

GSIN-1ug-flow-analysis 10/2 including median intensity

CTB

2024-10-21

```
data <- read.csv("/Users/csneffin/CSU Dropbox/Casey-Tyler Berezin/Cytek Data/GSIN-1ug-Data/With Median/
  rename(., "Sample"="X") %>%
  separate(., "Sample", sep = "-(?!mEGFP|Unsigned|Signed)", into=c("Date", "Exp", "Mixed", "Rack", "Gro
  select(Group:GFP..Cells.Median.GFP.A) %>%
  separate(., "Sample", sep = "\\s", extra = "merge", into=c("Well", "Sample")) %>%
  mutate(Group = as.factor(Group),
         Sample = factor(as.factor(Sample),
                        levels = c("Unstained (Cells)", "Unstained",
                                   "Green GFP (Cells)", "CMV-mEGFP-Unsigned",
                                   "CMV-mEGFP-Signed", "pV18", "pV19", "pV20")),
         Rep = as.integer(Rep),
         GFP..Cells.Mean.GFP.A = as.integer(GFP..Cells.Mean.GFP.A),
         GFP..Cells.Median.GFP.A = as.integer(GFP..Cells.Median.GFP.A),
         GFP..Cells.Min.GFP.A = as.integer(GFP..Cells.Min.GFP.A),
         GFP..Cells.Max.GFP.A = as.integer(GFP..Cells.Max.GFP.A)
        ) %>%
  mutate(Sample = str_replace_all(Sample, "-", "_")) #for multcompLetters
```

```
## Warning: Expected 7 pieces. Missing pieces filled with 'NA' in 2 rows [2, 3].
```

```
## Warning: There were 4 warnings in 'mutate()'.
## The first warning was:
## i In argument: 'GFP..Cells.Mean.GFP.A = as.integer(GFP..Cells.Mean.GFP.A)'.
## Caused by warning:
## ! NAs introduced by coercion
## i Run 'dplyr::last_dplyr_warnings()' to see the 3 remaining warnings.
```

```
head(data)
```

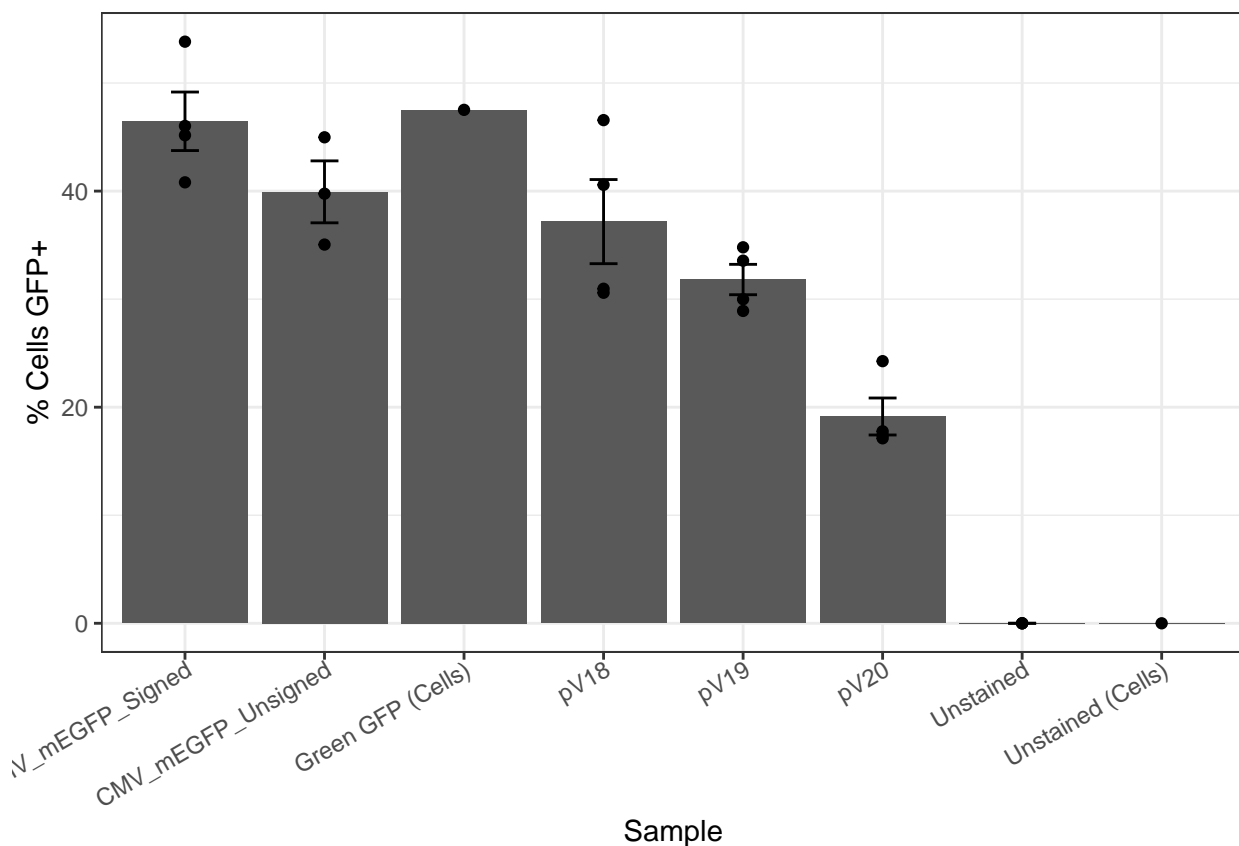
| | Group | Well | Sample | Rep | All.Events.Count |
|------|---------------------|---------------------------|--------------------|-----|------------------|
| ## 1 | GSIN | B2 | CMV_mEGFP_Unsigned | 2 | 10000 |
| ## 2 | Reference Group | A1 | Unstained (Cells) | NA | 10000 |
| ## 3 | Reference Group | A2 | Green GFP (Cells) | NA | 10000 |
| ## 4 | GSIN | A3 | CMV_mEGFP_Signed | 1 | 10000 |
| ## 5 | GSIN | A4 | pV18 | 1 | 10000 |
| ## 6 | GSIN | A5 | pV19 | 1 | 10000 |
| ## | All.Events...Parent | All.Events...Grand.Parent | All.Events...Total | | |
| ## 1 | 100 | | 100 | | 100 |
| ## 2 | 100 | | 100 | | 100 |
| ## 3 | 100 | | 100 | | 100 |

