```
dt.vdjdb = fread("rearr_model/VDJDB_fullP_rob_ageing.txt") %>%
  filter(gene == "TRB", mhc.class == "MHCI") %>%
  mutate(hla_spec = str_split_fixed(mhc.a, pattern = "[:,]", 2)[,1]) %>%
  select(cdr3, hla_spec, antigen.epitope, antigen.species) %>%
  group_by(antigen.epitope) %>%
  mutate(unique_cdrs = n()) %>%
  filter(unique_cdrs > 30) %>%
```

```
  select(cdr3, hla_spec, antigen.epitope, antigen.species, unique_cdrs)
```

Merge datasets

```
dt.aging.m = dt.aging %>%
  merge(dt.vdjdb) %>%
  merge(dt.aging.stats)
```

Summarise by epitope

```
dt.aging.s = dt.aging.m %>%
  group_by(sex, antigen.epitope, antigen.species, unique_cdrs) %>%
  summarise(occurrences = sum(occurrences) / unique_cdrs[1]) %>%
  merge(dt.aging.stats %>%
          group_by(sex) %>%
          summarise(occurrences_total = sum(as.numeric(occurrences_total))))
```

Evenness of epitope-specific TCR occurrences

```
dt.aging.s2 = as.data.table(dt.aging.m) %>%
  group_by(sample_id, sex, antigen.epitope, antigen.species, unique_cdrs) %>%
  summarise(occurrences = sum(occurrences)) %>%
  group_by(sample_id) %>%
  mutate(occurrences_share = occurrences / sum(occurrences) / unique_cdrs)
```

```
dt.p = data.table(antigen.epitope = unique(dt.aging.s2$antigen.epitope), p = 1, freq.ratio = 1) %>%
  merge(dt.aging.s2 %>% ungroup %>% select(antigen.species, antigen.epitope) %>% unique)

for (i in 1:nrow(dt.p)) {
  tmp = dt.aging.s2 %>% filter(antigen.epitope == dt.p$antigen.epitope[i])
  dt.p$freq.ratio[i] = with(tmp, mean(occurrences_share[which(sex == "M")]) / mean(occurrences_share[wh
  dt.p$p[i] = t.test(occurrences_share ~ sex, tmp)$p.value
}

dt.p$p = p.adjust(dt.p$p, method = "BH")
dt.p$len = nchar(as.character(dt.p$antigen.epitope))

kable(dt.p %>% arrange(p))
```

antigen.epitope	p	freq.ratio	antigen.species	len
ATDALMTGY	0.4643940	1.0594762	HCV	9
EIYKRWII	0.4643940	1.1539361	HIV-1	8
GILGFVFTL	0.4643940	0.9711087	InfluenzaA	9
LPRRSGAAGA	0.4643940	0.9658870	InfluenzaA	10
RAKFKQLL	0.4643940	1.0376307	EBV	8
RPRGEVRFL	0.4643940	0.9000810	HSV-2	9
VTEHDTLLY	0.4643940	0.9624216	CMV	9
KLVALGINAV	0.5758008	1.0423553	HCV	10
AMFWSVPTV	0.5813778	1.0584902	HomoSapiens	9
LLWNGPMAV	0.6627509	1.0349229	YellowFeverVirus	9
ELAGIGILTV	0.8028196	1.0090034	HomoSapiens	10
FLYNLLTRV	0.8028196	0.9383875	HomoSapiens	9
IIKDYGKQM	0.8028196	1.0840442	HIV-1	9

antigen.epitope	p	freq.ratio	antigen.species	len
YVLDHLIVV	0.8299260	0.9593484	EBV	9
EPLPQGQLTAY	0.8673266	1.0672729	EBV	11
FLKEKGGL	0.8673266	1.0315159	HIV-1	8
FPRPWLHGL	0.8673266	0.9761859	HIV-1	9
GPGHKARVL	0.8673266	1.0243763	HIV-1	9
HPKVSSEVHI	0.8673266	0.9599714	HIV-1	10
HSKKKCDEL	0.8673266	1.0682158	HCV	9
IPSINVHHY	0.8673266	1.0294460	CMV	9
KRWIILGLNK	0.8673266	1.0154329	HIV-1	10
KRWIIMGLNK	0.8673266	0.9814059	HIV-1	10
LPPIVAKEI	0.8673266	1.0366933	HIV-1	9
TPGPGVRYPL	0.8673266	0.9823725	HIV-1	10
TPQDLNTML	0.8673266	0.9795780	HIV-1	9
TPRVTGGGAM	0.8797360	1.0249844	CMV	10
GLCTLVAML	0.9043370	0.9974496	EBV	9
ISPRTLNAW	0.9043370	0.9931273	HIV-1	9
KAFSPEVIPMF	0.9043370	0.9945367	HIV-1	11
LLLGIGILV	0.9043370	0.9927336	HomoSapiens	9
NLVPMVATV	0.9043370	0.9955132	CMV	9
SLYNTVATL	0.9043370	0.9925401	HIV-1	9
CINGVCWTV	0.9135497	0.9932904	HCV	9

