Sex

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

```
## ------------------------------------------------------------------------
```

```
## data.table + dplyr code now lives in dtplyr.
## Please library(dtplyr)!
```

```
## ------------------------------------------------------------------------
```

```
##
## 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)
```

Load metadata

```
dt.hip.stats = fread("annotations/hip_stats.txt") %>%
  filter(!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 HIP data (time consuming, ~ 2mln clonotypes)

```
dt.hip = rbindlist(mclapply(as.list(dt.hip.stats$sample_id),
                   function(x) fread(paste0("annotations/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") %>%
  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)
```

Merge

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

Summarise by sex

```
dt.hip.s = dt.hip.m %>%
  group_by(sex, hla_spec, antigen.epitope, antigen.species) %>%
  summarise(occurrences = sum(occurrences), occurrences_total = sum(as.numeric(occurrences_total)))
```

Plot

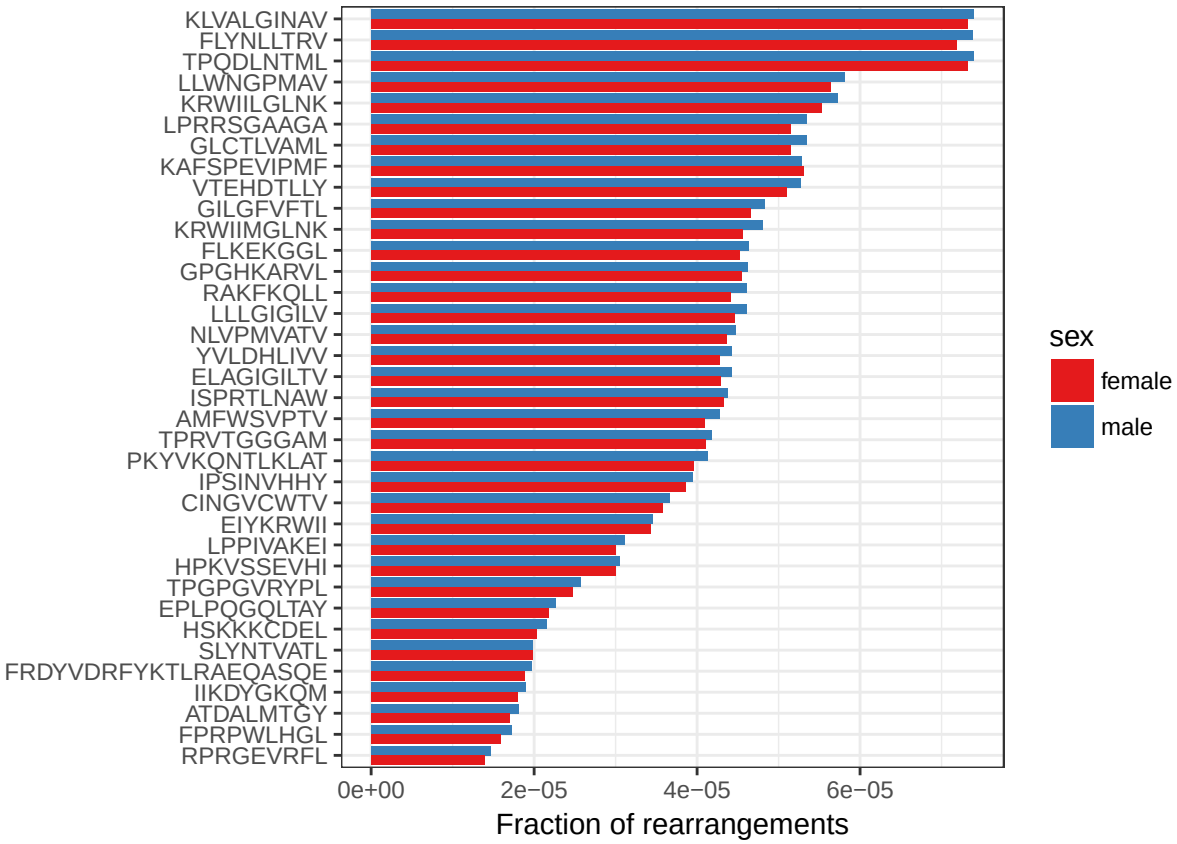
```
dt.hip.s.s = dt.hip.s %>%
  filter(sex == "male") %>%
  group_by(antigen.epitope) %>%
  summarise(freq = sum(occurrences) / sum(occurrences_total))

