Write-up

Experimental protocol for cytometry readings

Malaria parasite culture was first synchronized via sorbitol prior to preparation of each well plate. Each sample was given one of three different salivary treatments (male saliva, or female saliva, or PBS control), as well as one of two hematocrit treatments (0.5% or 1%); parasitemia was also the same across all wells/plates. Five different replicates were made for each treatment combination, resulting in 30 samples prepared for each well plate. Remaining wells on plate were either filled with water or left empty (if on the edge). Four well plates were prepared in total, each created and measured on different days.

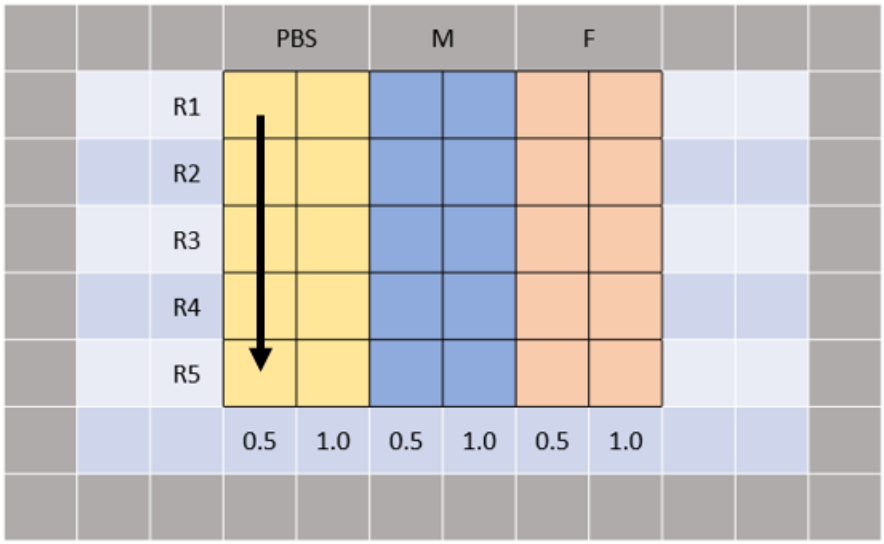


Figure 1: Well plate set up: five replicates (rows 1-5) are made for each combination of hematocrit treatments (0.5% and 1.0%) and salivary treatment (Female, Male, and PBS control). Replicate samples were generally filled in descending order (see arrow). Remaining non-edge wells are filled with water.

After two days (to allow for replication and investment), cell counts were taken via flow cytometry for each well plate. Samples were dyed to separate uninfected RBCs, asexual iRBCs, and gametocytes. First gate of cytometry results gated for singlets, second separates uRBCs from iRBCs (`iRBCs Count`, includes sexual and asexual), and the final gating separates uRBCs, asexual RBCs, and gametocytes (`iRBC_tdT+ Count`). We also calculate a proportion of cells commited to gametocytes (`Commit`) by dividing counts of gametocytes by counts of total iRBCs.

KEY

Key for variable names (in code).

- `iRBC_tdT+ Count` & `iRBCs Count`: Counts of gametocytes and total infected (asexual or gams) RBCs respectively, continuous. **This is the primary response variable.**
- `Treatment`: Salivary treatment, i.e. treatment of PBS control or male or female saliva (categorical, three levels: `PBS, F, M`). **This is the primary treatment of interest.**
- `Hematocrit`: Hematocrit treatment, i.e. 0.5% or 1.0% (*stored as categorical*, two levels: `0.5, 1.0`).

- `Plate` (ID): Which plate sample was from (categorical, four levels: `SGE-1, SGE-2, SGE-3, SGE-4`).
- `Replicate` (ID): Which replicate (row) of plate sample was from (categorical, five levels: `R1, R2, R3, R4, R5`)

RUNNING MODELS

For the following models, we use `glmer` from the `lme4` package to fit generalized linear mixed-effects model, as the data is binomial (either an iRBC is a gametocyte - success - or is not - failure) and requires a mixed-effects model (to account for the potential effects of the individual plates, rows, etc.).

Earlier (incorrect) models

A note on earlier reports regarding the results of the analysis - earlier I reported that results were not significant. However, after extensively retesting the code/models and considering the potential variables to include/exclude, this may be an incorrect conclusion based on erroneous code, displayed below. There are two issues with the model as written here:

1. Hematocrit is here is included as a random effect, and not a fixed effect. However, while incorrect in terms of what the model represents, based on outputs from models with this error, the error ultimately doesn't change what conclusions would be drawn.
2. The much bigger issue is what variable is being used for `weights`. For binomial data, `glmer` allows for two methods for imputing data. One is to input the response variable data as a data frame of two columns, the number of successes and number of failures (which would be number of gametocytes and non-gametocytes out of the counted infected RBCs). The other method is to input data as *proportions*, which necessitates use of the `weights` argument, which should be supplied with the total number of trials which are being divided into successes or failures - i.e., the total number of asexual and sexual iRBCs, or `iRBCs Count`. Unfortunately, it seems what was supplied to `weights` was instead `iRBC_tdT+ Count`, the total number of successes. This change leads to a completely different conclusion that there is no significant difference in commitment between samples exposed to PBS and samples exposed to salivary gland extracts. Current coding for the models instead uses the data frame input for the binomial data to prevent future issues.

For comparison, here is the output from utilizing `iRBC_tdT+ Count` (incorrect):

```
fit_past <- glmer(Commit ~ Treatment + (1 | Hematocrit) + (1 | Replicate) + (1 | Plate), weights = `iRBC
```

```
## Warning in eval(family$initialize, rho): non-integer #successes in a binomial
## glm!
```

```
summary(fit_past)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: Commit ~ Treatment + (1 | Hematocrit) + (1 | Replicate) + (1 |
##     Plate)
##    Data: flow_data
## Weights: iRBC_tdT+ Count
##
##       AIC       BIC    logLik -2*log(L)  df.resid
##     480.8     497.2    -234.4     468.8       108
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.1428 -0.3865 -0.1240  0.1931  1.8070
##
## Random effects:
```

```
##  Groups     Name        Variance Std.Dev.
##  Replicate  (Intercept) 0.003821 0.06181
##  Plate      (Intercept) 0.393212 0.62707
##  Hematocrit (Intercept) 0.002751 0.05245
## Number of obs: 114, groups:  Replicate, 5; Plate, 4; Hematocrit, 2
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.94691    0.32723  -9.006   <2e-16 ***
## TreatmentF   0.08187    0.05738   1.427    0.154
## TreatmentM   0.01592    0.05795   0.275    0.784
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) TrtmnF
## TreatmentF -0.096
## TreatmentM -0.096  0.578
```

Here is the output from utilizing `iRBCs Count` (correct - though not the entire model, just what variable is supplied to `weights`):

```
fit_past <- glmer(Commit ~ Treatment + (1 | Hematocrit) + (1 | Replicate) + (1 | Plate), weights = `iRBC
```

```
## Warning in eval(family$initialize, rho): non-integer #successes in a binomial
## glm!
```

```
summary(fit_past)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: Commit ~ Treatment + (1 | Hematocrit) + (1 | Replicate) + (1 |
##     Plate)
##    Data: flow_data
## Weights: iRBCs Count
##
##       AIC       BIC    logLik -2*log(L)  df.resid
##    1430.8    1447.2    -709.4    1418.8       108
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -7.9258 -1.5172 -0.2451  1.3515  7.7085
##
## Random effects:
##  Groups     Name        Variance Std.Dev.
##  Replicate  (Intercept) 0.005884 0.07671
##  Plate      (Intercept) 0.469318 0.68507
##  Hematocrit (Intercept) 0.003703 0.06085
## Number of obs: 114, groups:  Replicate, 5; Plate, 4; Hematocrit, 2
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.00321    0.34730  -8.647  < 2e-16 ***
## TreatmentF   0.04605    0.01660   2.775  0.00553 **
## TreatmentM  -0.01566    0.01653  -0.947  0.34346
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) TrtmnF
## TreatmentF -0.026
## TreatmentM -0.026  0.549
```

Base GLM mixed models

These models are all generally structured to have commitment proportion (or rather, number of gametocytes versus asexual iRBCs) as a response variable, salivary treatment and hematocrit treatment as fixed effects, and plate and replicate numbers as fixed effects. The code below also tests out minor variations in this structure, mainly changing the model to have replicate nested within plate ID, or changing the data to exclude certain plates. However, we've found that regardless of minor variations, there is an overal pattern that *there is no significant difference in commitment between samples exposed to male salivary gland extracts versus PBS control, but there is a significant difference between female saliva exposed samples versus PBS control.*

1. Treatment + Hematocrit as fixed effects, Plate & Replicate as random effects.

```
# using cbind of successes & failures
fit_1 <- glmer(cbind(`iRBC_tdT+ Count`, `iRBCs Count` - `iRBC_tdT+ Count`) ~ Treatment + Hematocrit + (
summary(fit_1)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: cbind(`iRBC_tdT+ Count`, `iRBCs Count` - `iRBC_tdT+ Count`) ~
##     Treatment + Hematocrit + (1 | Plate) + (1 | Replicate)
##    Data: flow_data
##
##       AIC       BIC    logLik -2*log(L)  df.resid
##    1441.1    1457.5    -714.5    1429.1       108
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -7.7886 -1.5115 -0.1215  1.4994  8.2032
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  Replicate (Intercept) 0.00606  0.07785
##  Plate     (Intercept) 0.47235  0.68728
## Number of obs: 114, groups:  Replicate, 5; Plate, 4
##
## Fixed effects:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -2.963396   0.345911  -8.567  < 2e-16 ***
## TreatmentF     0.049291   0.016653   2.960  0.00308 **
## TreatmentM    -0.008788   0.016571  -0.530  0.59588
## Hematocrit1.0 -0.096830   0.013850  -6.991 2.72e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
```

```
##             (Intr) TrtmnF TrtmnM
## TreatmentF  -0.027
## TreatmentM  -0.028  0.550
## Hematcrt1.0 -0.030  0.065  0.092
```

2. Treatment + Hematocrit as fixed effects, Plate & Replicate as random - same as before, but replicate is nested within plate (each plate has a `R1`, `R2`, etc.)

```
fit_1 <- glmer(cbind(`iRBC_tdT+ Count`, `iRBCs Count` - `iRBC_tdT+ Count`) ~ Treatment + Hematocrit + (
summary(fit_1)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: cbind(`iRBC_tdT+ Count`, `iRBCs Count` - `iRBC_tdT+ Count`) ~
##     Treatment + Hematocrit + (1 | Plate/Replicate)
##    Data: flow_data
##
##       AIC       BIC    logLik -2*log(L)  df.resid
##    1249.9    1266.3    -618.9    1237.9       108
##
## Scaled residuals:
##    Min     1Q Median     3Q    Max
## -4.608 -1.064 -0.058  1.085  7.730
##
## Random effects:
##  Groups          Name        Variance Std.Dev.
##  Replicate:Plate (Intercept) 0.01801  0.1342
##  Plate           (Intercept) 0.46209  0.6798
## Number of obs: 114, groups:  Replicate:Plate, 20; Plate, 4
##
## Fixed effects:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -2.96410    0.34175  -8.673  < 2e-16 ***
## TreatmentF     0.04648    0.01675   2.775  0.00551 **
## TreatmentM    -0.01922    0.01662  -1.156  0.24758
## Hematocrit1.0 -0.10153    0.01388  -7.316 2.56e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) TrtmnF TrtmnM
## TreatmentF  -0.028
## TreatmentM  -0.029  0.548
## Hematcrt1.0 -0.030  0.067  0.095
```