| | | | |
|------|----------------------|---------------------------|-------------------------------------|
| ## 4 | 100 | 100 | 100 |
| ## 5 | 100 | 100 | 100 |
| ## 6 | 100 | 100 | 100 |
| ## | All.Events.Count.uL | All.Events.Mean.GFP.A | All.Events.Max.GFP.A |
| ## 1 | 648.93 | 546934 | 5947098 |
| ## 2 | 93.27 | 2 | 997 |
| ## 3 | 604.23 | 622610 | 5879801 |
| ## 4 | 558.97 | 519580 | 5863031 |
| ## 5 | 465.12 | 543127 | 5840190 |
| ## 6 | 606.43 | 344771 | 5862167 |
| ## | All.Events.Min.GFP.A | All.Events.Median.GFP.A | Cells.Count Cells...Parent |
| ## 1 | -305 | 19075 | 6086 60.86 |
| ## 2 | -795 | 2 | 6908 69.08 |
| ## 3 | -317 | 51150 | 6969 69.69 |
| ## 4 | -219 | 20162 | 7373 73.73 |
| ## 5 | -270 | 35537 | 6704 67.04 |
| ## 6 | -455 | 4673 | 6758 67.58 |
| ## | Cells...Grand.Parent | Cells...Total | Cells.Count.uL Cells.Mean.GFP.A |
| ## 1 | 60.86 | 60.86 | 394.94 324837 |
| ## 2 | 69.08 | 69.08 | 64.43 -1 |
| ## 3 | 69.69 | 69.69 | 421.09 446286 |
| ## 4 | 73.73 | 73.73 | 412.13 414830 |
| ## 5 | 67.04 | 67.04 | 311.81 392441 |
| ## 6 | 67.58 | 67.58 | 409.82 231591 |
| ## | Cells.Max.GFP.A | Cells.Min.GFP.A | Cells.Median.GFP.A GFP..Cells.Count |
| ## 1 | 4727011 | -305 | 1902 2420 |
| ## 2 | 599 | -635 | -1 0 |
| ## 3 | 4973422 | -284 | 17314 3312 |
| ## 4 | 4665630 | -219 | 10659 3394 |
| ## 5 | 4856951 | -233 | 16684 3122 |
| ## 6 | 4628508 | -323 | 1808 2352 |
| ## | GFP..Cells...Parent | GFP..Cells...Grand.Parent | GFP..Cells...Total |
| ## 1 | 39.76 | 24.20 | 24.20 |
| ## 2 | 0.00 | 0.00 | 0.00 |
| ## 3 | 47.52 | 33.12 | 33.12 |
| ## 4 | 46.03 | 33.94 | 33.94 |
| ## 5 | 46.57 | 31.22 | 31.22 |
| ## 6 | 34.80 | 23.52 | 23.52 |
| ## | GFP..Cells.Count.uL | GFP..Cells.Mean.GFP.A | GFP..Cells.Max.GFP.A |
| ## 1 | 157.04 | 790364 | 4175170 |
| ## 2 | 0.00 | NA | NA |
| ## 3 | 200.12 | 899894 | 4161831 |
| ## 4 | 189.71 | 875486 | 4103897 |
| ## 5 | 145.21 | 799410 | 4186844 |
| ## 6 | 142.63 | 634359 | 4146679 |
| ## | GFP..Cells.Min.GFP.A | GFP..Cells.Median.GFP.A | |
| ## 1 | 29032 | 521636 | |
| ## 2 | NA | NA | |
| ## 3 | 28906 | 643616 | |
| ## 4 | 28882 | 602303 | |
| ## 5 | 28887 | 495711 | |
| ## 6 | 28892 | 299508 | |

```
ggplot(data = data, mapping = aes(x = Sample, y = GFP..Cells...Parent)) +
  geom_bar(position = "dodge", stat = "summary", fun.y = "mean") +
  stat_summary(geom="errorbar", width=0.2) +
  geom_point() +
  scale_y_continuous(name="% Cells GFP+") +
  theme_bw() %+replace%
  theme(axis.text.x = element_text(angle = 30, vjust = 1, hjust=1))
```

```
## Warning in geom_bar(position = "dodge", stat = "summary", fun.y = "mean"):
## Ignoring unknown parameters: 'fun.y'
```

```
## No summary function supplied, defaulting to 'mean_se()'
## No summary function supplied, defaulting to 'mean_se()'
```



```
ggplot(data = data, mapping = aes(x = Sample, y = GFP..Cells.Mean.GFP.A)) +
  geom_bar(position = "dodge", stat = "summary", fun.y = "mean") +
  stat_summary(geom="errorbar", width=0.2) +
  geom_point() +
  scale_y_continuous(name="Mean GFP Intensity") +
  theme_bw() %+replace%
  theme(axis.text.x = element_text(angle = 30, vjust = 1, hjust=1))
```

```
## Warning in geom_bar(position = "dodge", stat = "summary", fun.y = "mean"):
## Ignoring unknown parameters: 'fun.y'
```

