

T-cell survival: 3 year dynamic and YF data

Anastasia Pavlova

2022-10-12

```
meta.a <- read_tsv("../data/aging_3year/metadata.txt")

## Rows: 4 Columns: 3
## -- Column specification -----
## Delimiter: "\t"
## chr (2): file.name, donor
## dbl (1): age
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
meta.a

## # A tibble: 4 x 3
##   file.name      donor    age
##   <chr>          <chr>  <dbl>
## 1 A6-I200ob.txt.gz donor1    27
## 2 A6-I201ob.txt.gz donor1    30
## 3 A6-I202ob.txt.gz donor2    47
## 4 A5-S23.txt.gz   donor2    50

data.a <- meta.a %>%
  group_by(donor, age) %>%
  group_modify(~read_tsv(paste0("../data/aging_3year/", .x$file.name))) %>%
  ungroup %>%
  mutate(tcrkey = paste(v, cdr3nt)) %>%
  group_by(donor, age, tcrkey) %>%
  summarise(count = sum(count)) %>%
  ungroup

## Rows: 1812062 Columns: 11
## -- Column specification -----
## Delimiter: "\t"
## chr (5): cdr3nt, cdr3aa, v, d, j
## dbl (6): count, freq, VEnd, DStart, DEnd, JStart
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 758985 Columns: 11
## -- Column specification -----
## Delimiter: "\t"
## chr (5): cdr3nt, cdr3aa, v, d, j
## dbl (6): count, freq, VEnd, DStart, DEnd, JStart
##
```

```

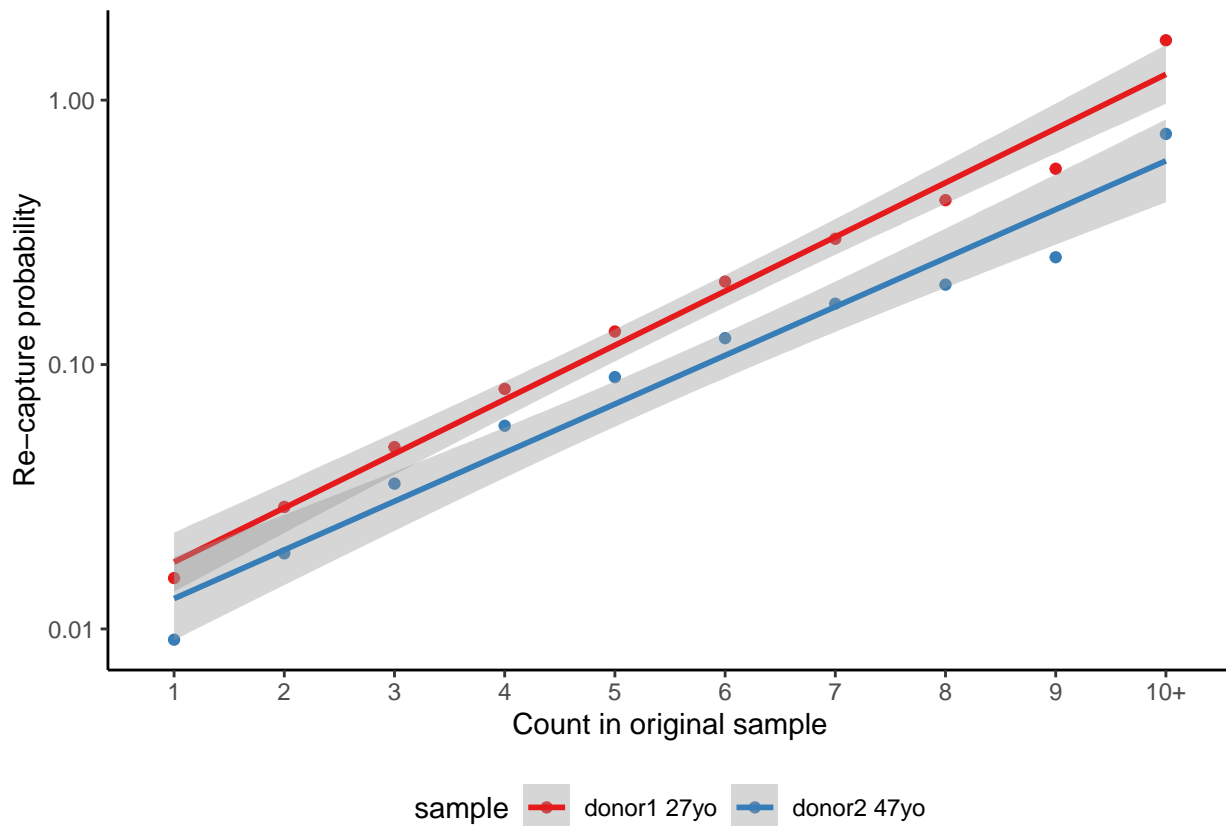
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 954170 Columns: 11