Excluding Plate 2

Plate 2 has multiple wells where events were clustering outside the expected are of the first gate, indicating potential issues with sampling (due to settlement of cells?). We re-run the models with the exclusion of Plate 2.

```
flow_data_minus2 <- flow_data[flow_data$Plate != "SGE-2", ]
```

1. Treatment + Hematocrit as fixed effects, Plate & Replicate as random

```r
# using cbind of successes & failures
fit_1 <- glmer(cbind(`iRBC_tdT+ Count`, `iRBCs Count` - `iRBC_tdT+ Count`) ~ Treatment + Hematocrit + (
summary(fit_1)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: cbind(`iRBC_tdT+ Count`, `iRBCs Count` - `iRBC_tdT+ Count`) ~
##     Treatment + Hematocrit + (1 | Plate) + (1 | Replicate)
##    Data: flow_data_minus2
##
##       AIC       BIC    logLik -2*log(L)  df.resid
##    1233.3    1248.2    -610.6    1221.3        82
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -7.5332 -1.5690 -0.1981  1.8170  8.0569
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  Replicate (Intercept) 0.006082 0.07798
##  Plate     (Intercept) 0.553862 0.74422
## Number of obs: 88, groups:  Replicate, 5; Plate, 3
##
## Fixed effects:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -2.821125   0.431557  -6.537 6.27e-11 ***
## TreatmentF     0.047207   0.017222   2.741  0.00612 **
## TreatmentM    -0.001138   0.017117  -0.066  0.94701
## Hematocrit1.0 -0.107353   0.014241  -7.538 4.77e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) TrtmnF TrtmnM
## TreatmentF  -0.022
## TreatmentM  -0.024  0.553
## Hematcrt1.0 -0.025  0.066  0.094
```

2. Treatment + Hematocrit as fixed effects, Plate & Replicate as randomf with Replicate is nested within Plate.

```r
fit_1 <- glmer(cbind(`iRBC_tdT+ Count`, `iRBCs Count` - `iRBC_tdT+ Count`) ~ Treatment + Hematocrit + (
summary(fit_1)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: cbind(`iRBC_tdT+ Count`, `iRBCs Count` - `iRBC_tdT+ Count`) ~
##     Treatment + Hematocrit + (1 | Plate/Replicate)
##    Data: flow_data_minus2
##
##       AIC       BIC    logLik -2*log(L)  df.resid
##    1037.7    1052.6    -512.9    1025.7        82
```

```
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.4427 -1.2963  0.0068  1.2662  7.6141
##
## Random effects:
##  Groups          Name        Variance Std.Dev.
##  Replicate:Plate (Intercept) 0.01983  0.1408
##  Plate           (Intercept) 0.54754  0.7400
## Number of obs: 88, groups:  Replicate:Plate, 15; Plate, 3
##
## Fixed effects:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -2.82778    0.42924  -6.588 4.46e-11 ***
## TreatmentF     0.04548    0.01726   2.634  0.00843 **
## TreatmentM    -0.01256    0.01717  -0.732  0.46435
## Hematocrit1.0 -0.11263    0.01427  -7.893 2.95e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) TrtmnF TrtmnM
## TreatmentF  -0.023
## TreatmentM  -0.024  0.553
## Hematcrt1.0 -0.025  0.069  0.097
```

Excluding Plate 4

Plate 4 likewise has issues where points would cluster outside the first gate (though unlike Plate 2, the main gate still had many points within them). It also had more cases where boundary between iRBCs and uRBCs were a bit more blurred in gate 2 and 3 (contrasted to Plate 2's sparser points, but still clear boundaries between points in versus outside of gate 2 & 3). Models here exlucde Plate 4 (but have Plate 2).

```
flow_data_minus4 <- flow_data[flow_data$Plate != "SGE-4", ]
```

1. Treatment + Hematocrit as fixed effects, Plate & Replicate as random

```
# using cbind of successes & failures
fit_1 <- glmer(cbind(`iRBC_tdT+ Count`, `iRBCs Count` - `iRBC_tdT+ Count`) ~ Treatment + Hematocrit + (
summary(fit_1)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: cbind(`iRBC_tdT+ Count`, `iRBCs Count` - `iRBC_tdT+ Count`) ~
##     Treatment + Hematocrit + (1 | Plate) + (1 | Replicate)
##    Data: flow_data_minus4
##
##       AIC       BIC    logLik -2*log(L)  df.resid
##    1115.3    1129.9    -551.7    1103.3        78
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -7.9123 -1.3766 -0.1844  1.4945  6.4651
##
```

```
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  Replicate (Intercept) 0.006063 0.07787
##  Plate     (Intercept) 0.262285 0.51214
## Number of obs: 84, groups:  Replicate, 5; Plate, 3
##
## Fixed effects:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -2.6737259  0.2982485  -8.965  < 2e-16 ***
## TreatmentF     0.0610039  0.0169177   3.606 0.000311 ***
## TreatmentM     0.0003148  0.0168893   0.019 0.985131
## Hematocrit1.0 -0.0873154  0.0140281  -6.224 4.84e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) TrtmnF TrtmnM
## TreatmentF  -0.032
## TreatmentM  -0.033  0.553
## Hematcrt1.0 -0.034  0.065  0.094
```

2. Treatment + Hematocrit as fixed effects, Plate & Replicate as randomf with Replicate is nested within Plate.

```
fit_1 <- glmer(cbind(`iRBC_tdT+ Count`, `iRBCs Count` - `iRBC_tdT+ Count`) ~ Treatment + Hematocrit + (
summary(fit_1)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: cbind(`iRBC_tdT+ Count`, `iRBCs Count` - `iRBC_tdT+ Count`) ~
##     Treatment + Hematocrit + (1 | Plate/Replicate)
##    Data: flow_data_minus4
##
##       AIC       BIC    logLik -2*log(L)  df.resid
##     929.0     943.6    -458.5     917.0        78
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.4555 -0.9690 -0.0238  1.0061  4.6916
##
## Random effects:
##  Groups          Name        Variance Std.Dev.
##  Replicate:Plate (Intercept) 0.01754  0.1324
##  Plate           (Intercept) 0.24307  0.4930
## Number of obs: 84, groups:  Replicate:Plate, 15; Plate, 3
##
## Fixed effects:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -2.671969   0.287261  -9.302  < 2e-16 ***
## TreatmentF     0.058676   0.017004   3.451 0.000559 ***
## TreatmentM    -0.009146   0.016936  -0.540 0.589184
## Hematocrit1.0 -0.091949   0.014057  -6.541 6.11e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Correlation of Fixed Effects:
##             (Intr) TrtmnF TrtmnM
## TreatmentF  -0.035
## TreatmentM  -0.035  0.551
## Hematcrt1.0 -0.035  0.068  0.098
```

OTHER PLOTS

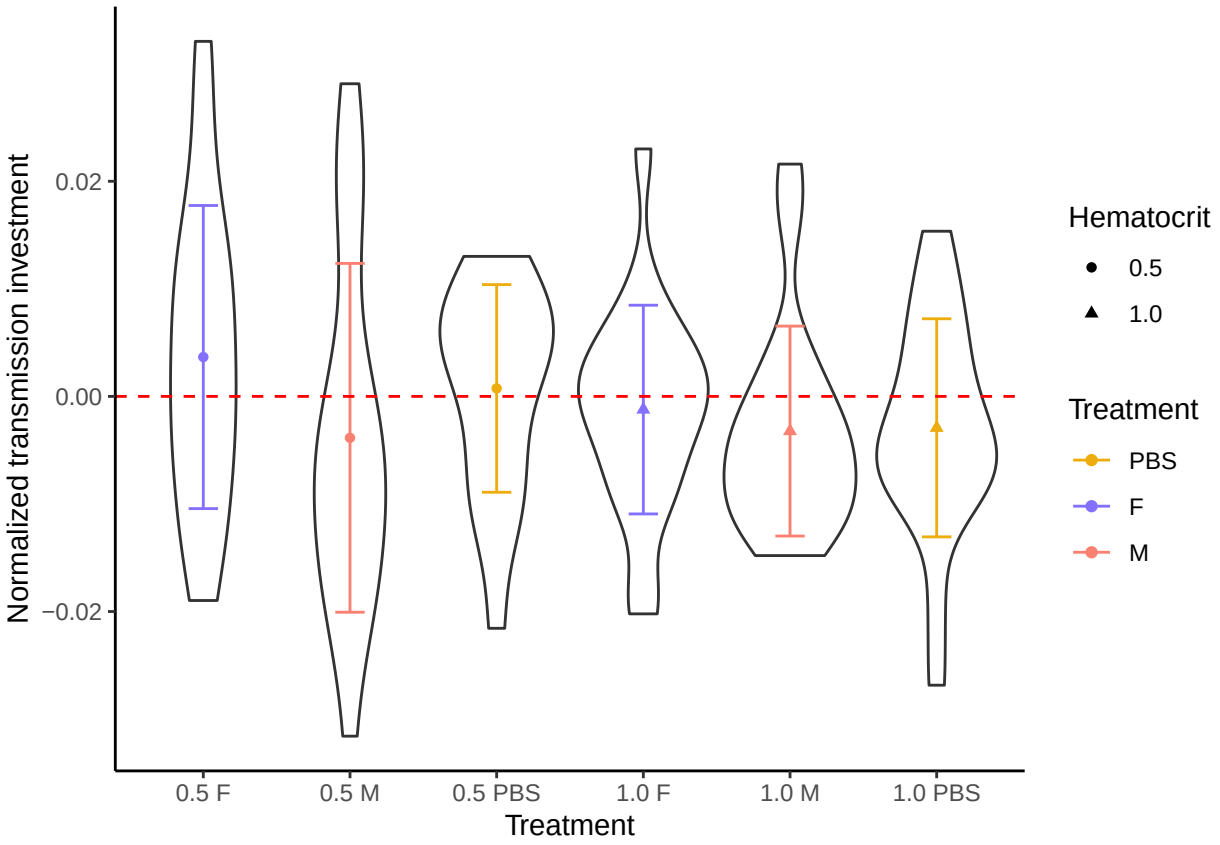
Comparison of commitment

Plot of (normalized) commitment proportion among salivary and hematocrit treatments. Violin plot shows distribution, the colored points show mean proportion of iRBCs in sample consisting of gametocytes (commitment), with error bars showing standard deviation. Normalization was done by subtracting weighted mean of control (weighted by singlets counts) from all data points, a process that was done individually for each plate to account for singlets count differences between plates.

```
normalized <- flow_data

for (plate in unique(flow_data$Plate)) {
  # get weighted mean of control for the plate
  control_commits <- normalized[normalized$Plate == plate & normalized$Treatment == "PBS", c("Commit")]
  weight_control <- normalized[normalized$Plate == plate & normalized$Treatment == "PBS", c("Singlets C
  mean_control <- weighted.mean(control_commits,weight_control)

  # normalize
  normalized[normalized$Plate == plate, c("Commit")] <- normalized[normalized$Plate == plate, c("Commit
}



normalized_mean <- normalized %>%
  group_by(Treatment, Hematocrit) %>%
  summarise(mean_commit = mean(Commit), sd = sd(Commit))
```

```
## `summarise()` has grouped output by 'Treatment'. You can override using the
## `.groups` argument.
```

```
normalized_mean$Hem_Treat <- paste(normalized_mean$Hematocrit, normalized_mean$Treatment)


# png("saliva.png", width = 4, height = 3, units = "in", res = 300)
ggplot(data = normalized_mean, aes(x=`Hem_Treat`, y=mean_commit)) +
  geom_violin(data = normalized, aes(x=`Hem_Treat`, y=`Commit`)) +
  geom_point(data = normalized_mean, aes(x=`Hem_Treat`, y=mean_commit, shape = Hematocrit, color = Trea
  geom_errorbar(data = normalized_mean, aes(ymin=mean_commit-sd, ymax=mean_commit+sd, color = Treatment
  scale_color_manual(values=c("darkgoldenrod2",
                              "lightslateblue",
                              "salmon")) +
  geom_hline(yintercept=0, linetype="dashed", color = "red") +
  # ggtitle("Commitment proportion per group (normalized)") +
  labs(y= "Normalized transmission investment", x = "Treatment") +
  theme_classic() +
  stat_summary(fun.data=mean_sdl,
               geom="errorbar", fun.args = list(mult = 1), color="black")
```