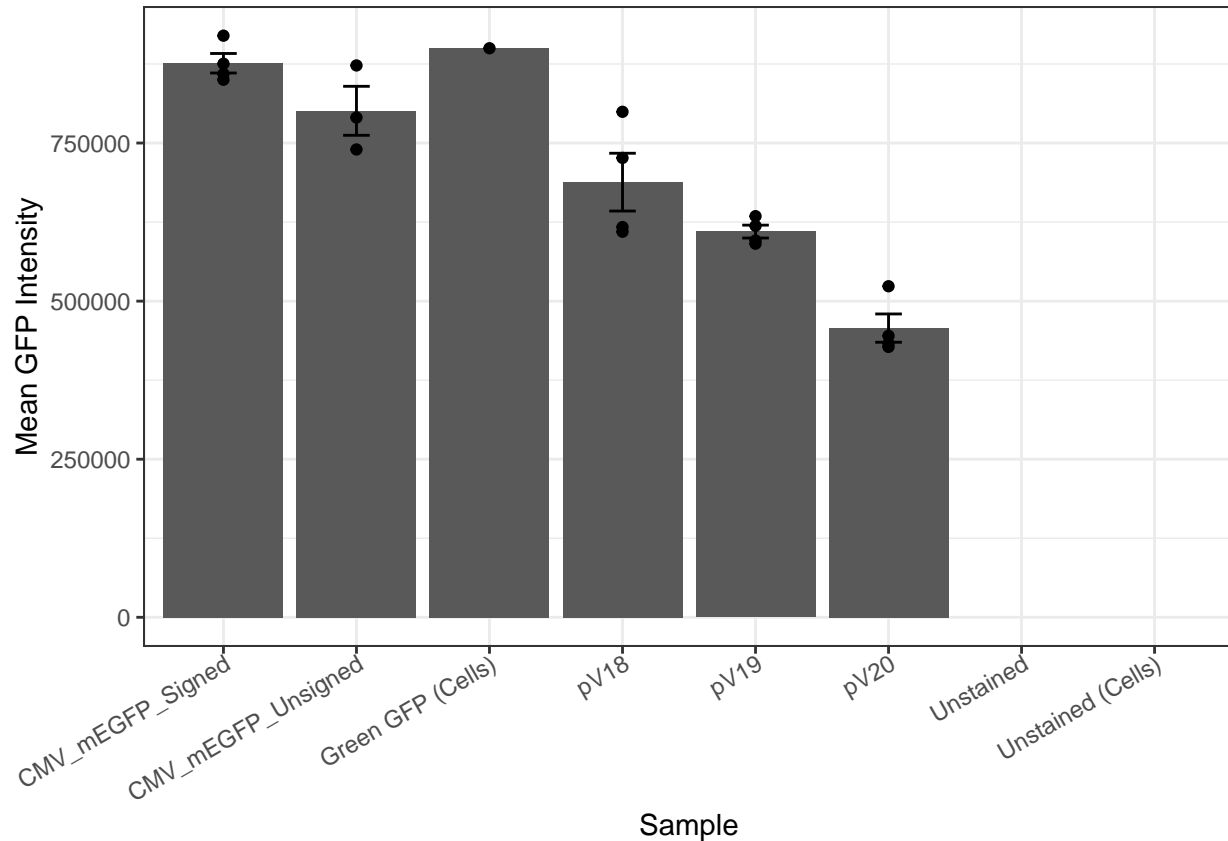
```
## Warning: Removed 4 rows containing non-finite values ('stat_summary()').

## No summary function supplied, defaulting to 'mean_se()'

## Warning: Removed 4 rows containing non-finite values ('stat_summary()').

## No summary function supplied, defaulting to 'mean_se()'

## Warning: Removed 4 rows containing missing values ('geom_point()').
```



```
ggplot(data = data, mapping = aes(x = Sample, y = GFP..Cells.Median.GFP.A)) +
  geom_bar(position = "dodge", stat = "summary", fun.y = "mean") +
  stat_summary(geom="errorbar", width=0.2) +
  geom_point() +
  scale_y_continuous(name="Median GFP Intensity") +
  theme_bw() %+replace%
  theme(axis.text.x = element_text(angle = 30, vjust = 1, hjust=1))
```