## -- Column specification -----
## Delimiter: "\t"
## chr (5): cdr3nt, cdr3aa, v, d, j
## dbl (6): count, freq, VEnd, DStart, DEnd, JStart
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 137624 Columns: 11
## -- Column specification -----
## Delimiter: "\t"
## chr (5): cdr3nt, cdr3aa, v, d, j
## dbl (6): count, freq, VEnd, DStart, DEnd, JStart
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## `summarise()` has grouped output by 'donor', 'age'. You can override using the `.groups` argument.
data.a.m <- data.a %>%
  filter(age %in% c(27, 47)) %>%
  left_join(data.a %>%
    filter(age %in% c(30, 50)) %>%
    select(donor, tcrkey) %>%
    unique %>%
    mutate(overlapping = T)) %>%
  mutate(overlapping = !is.na(overlapping))

## Joining with `by = join_by(donor, tcrkey)`
data.a.s <- data.a.m %>%
  mutate(count = pmin(10, count)) %>%
  group_by(sample = paste0(donor, " ", age, "yo"), count) %>%
  summarise(alpha = sum(overlapping),
    beta = sum(!overlapping))

## `summarise()` has grouped output by 'sample'. You can override using the
## `.groups` argument.
p1 <- data.a.s %>%
  ggplot(aes(x = count, y = alpha / beta,
    color = sample)) +
  geom_point(aes(x = as.factor(
    ifelse(count > 9, "10+", count) %>%
    fct_reorder(count)
  ))) +
  geom_smooth(method = "lm") +
  scale_color_brewer(palette = "Set1") +
  scale_y_log10("Re-capture probability") +
  xlab("Count in original sample") +
  theme_classic() +
  theme(legend.position = "bottom")
p1

## `geom_smooth()` using formula = 'y ~ x'