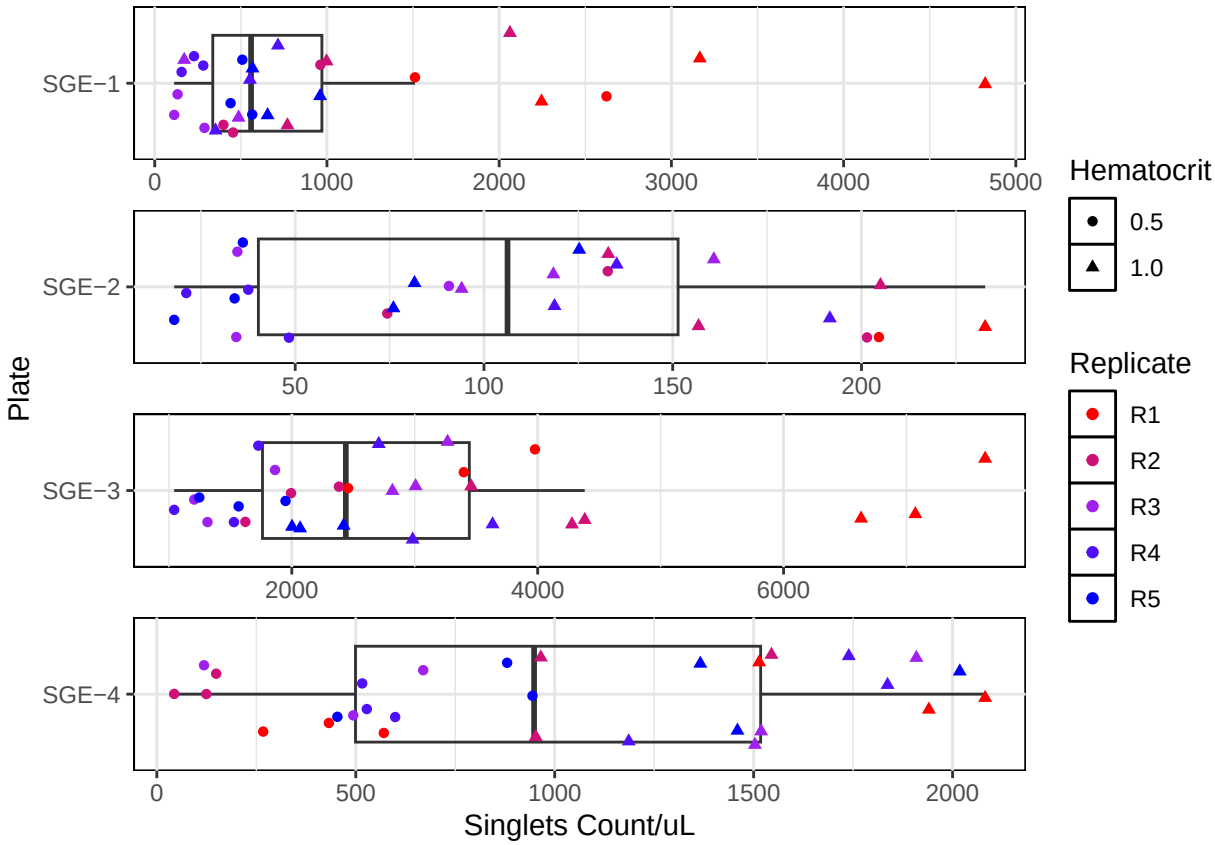

```
# dev.off()
```

Plot of singlets counts per uL across different plates, colored by replicate (row) the sample was located in. Earlier rows tend to have higher singlets counts for all plates except plate 4, showing a potential effect of replicate ID on the result.

```
fun_color_range <- colorRampPalette(c("red","purple","blue"))
colors <- fun_color_range(5)

# png("replicate.png", width = 4, height = 4, units = "in", res = 300)
ggplot(data = flow_data) +
  facet_wrap(. ~ Plate, scales = "free", nrow = 4) +
  geom_boxplot(alpha = 0, aes(y=Plate, x=`Singlets Count/uL`)) +
  geom_jitter(aes(y=1, x=`Singlets Count/uL`, color= Replicate, shape = Hematocrit)) +
  scale_color_manual(values=colors) +
  theme(strip.text.x = element_blank(),
        panel.background = element_rect(fill = "white",
                                        colour = "black",
                                        linewidth = 0.5, linetype = "solid"),
        panel.grid.major = element_line(size = 0.5, linetype = 'solid',
                                        colour = "grey90"),
        panel.grid.minor = element_line(size = 0.25, linetype = 'solid',
                                        colour = "grey90"))
```

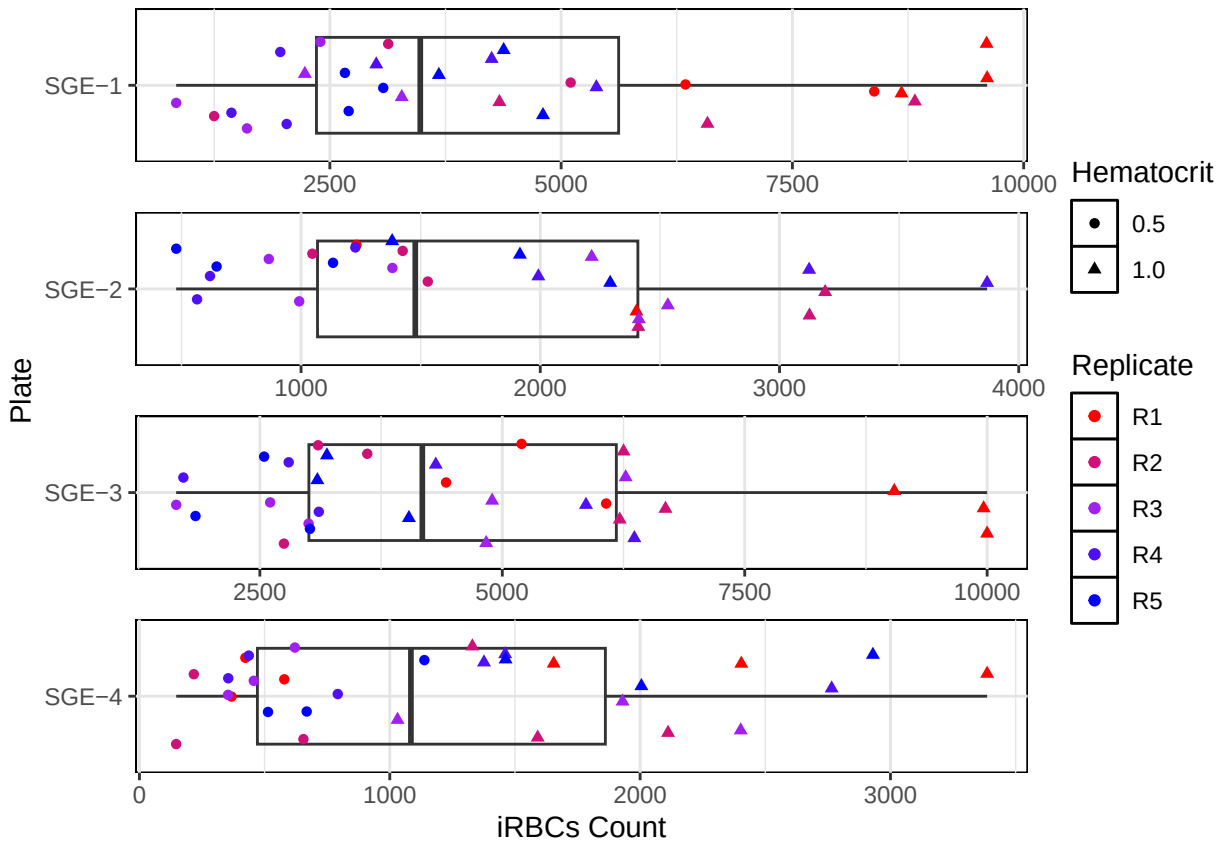
```
## Warning: The `size` argument of `element_line()` is deprecated as of ggplot2 3.4.0.
## i Please use the `linewidth` argument instead.
## This warning is displayed once every 8 hours.
```