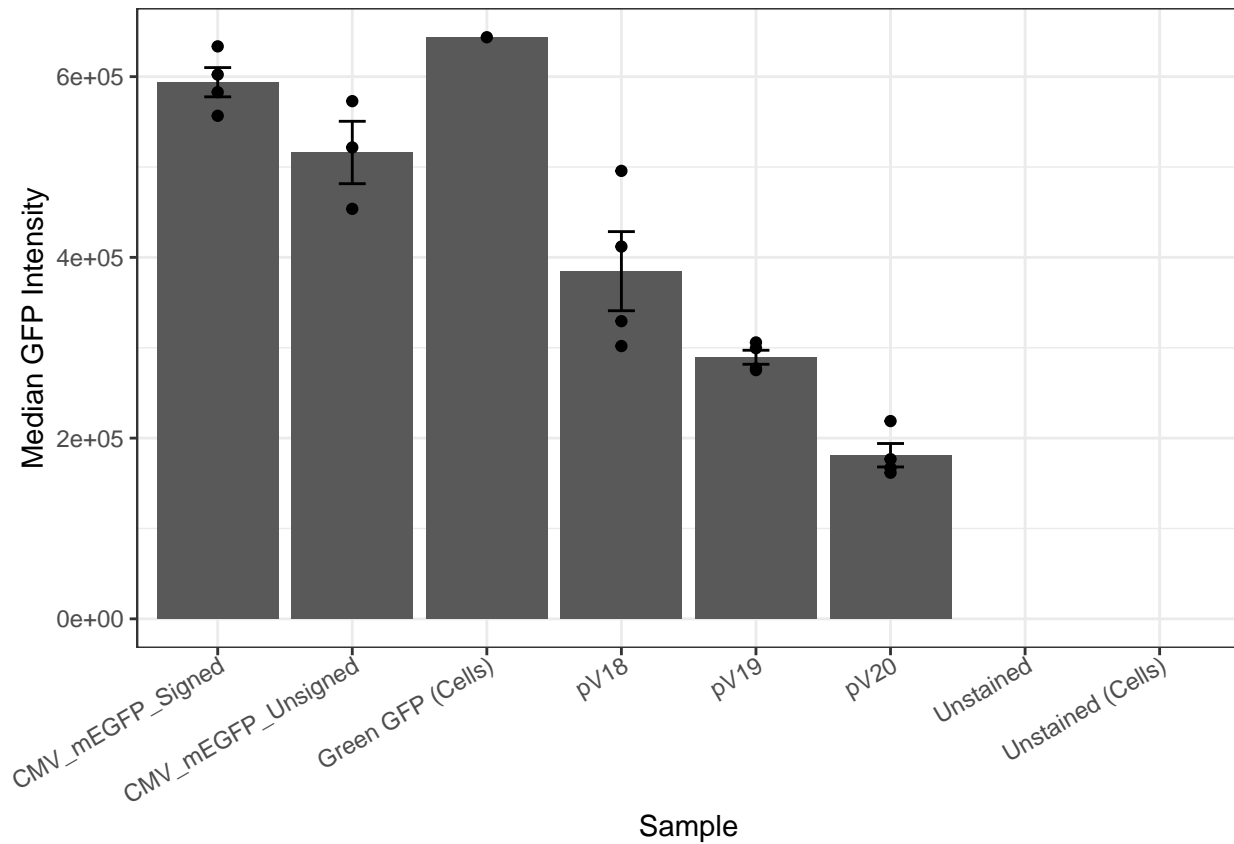
```
## Warning in geom_bar(position = "dodge", stat = "summary", fun.y = "mean"): Ignoring unknown parameters:
## Removed 4 rows containing non-finite values ('stat_summary()').

## No summary function supplied, defaulting to 'mean_se()'

## Warning: Removed 4 rows containing non-finite values ('stat_summary()').
```

```
## No summary function supplied, defaulting to 'mean_se()'
```

```
## Warning: Removed 4 rows containing missing values ('geom_point()').
```



```
data %>% group_by(Sample) %>% summarise(
  mean_int = mean(GFP..Cells.Mean.GFP.A),
  sd_mean_int = sd(GFP..Cells.Mean.GFP.A),
  med_int = median(GFP..Cells.Median.GFP.A),
  min_int = min(GFP..Cells.Min.GFP.A),
  max_int = max(GFP..Cells.Max.GFP.A),
  mean_perc = mean(GFP..Cells...Parent),
  sd_perc = sd(GFP..Cells...Parent),
  min_perc = min(GFP..Cells...Parent),
  max_perc = max(GFP..Cells...Parent)
)
```

```
## # A tibble: 8 x 10
##   Sample mean_int sd_mean_int med_int min_int max_int mean_perc sd_perc min_perc
##   <chr>      <dbl>      <dbl>  <dbl>  <int>  <int>      <dbl>  <dbl>  <dbl>
## 1 CMV_m~  876197    30881. 592564  28880 4187964    46.5    5.42   40.8
## 2 CMV_m~  800976    67078. 521636  28983 4180302    39.9    4.97   35.0
## 3 Green~  899894      NA 643616  28906 4161831    47.5    NA    47.5
## 4 Unsta~    NA      NA    NA      NA      NA      0      0      0
## 5 Unsta~    NA      NA    NA      NA      NA      0    NA      0
## 6 pV18    688178    91368. 370682.  28880 4186844    37.2    7.79   30.6
```