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

tmp = dt.hip.s %>%
  group_by(sex, antigen.epitope) %>%
  summarise(freq = sum(occurrences) / sum(occurrences_total)) %>%
  dcast(antigen.epitope~sex, value.var= "freq")
freq.ratios = tmp[,3] / tmp[,2]
m=mean(freq.ratios)
ci = qnorm(0.975)*sd(freq.ratios)/sqrt(length(freq.ratios))
paste(round(m,2), round(m-ci,2), round(m+ci,2))
```

```
## [1] "1.03 1.03 1.04"
```

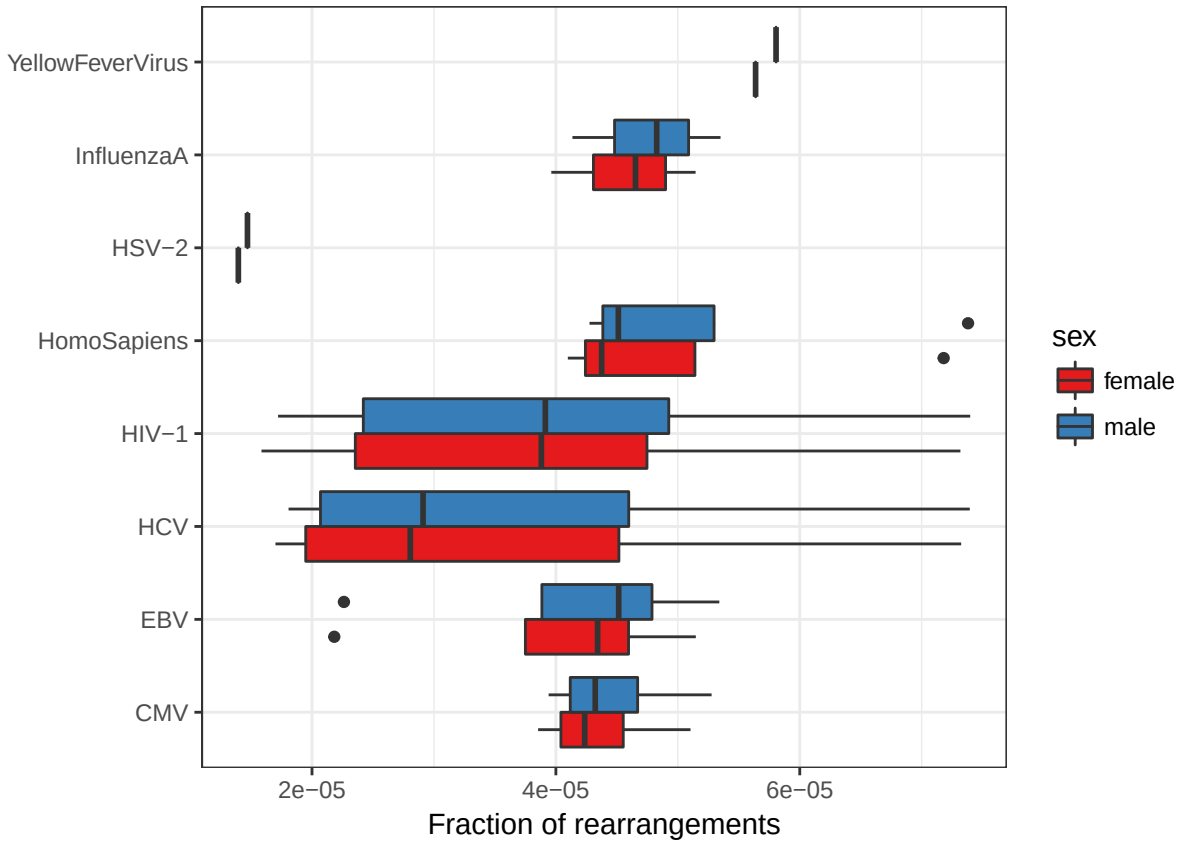
```
wilcox.test(occurrences / occurrences_total ~ sex, dt.hip.s, paired=T)
```