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

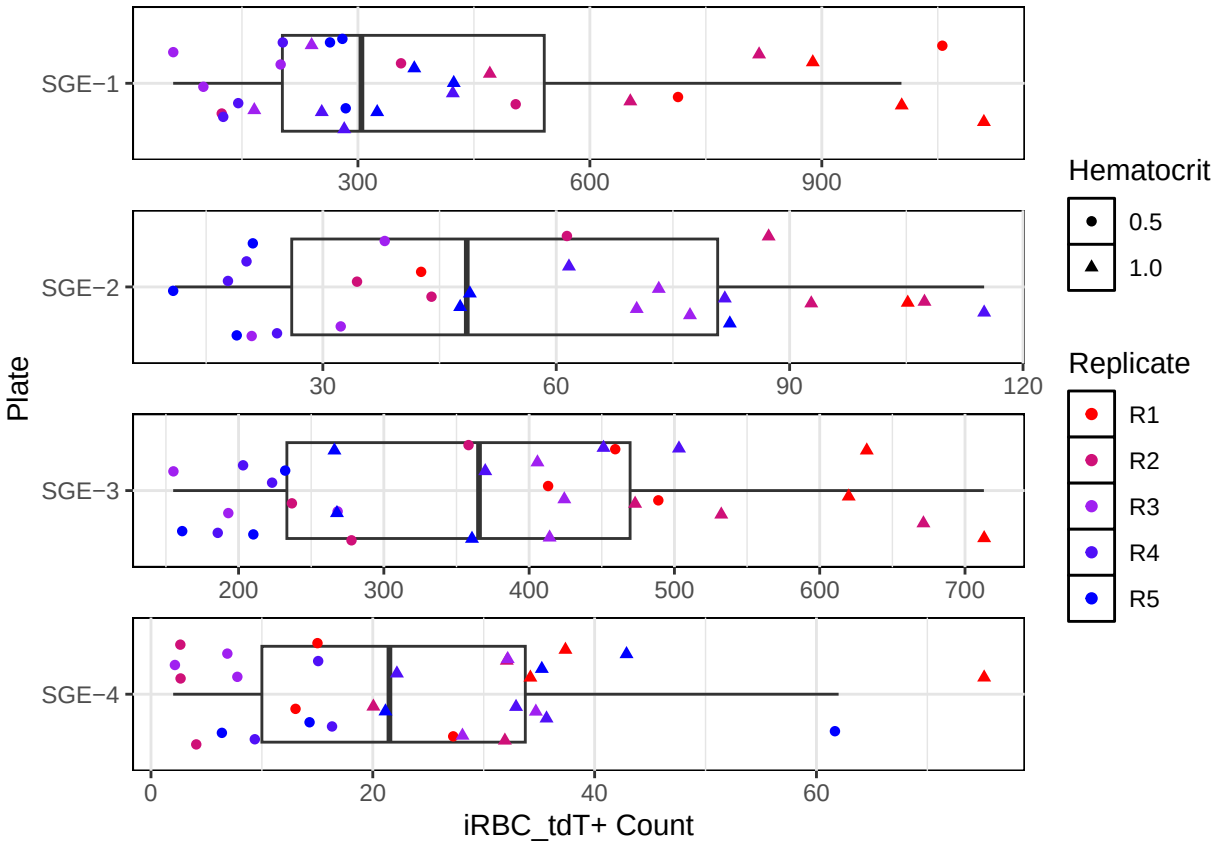


Similar plot showing iRBC counts across plates.

```
fun_color_range <- colorRampPalette(c("red","purple","blue"))
colors <- fun_color_range(5)

# png("replicate.png", width = 4, height = 4, units = "in", res = 300)
ggplot(data = flow_data) +
  facet_wrap(. ~ Plate, scales = "free", nrow = 4) +
  geom_boxplot(alpha = 0, aes(y=Plate, x=`iRBCs Count`)) +
  geom_jitter(aes(y=1, x=`iRBCs Count`, color= Replicate, shape = Hematocrit)) +
  scale_color_manual(values=colors) +
  theme(strip.text.x = element_blank(),
        panel.background = element_rect(fill = "white",
                                        colour = "black",
                                        linewidth = 0.5, linetype = "solid"),
        panel.grid.major = element_line(size = 0.5, linetype = 'solid',
                                        colour = "grey90"),
        panel.grid.minor = element_line(size = 0.25, linetype = 'solid',
                                        colour = "grey90"))
```




```
# png("replicate.png", width = 4, height = 4, units = "in", res = 300)
ggplot(data = flow_data) +
  facet_wrap(. ~ Plate, scales = "free", nrow = 4) +
  geom_boxplot(alpha = 0, aes(y=Plate, x=`iRBC_tdT+ Count`)) +
  geom_jitter(aes(y=1, x=`iRBC_tdT+ Count`, color= Replicate, shape = Hematocrit)) +
  scale_color_manual(values=colors) +
  theme(strip.text.x = element_blank(),
        panel.background = element_rect(fill = "white",
                                        colour = "black",
                                        linewidth = 0.5, linetype = "solid"),
        panel.grid.major = element_line(size = 0.5, linetype = 'solid',
                                        colour = "grey90"),
        panel.grid.minor = element_line(size = 0.25, linetype = 'solid',
                                        colour = "grey90"))
```



Cytometry clustering??

Some of the cytometry plots showed issues that may indicate that the results (and any resulting analyses) may be untrustworthy. Here is a decent flow cytometry result for reference:

Some issues with the cytometry data include the following:

- Plots show 'bleeding' between boundaries: gates were left to presets to avoid biasing data (with the exception of the first gate on plate 4), but on some wells the boundaries between iRBCs and uRBCs is more ambiguous, where singlets may arguably be on one side or the other. While this largely doesn't affect gametocyte counts, moving the gating up or down will decrease/increase counts of asexual iRBCs, which in turn increases/decreases the commitment proportion.
- Plots show comparatively sparse amounts of singlets/events. While events clearly fit within the gating (perhaps barring initial gating of singlets from events, see below), the number of points is relatively sparse overall.
- Plots show many events lying outside the initial gating which separates all events (all points recorded by the cytometer) from legitimate singlets (cells). This may potentially be an indication of dead cells (and thus potential issues during the procedure). The level of clustering outside of the first gate varies from sample to sample. Some plots show only show a small cluster of events outside the first gate (first example below), but otherwise the majority of the events fit within the gates. However, other plots show a majority of the points lying outside the gates (second and third example below), potentially leading to a large portion of points lying outside the first gate.

The four plots below shows the counts of all events (dark gray bars) compared to the singlets count (colored portion) for each individual well for all four plates.

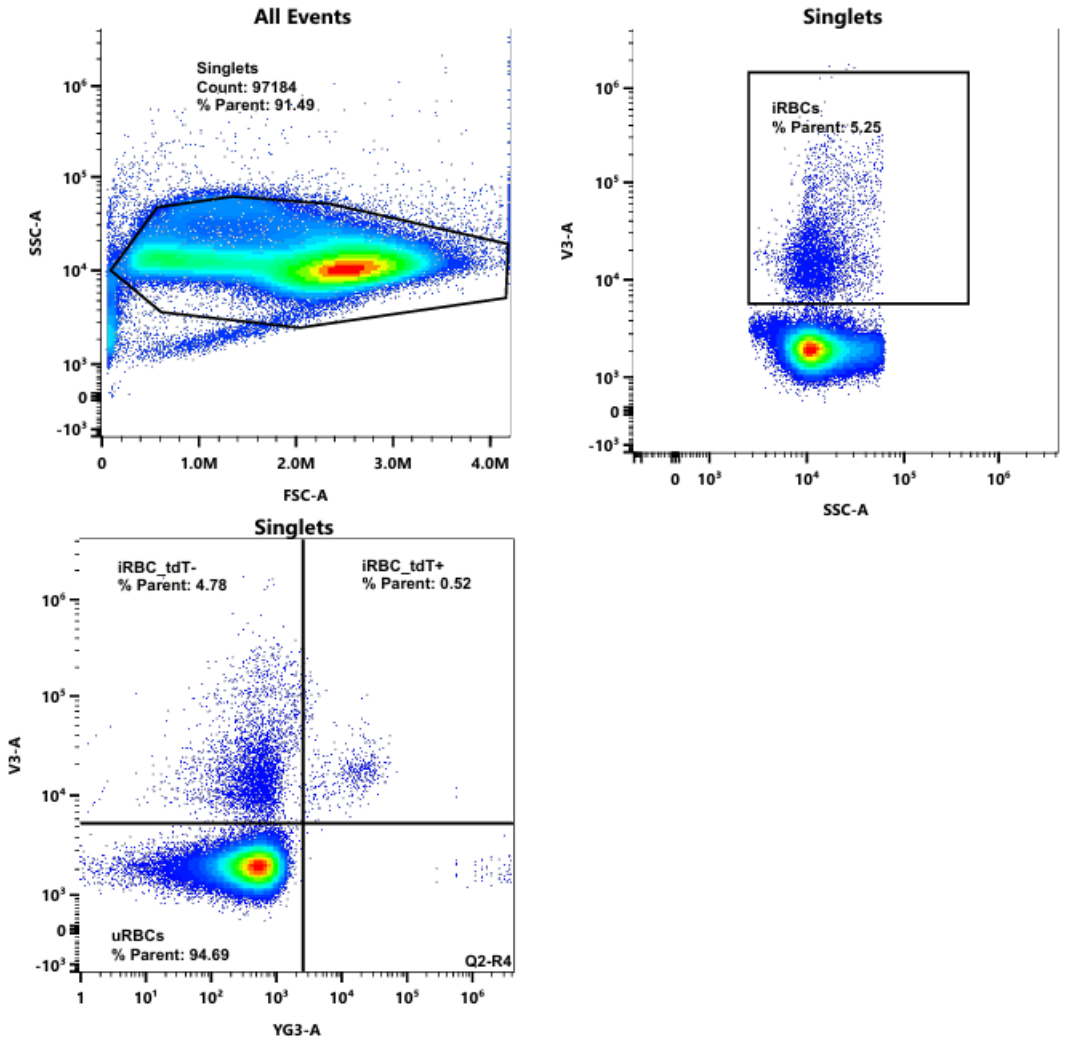


Figure 2: Reliable flow cytometry plot: shows singlets mostly appearing within the boundaries of the first gates and iRBCs clearly separated from the uRBCs/sexual and asexual iRBCs clearly separated from each other.

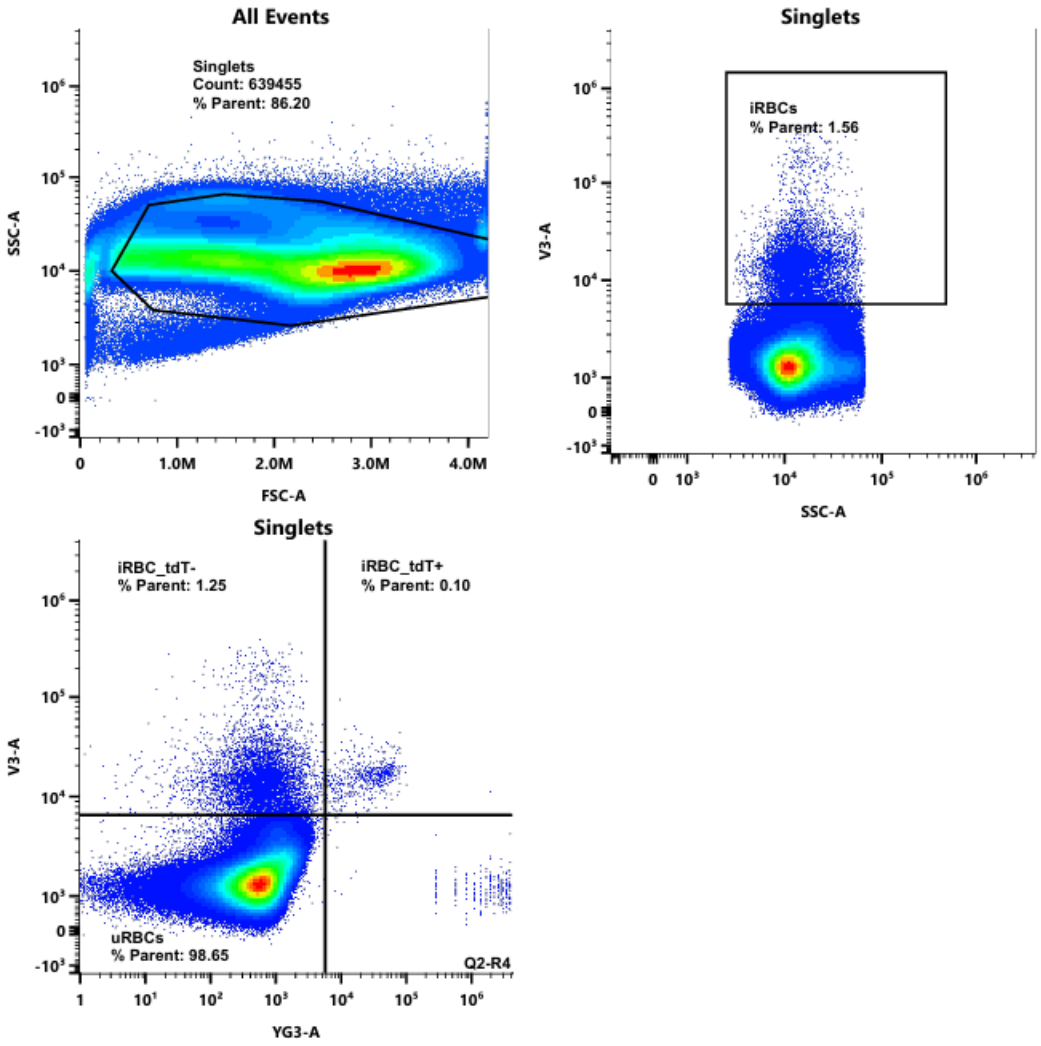


Figure 3: Separation of events at the second (upper right) and third gate (lower left) is more questionable.