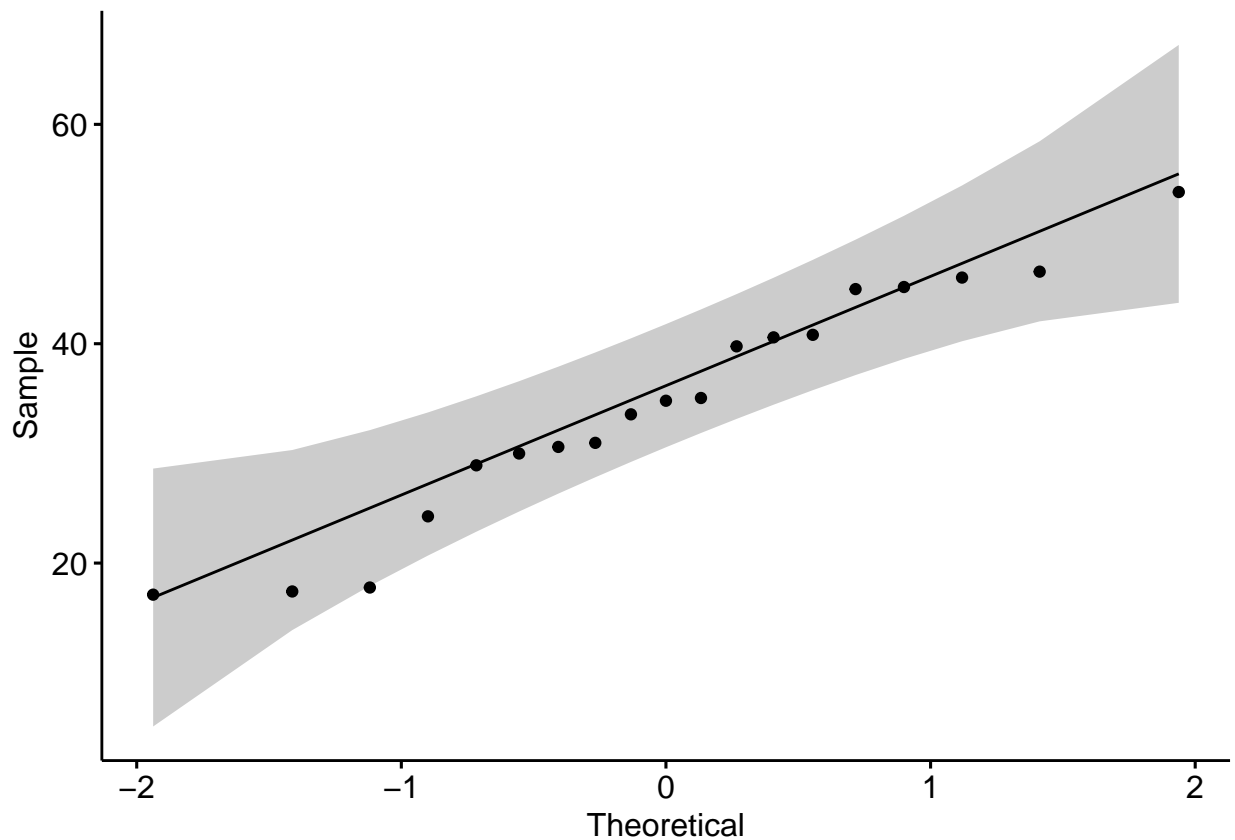
```
## 7 pV19      609985      20377. 288360.   28887 4193757      31.8    2.81    28.9
## 8 pV20      457316      44747. 171858.   28951 4187061      19.1    3.42    17.1
## # i 1 more variable: max_perc <dbl>
```

```
sample_data_only <- data %>% filter(Group != "Reference Group")

green_data_only <- data %>% filter(Sample != "Unstained")

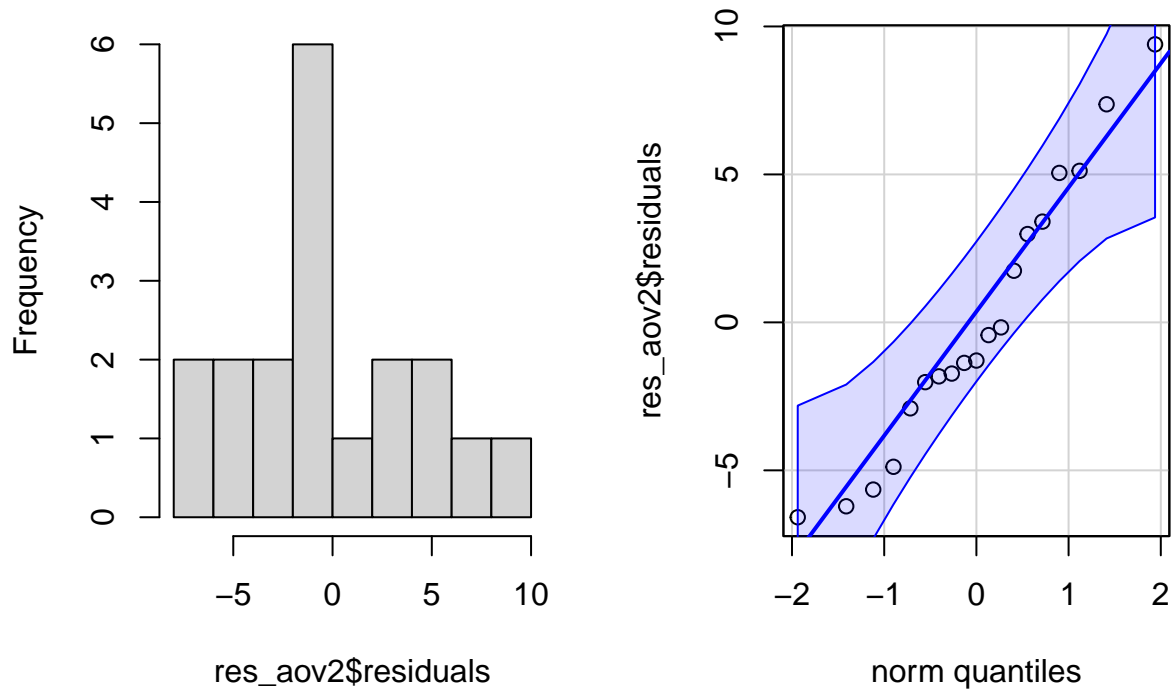
green_samples_only <- sample_data_only %>% filter(Sample != "Unstained")
```

```
#Checking ANOVA assumptions for % GFP cells
ggqqplot(green_samples_only$GFP..Cells...Parent)
```



```
res_aov2 <- aov(GFP..Cells...Parent ~ Sample,
  data = green_samples_only
)
par(mfrow = c(1, 2)) # combine plots
# histogram
hist(res_aov2$residuals)
# QQ-plot
qqPlot(res_aov2$residuals,
  id = FALSE # id = FALSE to remove point identification
)
```