```
##
##  Wilcoxon signed rank test
##
## data:  occurrences/occurrences_total by sex
## V = 3, p-value = 7.276e-11
## alternative hypothesis: true location shift is not equal to 0
```

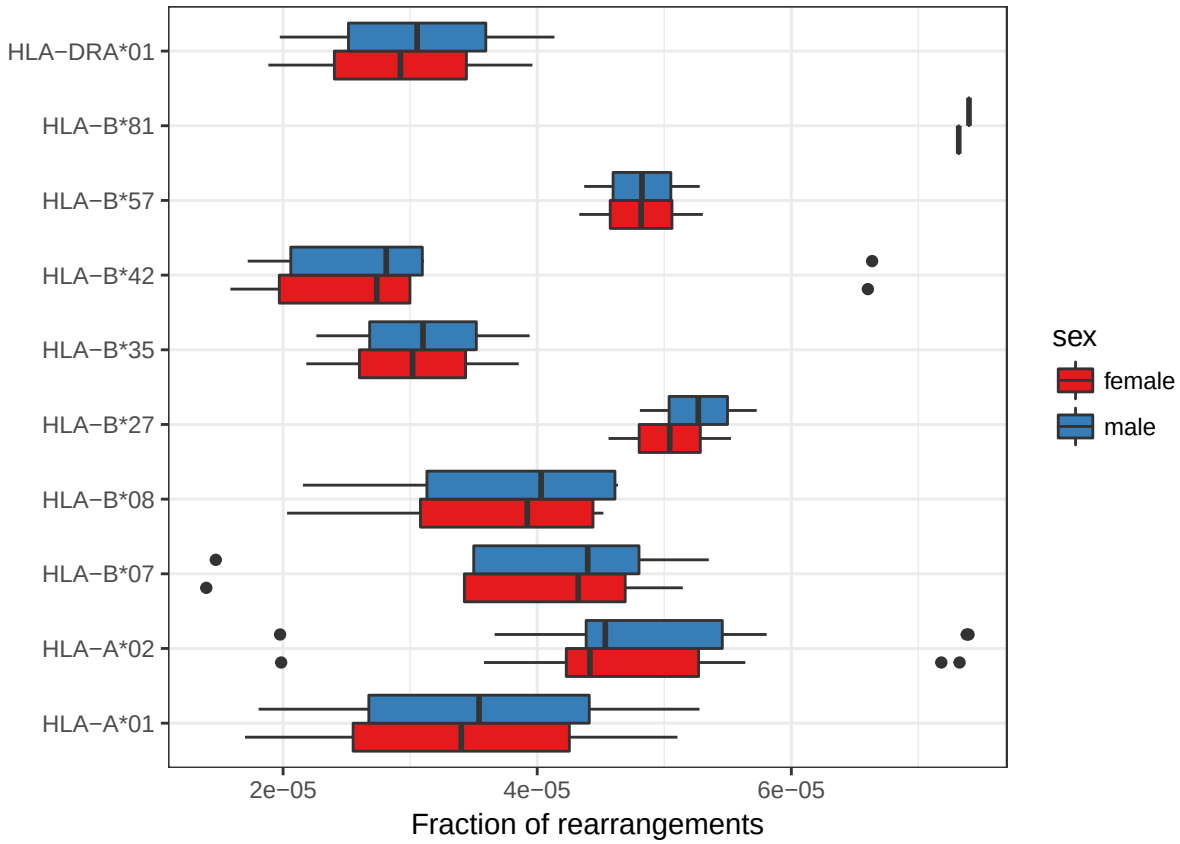
```
ggplot(dt.hip.s, aes(x = antigen.epitope, fill = sex, y = occurrences / occurrences_total)) +
  geom_bar(stat="identity", position = "dodge") +
  coord_flip() +
  scale_fill_brewer(palette = "Set1") +
  xlab("") + ylab("Fraction of rearrangements") +
  theme_bw()
```




```
ggplot(dt.hip.s, aes(x = antigen.species, fill = sex, y = occurrences / occurrences_total)) +
  geom_boxplot(aes(group = paste(antigen.species, sex))) +
  coord_flip() +
  scale_fill_brewer(palette = "Set1") +
  xlab("") + ylab("Fraction of rearrangements") +
  theme_bw()
```




```
ggplot(dt.hip.s, aes(x = hla_spec, fill = sex, y = occurrences / occurrences_total)) +
  geom_boxplot(aes(group = paste(hla_spec, sex))) +
  coord_flip() +
  scale_fill_brewer(palette = "Set1") +
  xlab("") + ylab("Fraction of rearrangements") +
  theme_bw()
```


HLA−DRA*01
HLA−B*81
HLA−B*57
HLA−B*42
HLA−B*35
HLA−B*27
HLA−B*08
HLA−B*07
HLA−A*02
HLA−A*01
2e−05
4e−05
6e−05
Fraction of rearrangements
sex
female
male