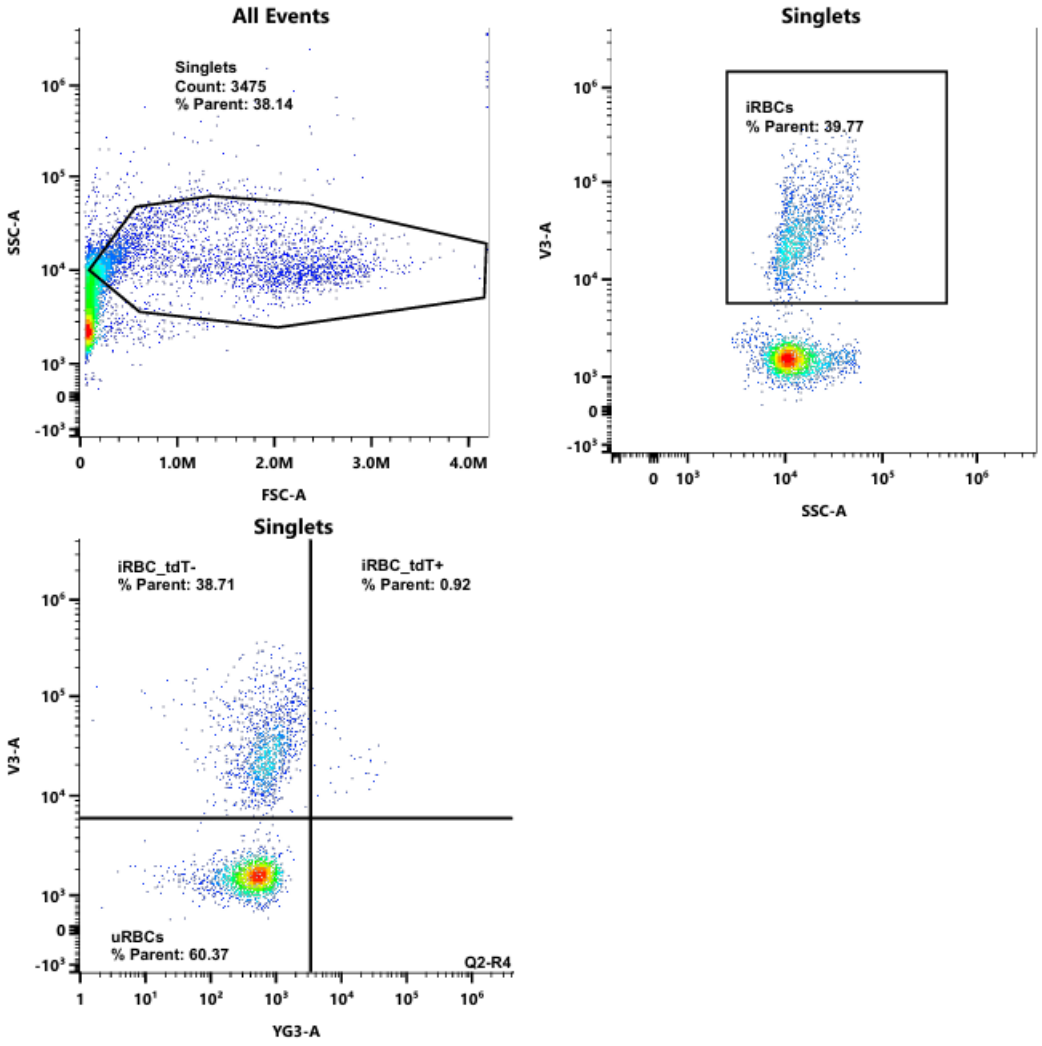


Figure 4: Plot shows sparse number of events/singlets.

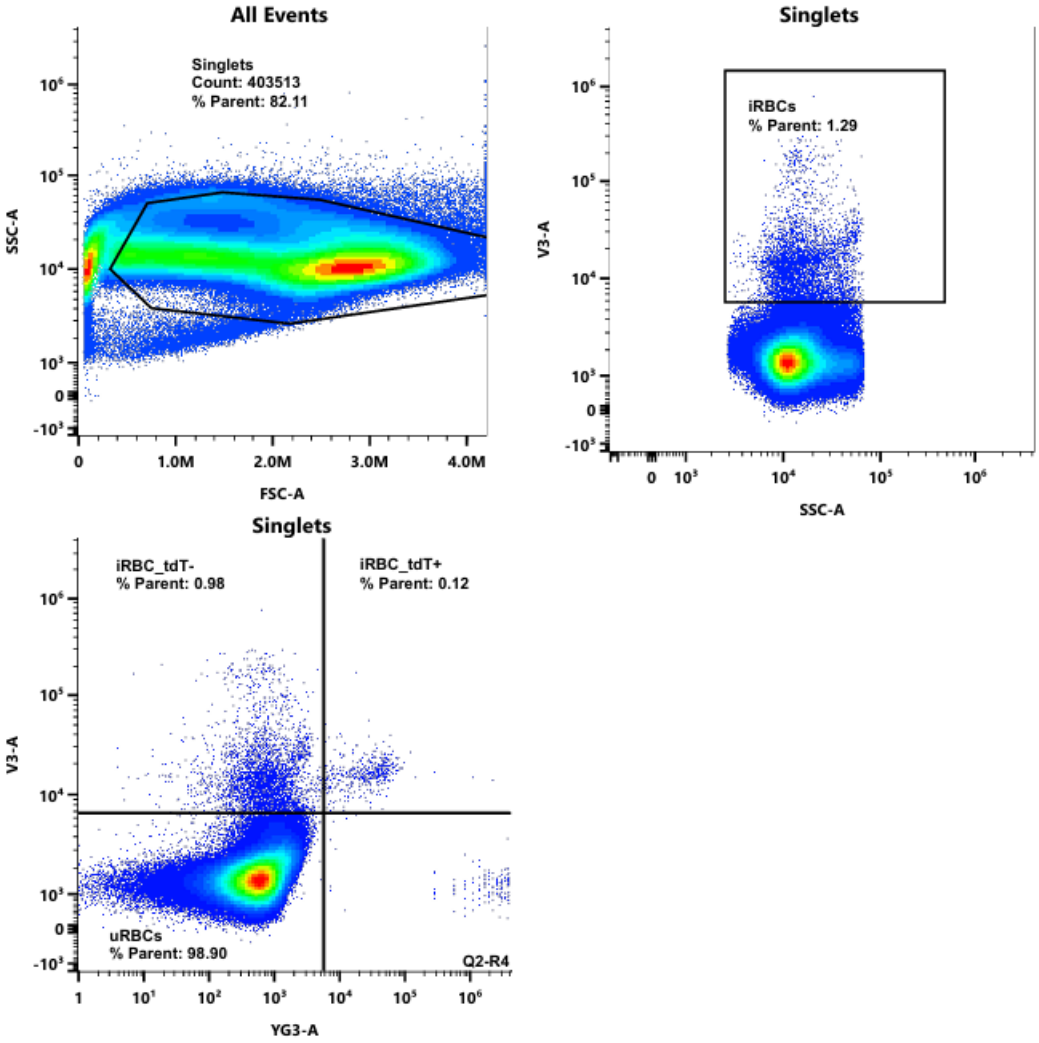


Figure 5: Upper left plot shows 'light' clustering outside the first gate. Points outside the gate are typically found to the very left of the plot.

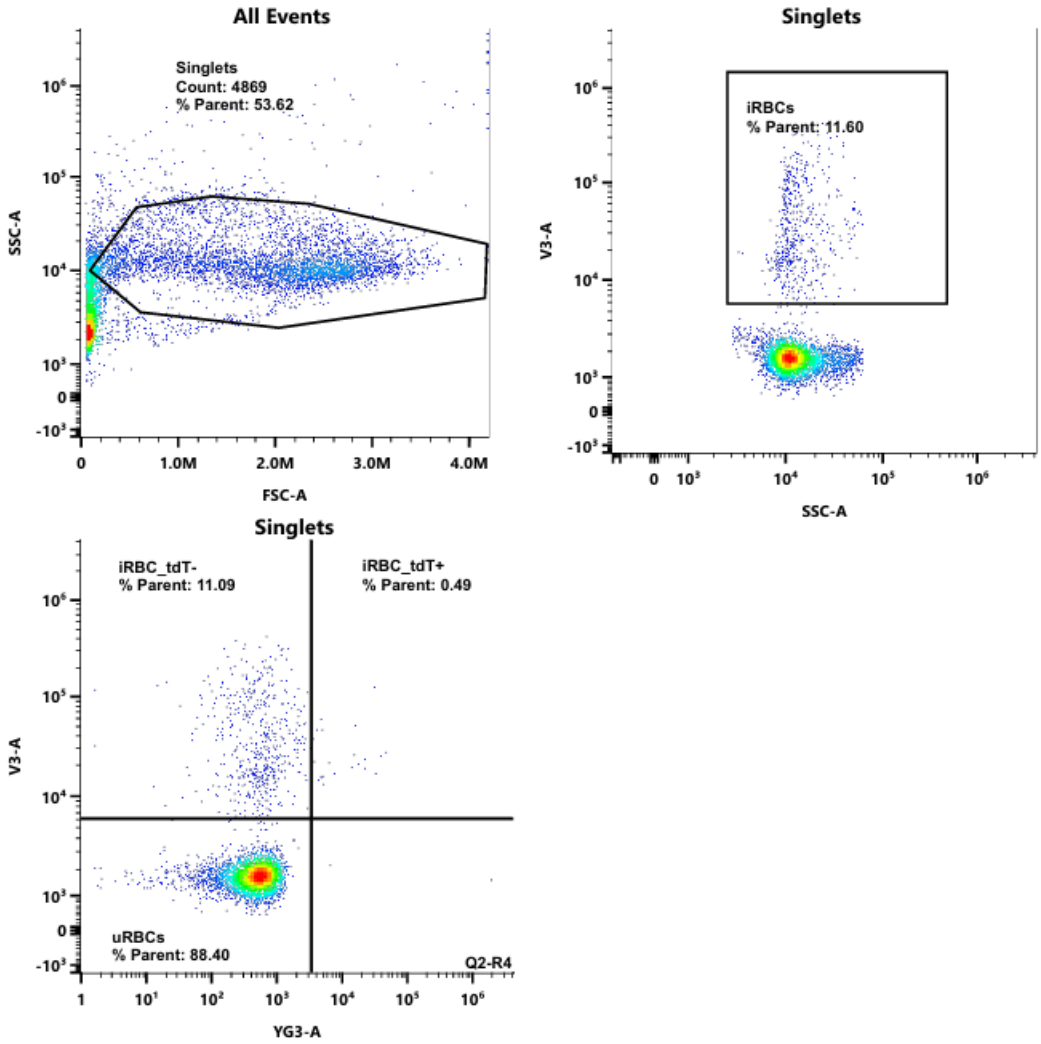


Figure 6: Upper left plot shows a majority of points clustering outside the first gate (again, to the left, slightly lower), resulting in sparser plots fo singlets.