Histogram of res_aov2\$residual

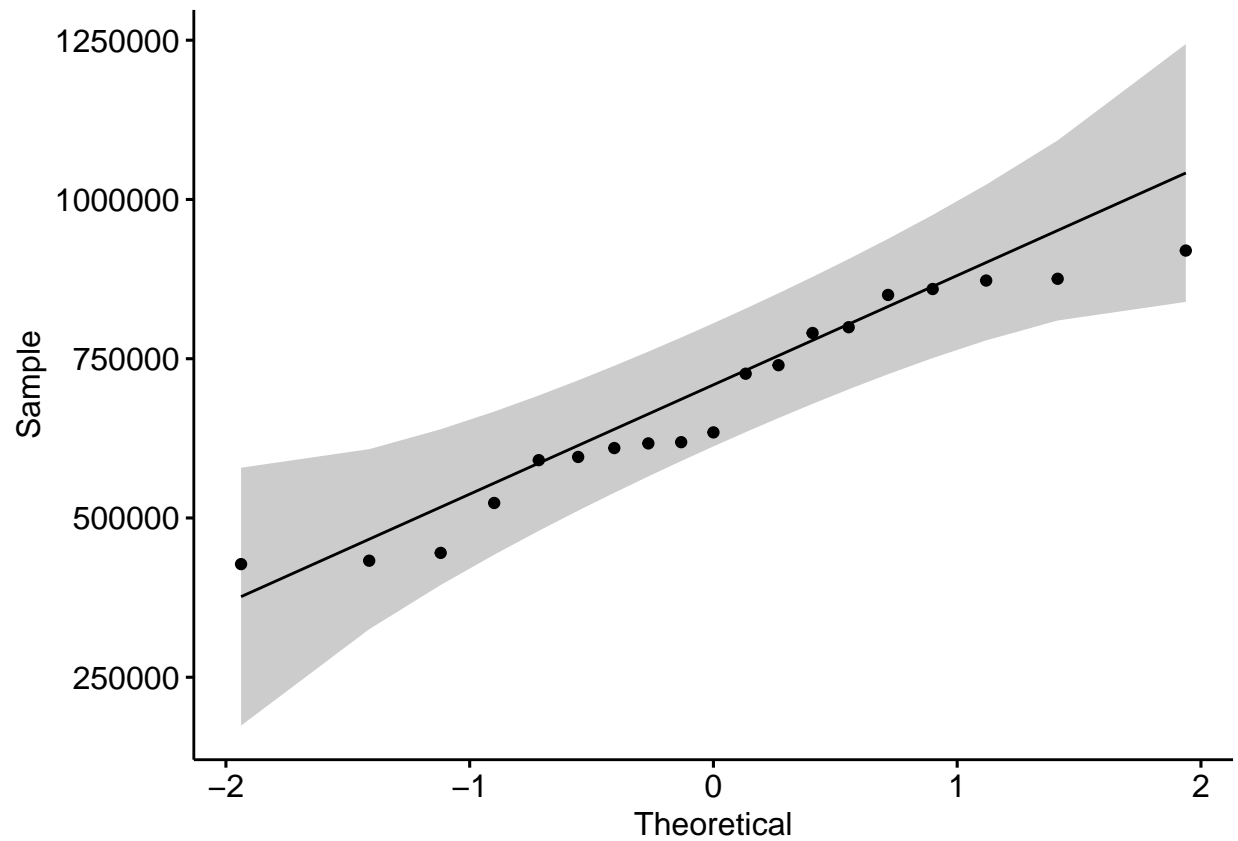


```
shapiro.test(green_samples_only$GFP..Cells...Parent)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  green_samples_only$GFP..Cells...Parent
## W = 0.9552, p-value = 0.4819
```