```



```
meta.v <- read_tsv("../data/yellow_fever_vac/metadata.txt") %>%
  filter(time %in% c(-1, 0))

data.v <- meta.v %>%
  filter(replica == "F1") %>%
  group_by(donor, time, replica) %>%
  group_modify(~ read_tsv(paste0("../data/yellow_fever_vac/", .x$file.name))) %>%
  mutate(cdr3 = `N. Seq. CDR3`,
         v = str_split_fixed(`All V hits`, fixed("*"), 2)[,1],
         count = `Clone count`) %>%
  group_by(donor, time, replica, cdr3, v) %>%
  summarise(count = sum(count)) %>%
  ungroup

## Warning: One or more parsing issues, call `problems()` on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)
## One or more parsing issues, call `problems()` on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)

data.v <- data.v %>%
  mutate(quantile = case_when(
    count == 1 ~ "Singleton",
    count == 2 ~ "Doubleton",
    count == 3 ~ "Tripleton",
```

```

    T ~ "Large"
  )) %>%
  mutate(quantile = factor(quantile,
                           c("Singleton",
                              "Doubleton", "Tripletion",
                              "Large"))) %>%

  select(-count)

data.v.m <- left_join(
  data.v %>%
    filter(time == -1) %>%
    select(-time, -replica),
  data.v %>%
    filter(time == 0) %>%
    mutate(found = T) %>%
    select(-quantile, -time, -replica),
  by = c("donor", "cdr3", "v")
) %>%
  mutate(found = !is.na(found))

data.v.s <- data.v.m %>%
  group_by(donor, quantile) %>%
  summarise(alpha = sum(found),
            beta = sum(!found)) %>%
  left_join(data.v %>%
    filter(time == -1) %>%
    group_by(donor) %>%
    summarise(clonotypes.pre = length(unique(paste(v, cdr3)))) %>%
  left_join(data.v %>%
    filter(time == 0) %>%
    group_by(donor) %>%
    summarise(clonotypes.post = length(unique(paste(v, cdr3))))

## `summarise()` has grouped output by 'donor'. You can override using the
## `.groups` argument.
## Joining with `by = join_by(donor)`
## Joining with `by = join_by(donor)`

p2 <- data.v.s %>%
  ggplot(aes(x = clonotypes.post / clonotypes.pre,
            y = alpha / (alpha + beta),
            group = quantile,
            color = quantile)) +
  geom_smooth(method = "lm", aes(), size = 1) +
  geom_point(size = 5) +
  geom_text(aes(label = donor), size = 2, color = "white") +
  scale_x_log10("Clones at day 0 / clones at day -7") +
  scale_y_log10("Capture probability") +
  scale_color_brewer("", palette = "Set2") +
  theme_classic() +
  theme(legend.position = "bottom")

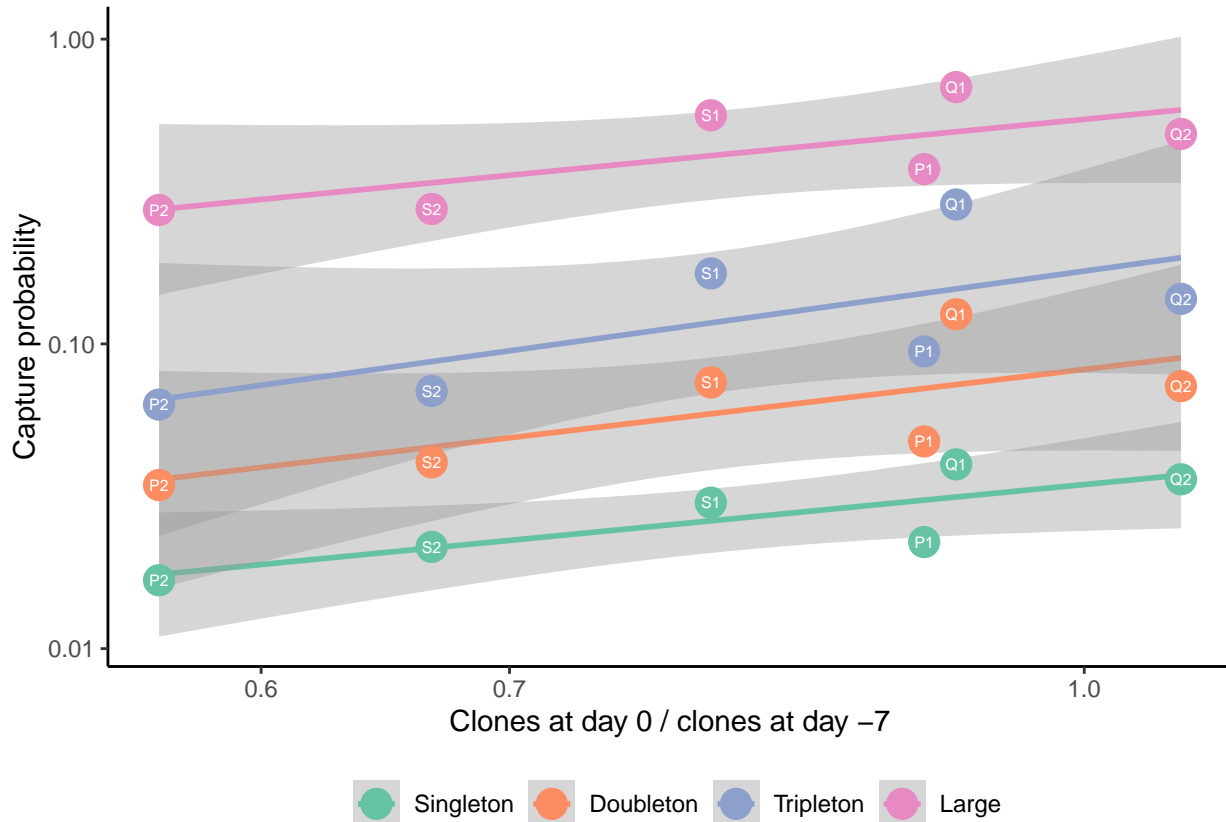
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.