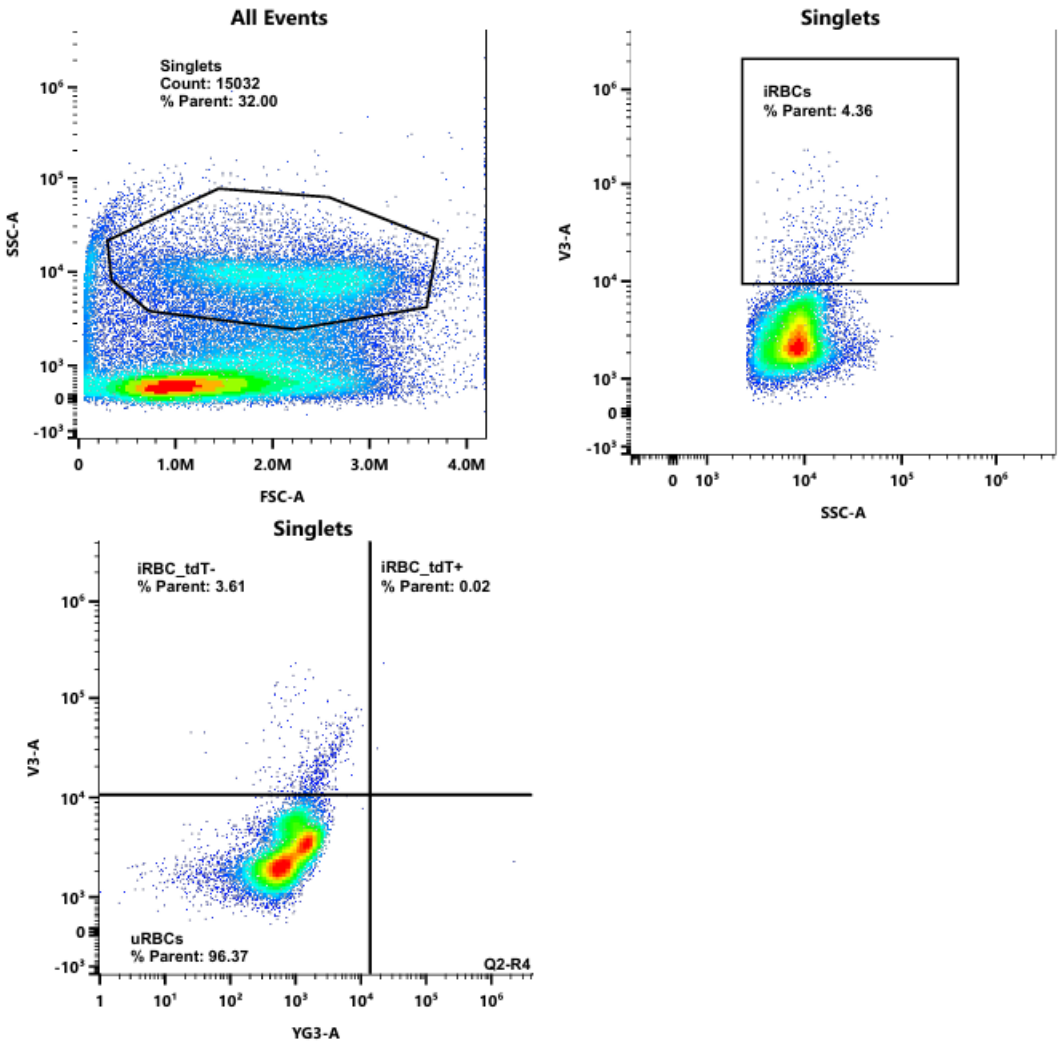
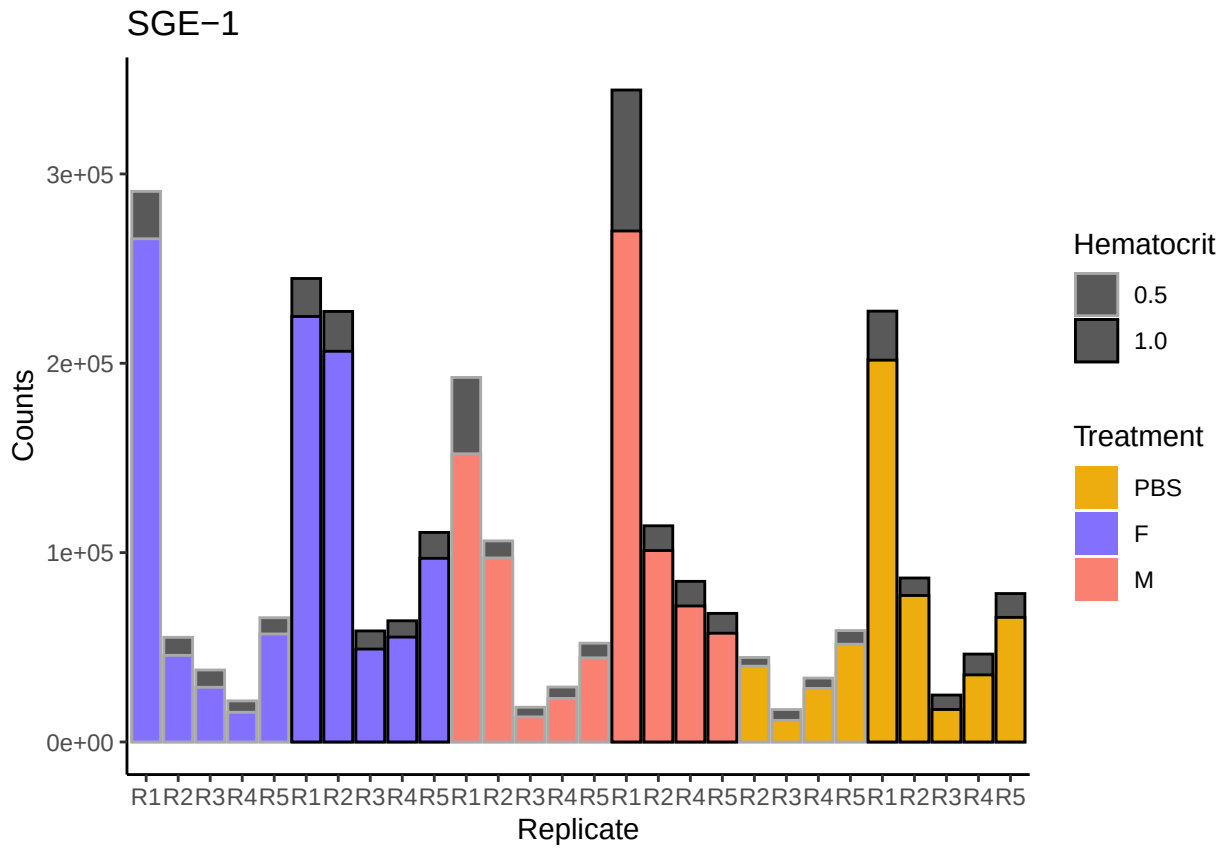
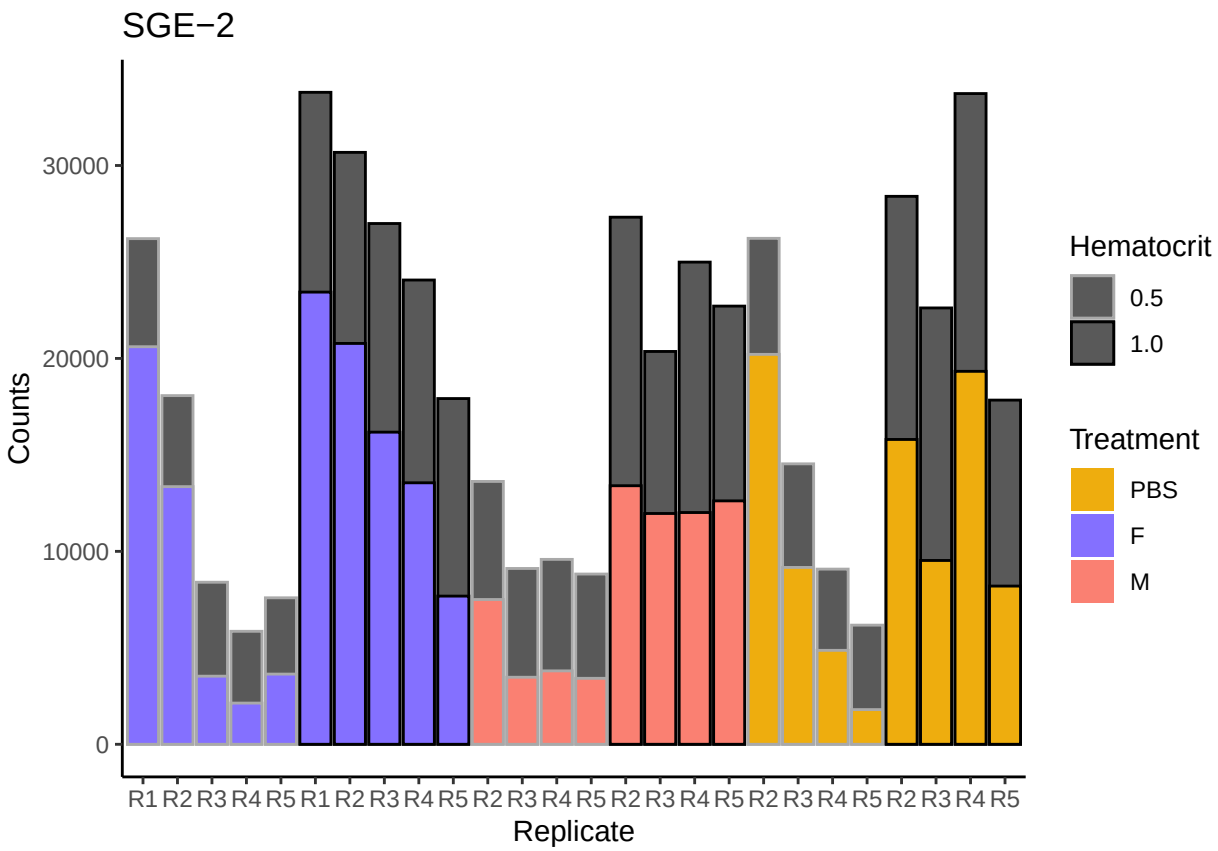


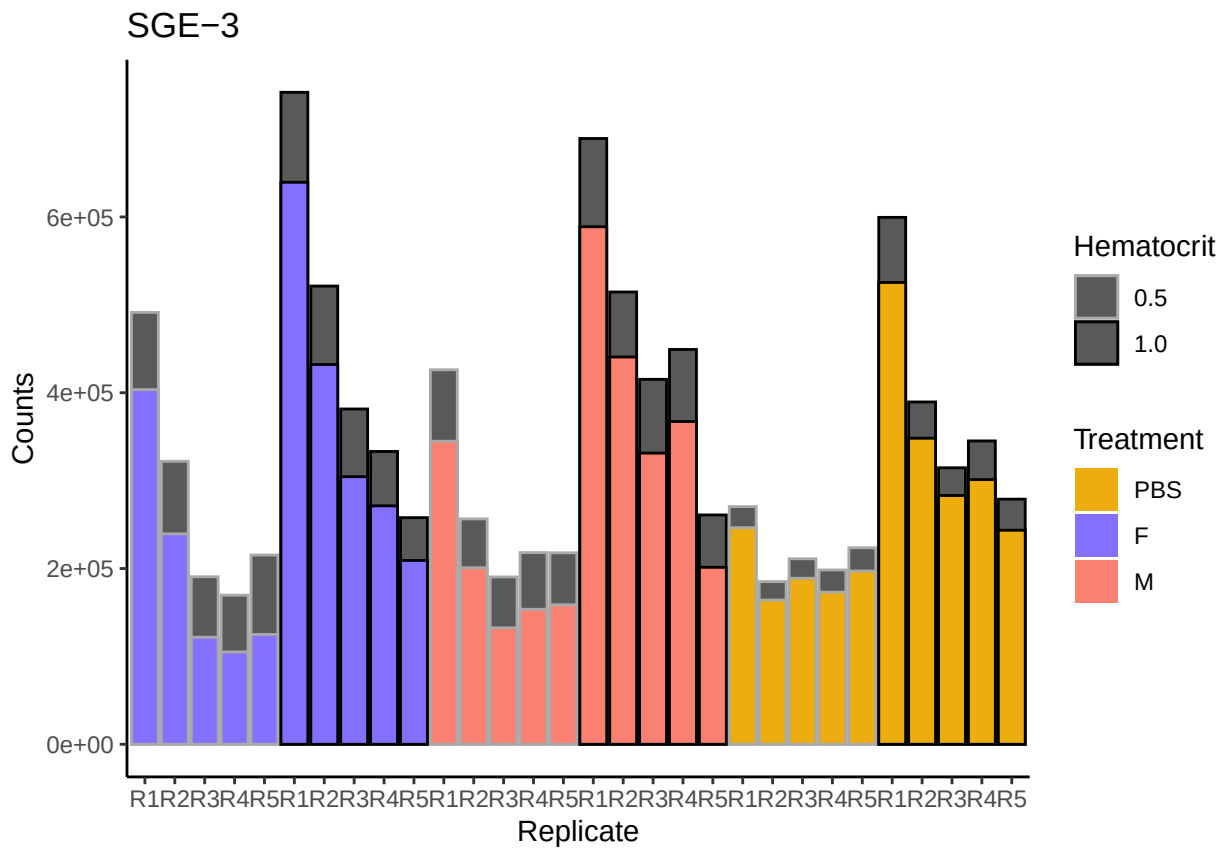
Figure 7: An unusual example of clustering outside the gate, where most points are found clustering on the bottom left of the plot. This sample (from plate 4, female saliva & 0.5% hematocrit treatment row 2) is unusual, and its clustering pattern only shared with two other samples, both on plate 4 (female saliva & 1.0% hematocrit treatment row 5, and PBS control & 1.0% hematocrit treatment row 4).