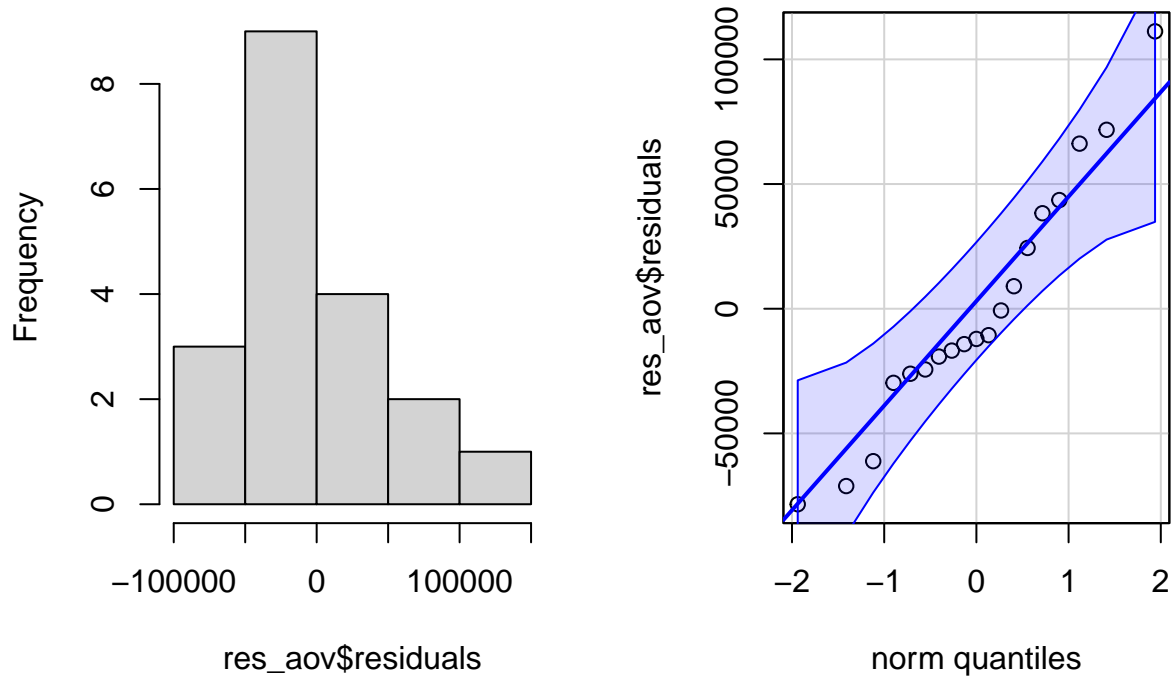
```
#Normally distributed!
```

```
#Checking ANOVA assumptions for Mean Intensity
ggqqplot(green_samples_only$GFP..Cells.Mean.GFP.A)
```



```
res_aov <- aov(GFP..Cells.Mean.GFP.A ~ Sample,
  data = green_samples_only
)
par(mfrow = c(1, 2)) # combine plots
# histogram
hist(res_aov$residuals)
# QQ-plot
qqPlot(res_aov$residuals,
  id = FALSE # id = FALSE to remove point identification
)
```


Histogram of res_aov\$residuals

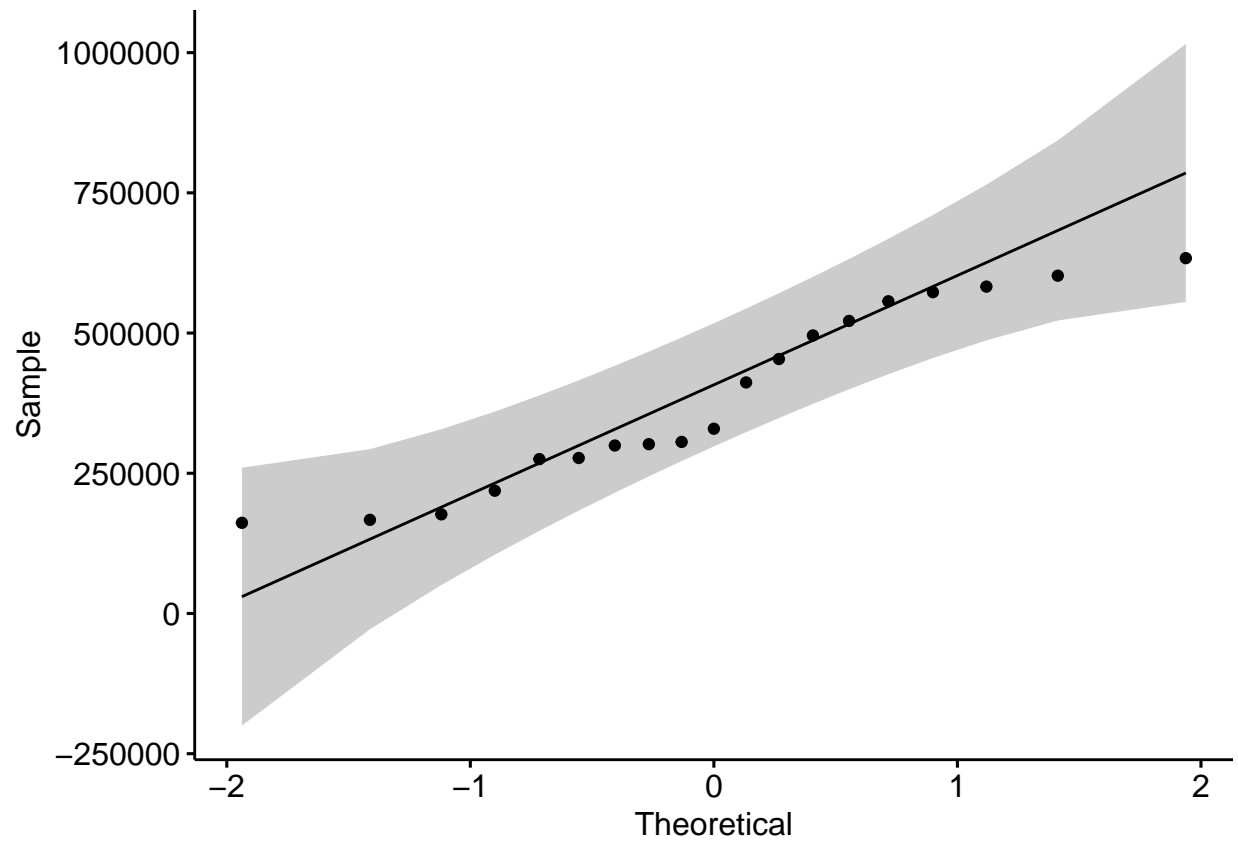


```
shapiro.test(green_samples_only$GFP..Cells.Mean.GFP.A)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  green_samples_only$GFP..Cells.Mean.GFP.A
## W = 0.93011, p-value = 0.174
```