```

```
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
p2
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



```
mdl <- lm(LogRecaptureProb ~ Quantile + LogDiversityRatio,
  data.v.s %>%
    mutate(LogRecaptureProb = log10(alpha / (alpha + beta)),
           LogDiversityRatio = log10(clonotypes.post / clonotypes.pre),
           Quantile = quantile))
summary(mdl)
```

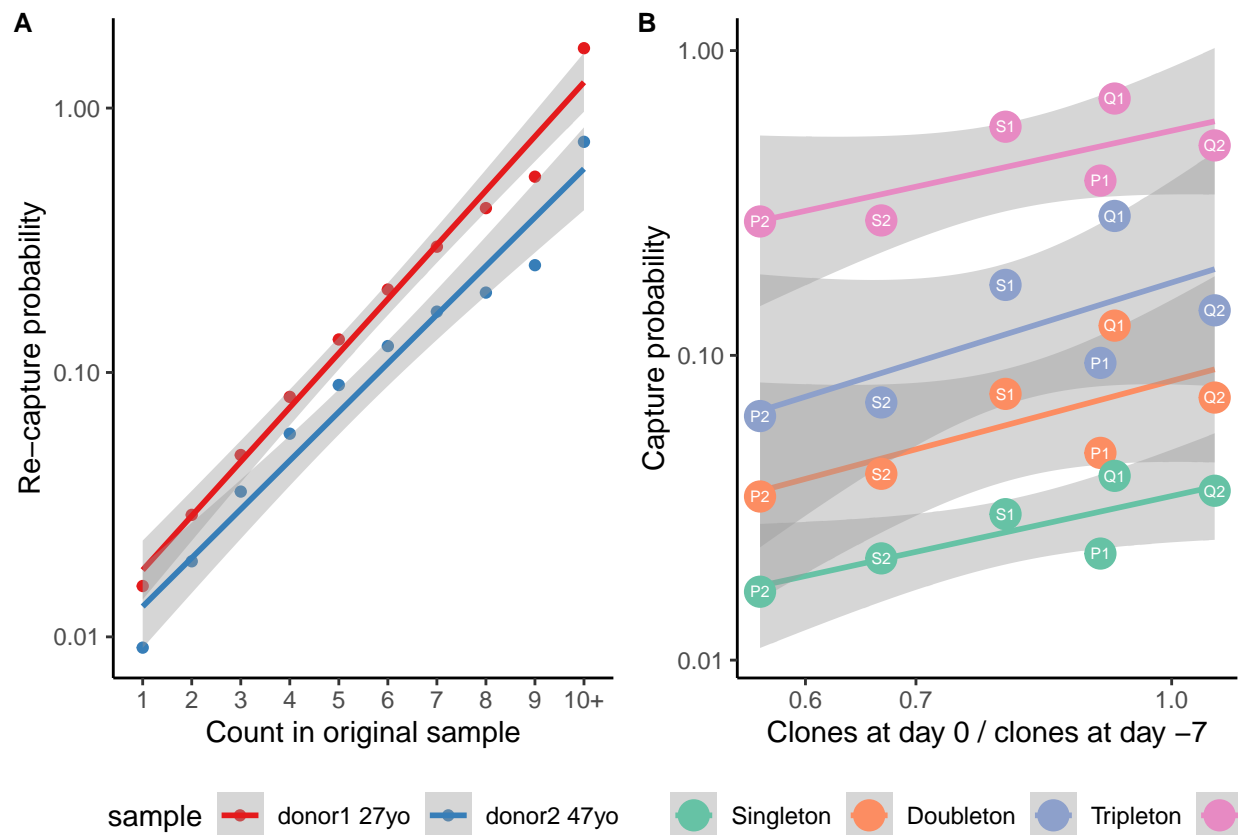
```
##
## Call:
## lm(formula = LogRecaptureProb ~ Quantile + LogDiversityRatio,
##     data = data.v.s %>% mutate(LogRecaptureProb = log10(alpha/(alpha +
##         beta)), LogDiversityRatio = log10(clonotypes.post/clonotypes.pre),
##         Quantile = quantile))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.17411 -0.09833 -0.03318  0.09707  0.29610
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -1.44294    0.06522  -22.123 5.05e-15 ***
## QuantileDoubleton  0.35227    0.08184   4.304 0.000382 ***
```

```
## QuantileTripletion 0.65180 0.08184 7.964 1.79e-07 ***
## QuantileLarge 1.19834 0.08184 14.643 8.40e-12 ***
## LogDiversityRatio 1.37663 0.31264 4.403 0.000305 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1417 on 19 degrees of freedom
## Multiple R-squared: 0.9294, Adjusted R-squared: 0.9145
## F-statistic: 62.5 on 4 and 19 DF, p-value: 1.142e-10
```

```
plot_grid(p1, p2,
  labels = c("A", "B"),
  label_size = 10) -> fig3
```

```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
```

```
fig3
```



```
ggsave("fig3.pdf", fig3)
```

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