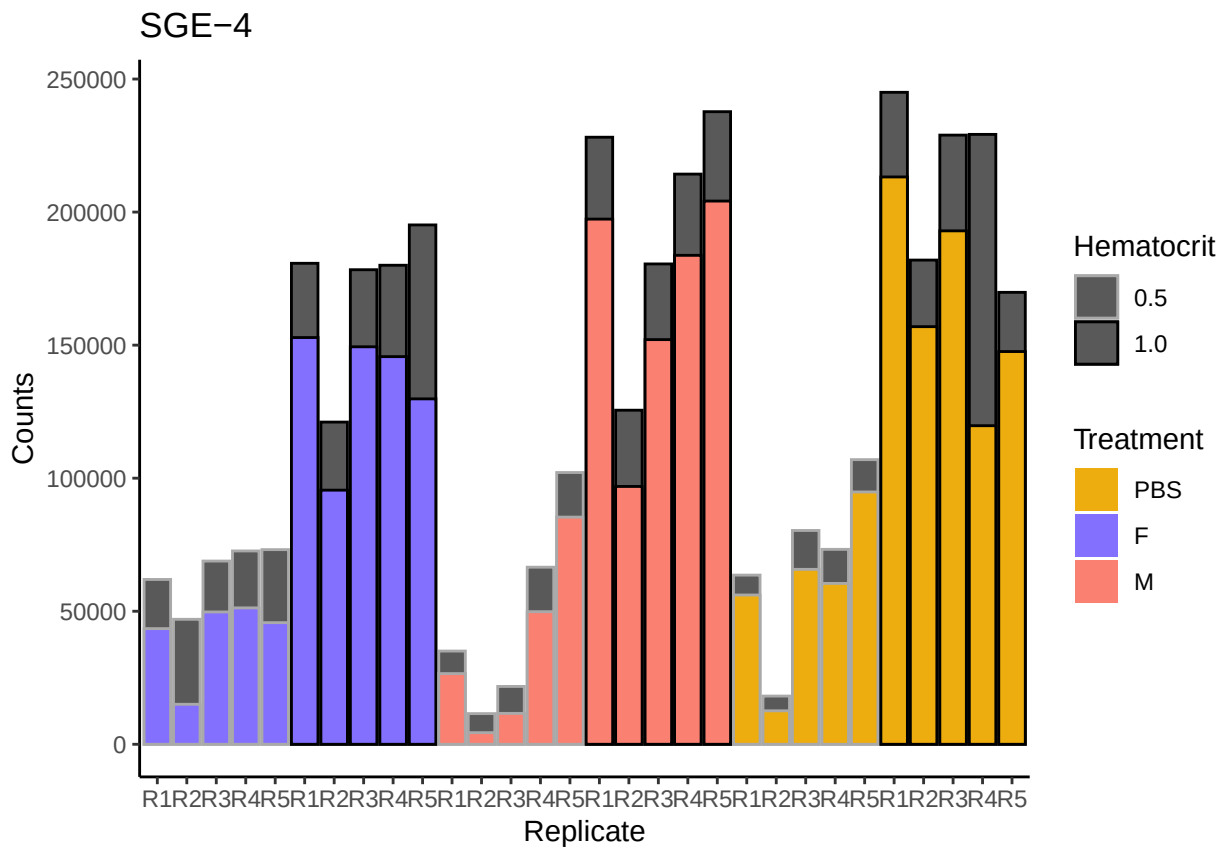
```
flow_data <- flow_data[order(flow_data$Treatments_FULL),]
```

```
for (plate in unique(flow_data$Plate)) {
  dataset = flow_data[flow_data$Plate == plate,]

  print(ggplot() +
      geom_bar(data = dataset, aes(y=`All Events Count`, x=Treatments_FULL, color = Hematocrit), stat='
      geom_bar(data = dataset, aes(y=`Singlets Count`, x=Treatments_FULL, fill = Treatment, color = Hema
      ggtitle(plate) +
      ylab("Counts") + xlab("Replicate") +
      scale_fill_manual(values=c("darkgoldenrod2",
                                 "lightslateblue",
                                 "salmon")) +
      scale_color_manual(values=c("darkgray",
                                  "black")) +
      scale_x_discrete(labels= dataset$Replicate) +
      theme_classic()
  )
}
```



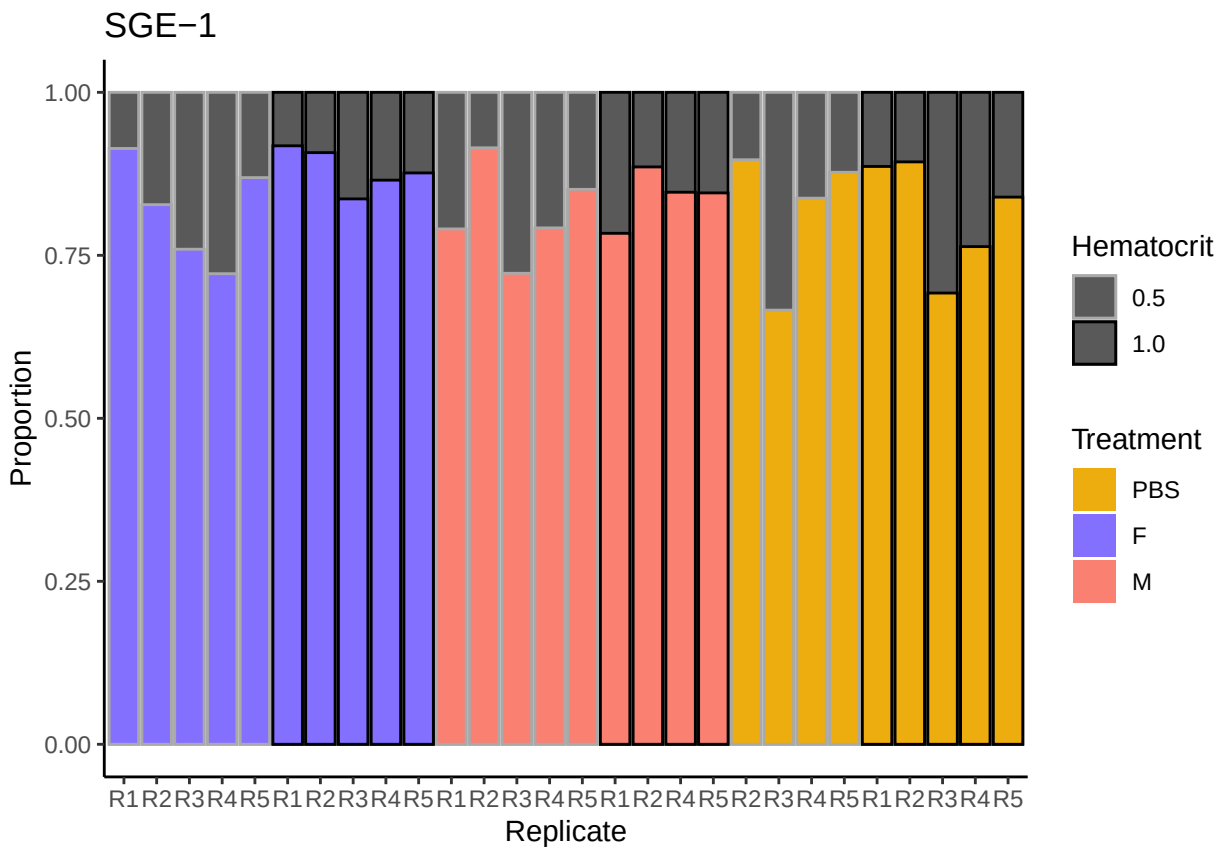

SGE−2
Counts
30000
20000
10000
0
R1 R2 R3 R4 R5 R1 R2 R3 R4 R5 R2 R3 R4 R5 R2 R3 R4 R5 R2 R3 R4 R5 R2 R3 R4 R5
Replicate
Hematocrit
0.5
1.0
Treatment
PBS
F
M

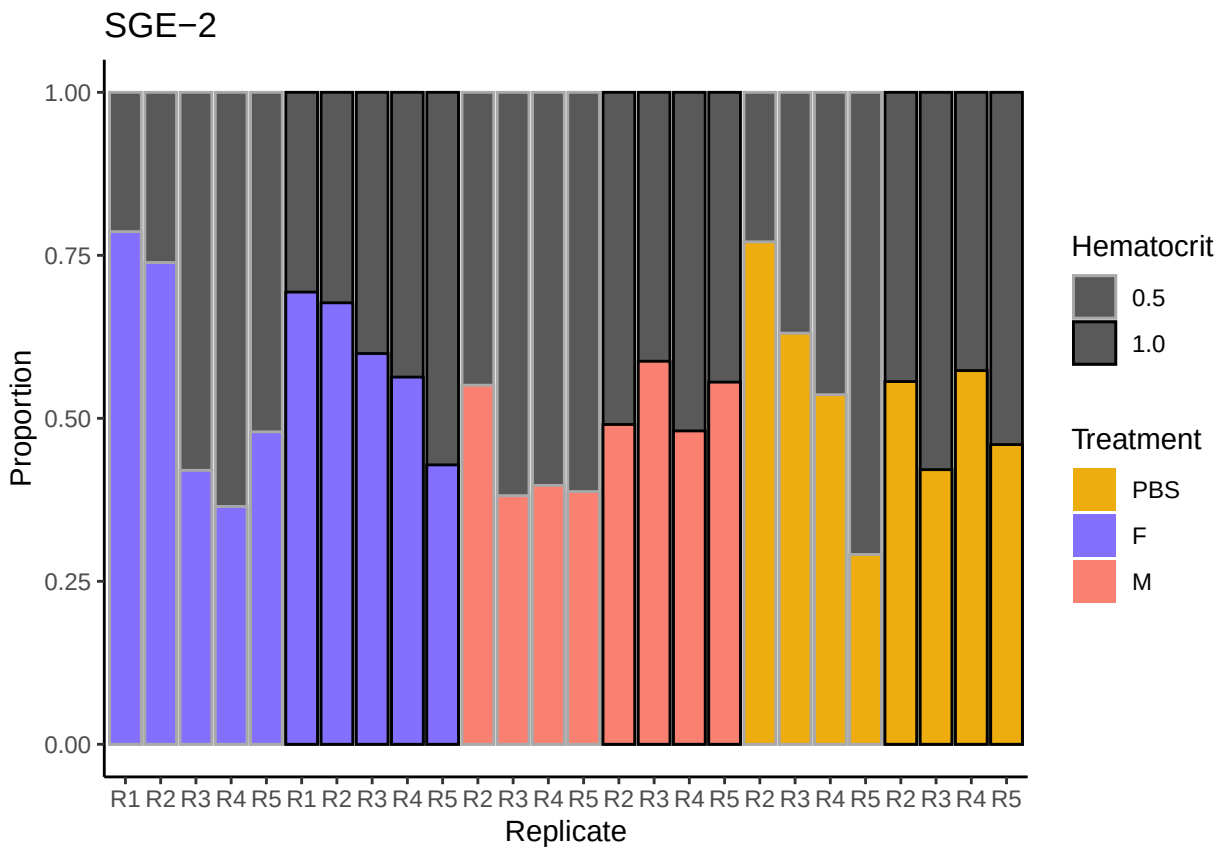

SGE−3
Counts
6e+05
4e+05
2e+05
0e+00
R1R2R3R4R5R1R2R3R4R5R1R2R3R4R5R1R2R3R4R5R1R2R3R4R5
Replicate
Hematocrit
0.5
1.0
Treatment
PBS
F
M