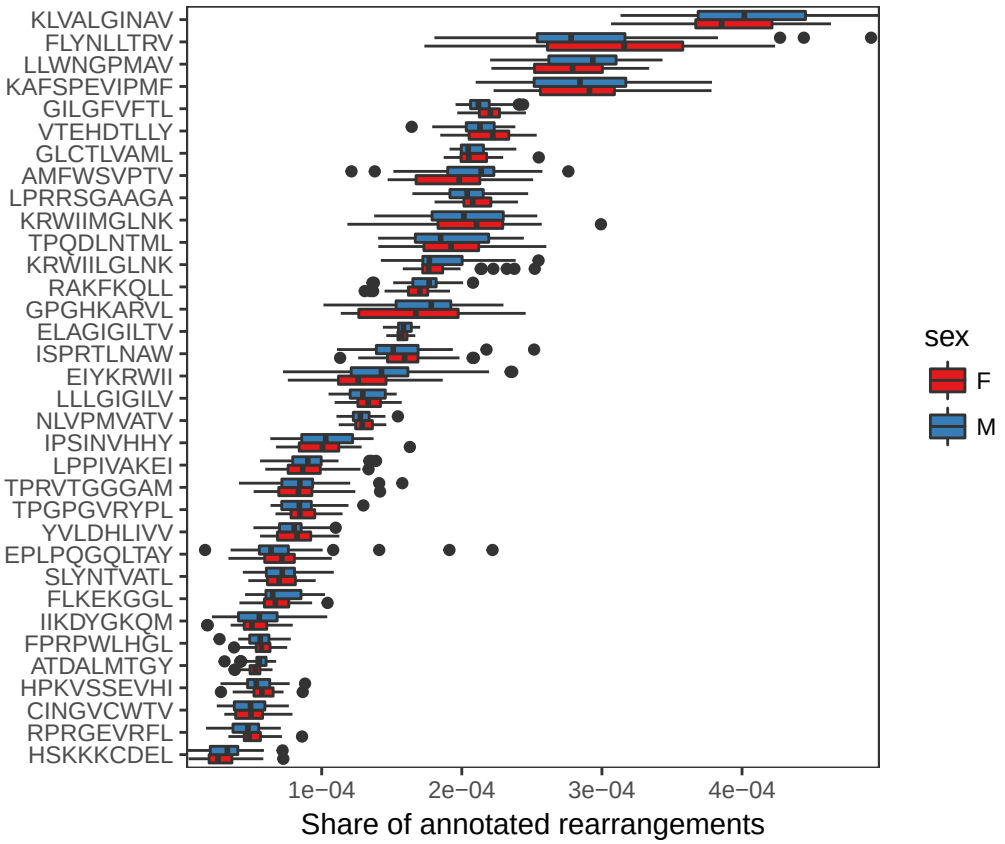
```
good_epi = (dt.p %>% filter(p < 0.05))$antigen.epitope
dt.aging.s2 = dt.aging.s2 %>%
  mutate(antigen.epitope = ifelse(antigen.epitope %in% good_epi, paste(antigen.epitope, "(*)"),antigen.
```

```
dt.aging.s2.s = dt.aging.s2 %>%
  group_by(antigen.epitope) %>%
  summarise(freq = mean(occurrences_share[which(sex == "M")]))

dt.aging.s2$antigen.epitope = factor(dt.aging.s2$antigen.epitope,
                                     levels = dt.aging.s2.s$antigen.epitope[order(dt.aging.s2.s$freq)])

dt.aging.s2$epi.len = nchar(as.character(dt.aging.s2$antigen.epitope))

ggplot(dt.aging.s2, aes(x = antigen.epitope, group = paste(antigen.epitope,sex),
                        fill = sex,
                       y = occurrences_share)) +
  geom_boxplot() +
  coord_flip() +
  scale_fill_brewer(palette = "Set1") +
  xlab("") + scale_y_continuous("Share of annotated rearrangements",
                                expand = c(0,0)) +
  theme_bw()  +
  theme(aspect = 1.1,
        panel.grid.major = element_blank(), panel.grid.minor = element_blank())
```


KLVALGINAV
FLYNLLTRV
LLWNGPMAV
KAFSPEVIPMF
GILGFVFTL
VTEHDTLLY
GLCTLVAML
AMFWSVPTV
LPRRSGAAGA
KRWIIMGLNK
TPQDLNTML
KRWIILGLNK
RAKFKQLL
GPGHKARVL
ELAGIGILTV
ISPRTLNAW
EIYKRWII
LLLGIGILV
NLVPMVATV
IPSINVHHY
LPPIVAKEI
TPRVTGGGAM
TPGPGVRYPL
YVLDHLIVV
EPLPQGQLTAY
SLYNTVATL
FLKEKGGL
IIKDYGKQM
FPRPWLHGL
ATDALMTGY
HPKVSSEVHI
CINGVCWTV
RPRGEVRFL
HSKKKCDEL
1e−04
2e−04
3e−04
4e−04
Share of annotated rearrangements
sex
F
M