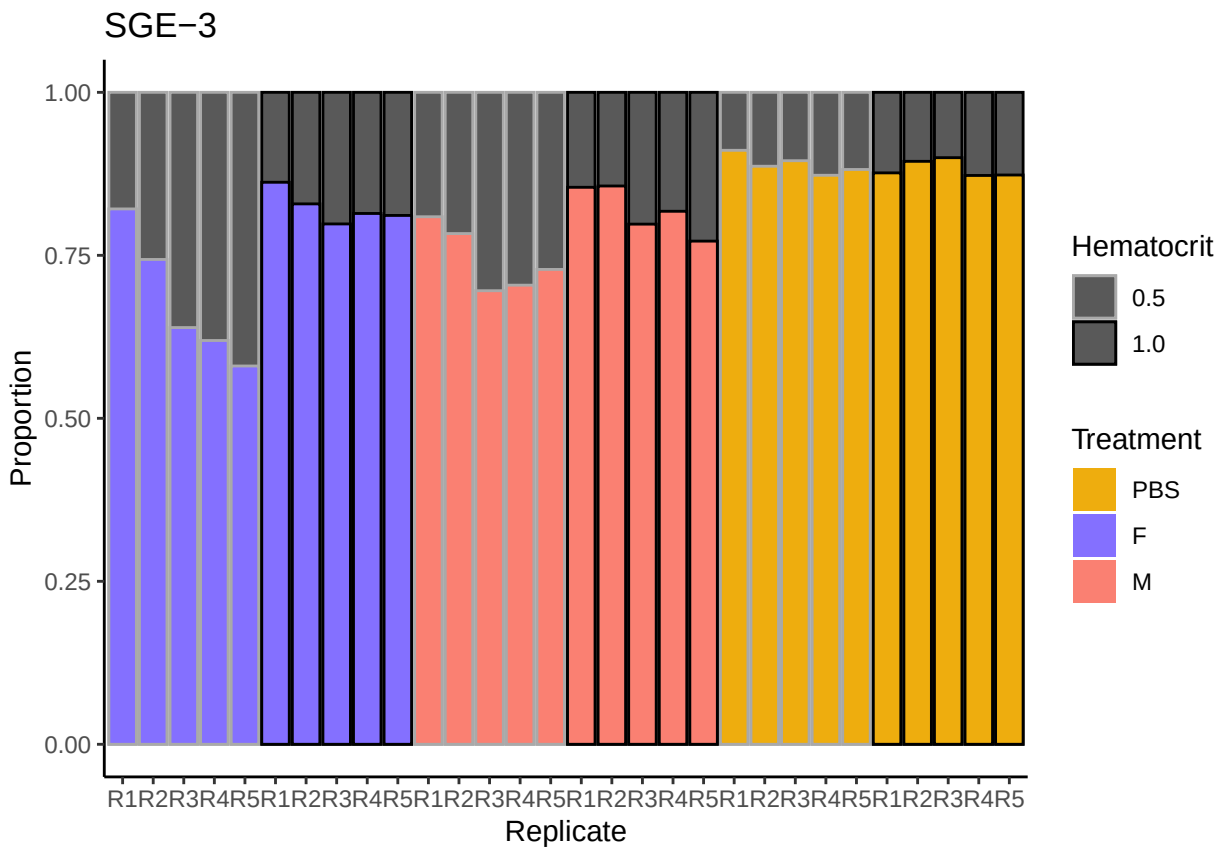


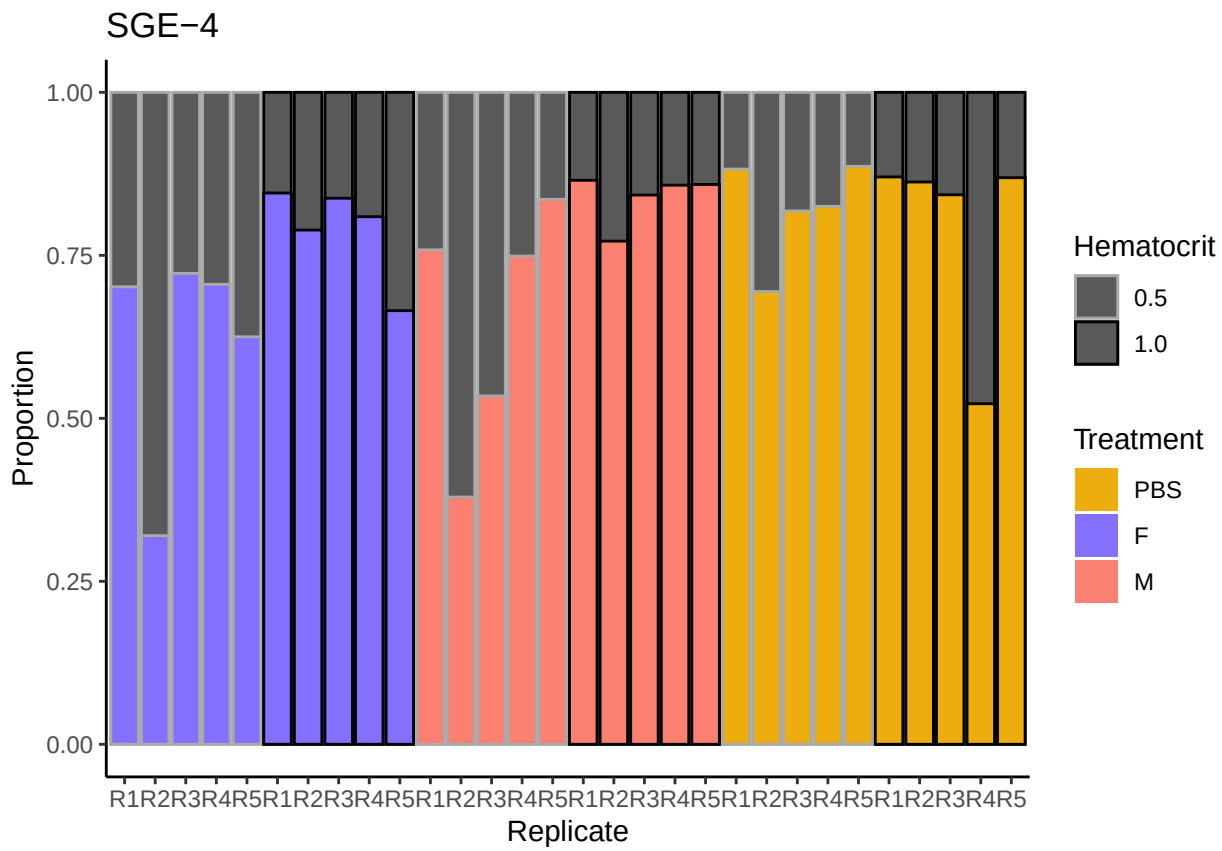
The next four plots below again shows the counts of all events (dark gray bars) compared to the singlets count (colored portion), but this time they are shown as proportions. Note that plate 2 seems to have abnormally low proportion of events counted as singlets, and plate 4, while mostly normal, has a few samples with abnormally low proportions, in particular row 2 of both the 0.5% hematocrit and female saliva treatment and the 0.5% hematocrit and male saliva treatment.

```
for (plate in unique(flow_data$Plate)) {
  dataset = flow_data[flow_data$Plate == plate,]

  print(ggplot() +
      geom_bar(data = dataset, aes(y=1, x=Treatments_FULL, color = Hematocrit), stat='identity') +
      geom_bar(data = dataset, aes(y=`Singlets Count`/`All Events Count`, x=Treatments_FULL, fill = Trea
      ggtitle(plate) +
      ylab("Proportion") + xlab("Replicate") +
      scale_fill_manual(values=c("darkgoldenrod2",
                                 "lightslateblue",
                                 "salmon")) +
      scale_color_manual(values=c("darkgray",
                                  "black")) +
      scale_x_discrete(labels= dataset$Replicate) +
      theme_classic()
  )
}
```


SGE−1
1.00
0.75
0.50
0.25
0.00
Proportion
R1 R2 R3 R4 R5 R1 R2 R3 R4 R5 R1 R2 R3 R4 R5 R1 R2 R4 R5 R2 R3 R4 R5 R1 R2 R3 R4 R5
Replicate
Hematocrit
0.5
1.0
Treatment
PBS
F
M


SGE−2
1.00
0.75
0.50
0.25
0.00
Proportion
R1 R2 R3 R4 R5 R1 R2 R3 R4 R5 R2 R3 R4 R5 R2 R3 R4 R5 R2 R3 R4 R5 R2 R3 R4 R5
Replicate
Hematocrit
0.5
1.0
Treatment
PBS
F
M


SGE−3
1.00
0.75
0.50
0.25
0.00
Proportion
R1R2R3R4R5R1R2R3R4R5R1R2R3R4R5R1R2R3R4R5R1R2R3R4R5
Replicate
Hematocrit
0.5
1.0
Treatment
PBS
F
M



In plates 1-3, we see that the number of events and singlets counts decreases with replicate/row number (i.e. order the samples were made, with 1 being the first made and 5 being the last). We also generally see that the proportion of events counted as singlets decreases with replicate ID.

Testing if percent of events captured as singlets affects proportion committed:

```
fit_events <- glmer(cbind(`iRBC_tdT+ Count`, `iRBCs Count` - `iRBC_tdT+ Count`) ~ Treatment + Hematocri
summary(fit_events)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: cbind(`iRBC_tdT+ Count`, `iRBCs Count` - `iRBC_tdT+ Count`) ~
##     Treatment + Hematocrit + PercentSinglets + (1 | Plate/Replicate)
##    Data: flow_data
##
##       AIC       BIC    logLik -2*log(L)  df.resid
##    1243.1    1262.2    -614.5    1229.1       107
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.3915 -1.1049  0.1134  1.0734  7.5142
##
## Random effects:
##  Groups          Name        Variance Std.Dev.
##  Replicate:Plate (Intercept) 0.0141   0.1188
##  Plate           (Intercept) 0.4380   0.6618
```

```
## Number of obs: 114, groups:  Replicate:Plate, 20; Plate, 4
##
## Fixed effects:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)     -3.245388   0.345819  -9.385  < 2e-16 ***
## TreatmentF       0.057463   0.017147   3.351 0.000805 ***
## TreatmentM       0.001366   0.018017   0.076 0.939565
## Hematocrit1.0   -0.111438   0.014258  -7.816 5.45e-15 ***
## PercentSinglets  0.368306   0.124070   2.969 0.002992 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) TrtmnF TrtmnM Hmt1.0
## TreatmentF  -0.086
## TreatmentM  -0.132  0.577
## Hematcrt1.0  0.034  0.009 -0.007
## PrcntSnglts -0.274  0.214  0.386 -0.230
```