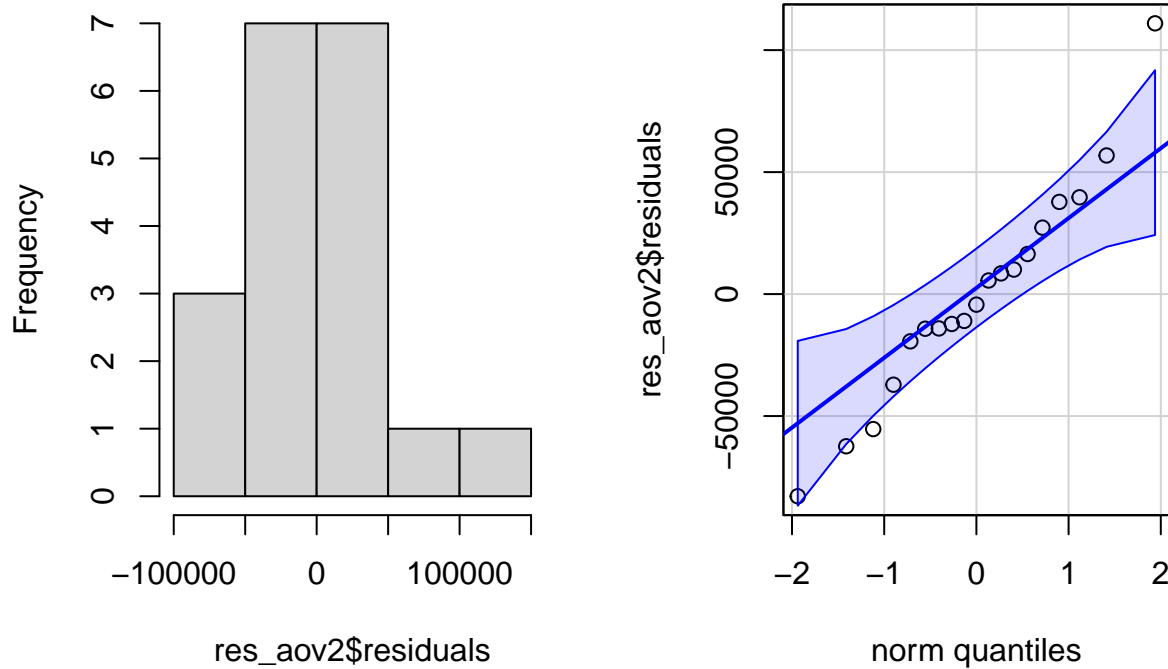
```
#Normal
```

```
#Checking ANOVA assumptions for Median Intensity
ggqqplot(green_samples_only$GFP..Cells.Median.GFP.A)
```



```
res_aov2 <- aov(GFP..Cells.Median.GFP.A ~ Sample,
  data = green_samples_only
)
par(mfrow = c(1, 2)) # combine plots
# histogram
hist(res_aov2$residuals)
# QQ-plot
qqPlot(res_aov2$residuals,
  id = FALSE # id = FALSE to remove point identification
)
```

Histogram of res_aov2\$residual



```
shapiro.test(green_samples_only$GFP..Cells.Median.GFP.A)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  green_samples_only$GFP..Cells.Median.GFP.A
## W = 0.91579, p-value = 0.0946
```

```
#normal
```

```
#Checking if % GFP cells is stat diff across groups
```

```
#Comparing all sample groups only to unsigned control group
```

```
DunnettTest(green_samples_only$GFP..Cells...Parent, green_samples_only$Sample, control="CMV_mEGFP_Unsigned")
```

```
##
##  Dunnett's test for comparing several treatments with a control :
##    95% family-wise confidence level
##
## $CMV_mEGFP_Unsigned
##
##          diff      lwr.ci      upr.ci      pval
## CMV_mEGFP_Signed-CMV_mEGFP_Unsigned  6.530 -4.285666 17.345666 0.31213
## pV18-CMV_mEGFP_Unsigned             -2.755 -13.570666  8.060666 0.87902
## pV19-CMV_mEGFP_Unsigned             -8.115 -18.930666  2.700666 0.16670
```

```
## pV20-CMV_mEGFP_Unsigned          -20.790 -31.605666 -9.974334 0.00032 ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Checking if mean intensity is stat diff across groups
```

```
#Comparing all sample groups only to unsigned control group
```

```
DunnettTest(green_samples_only$GFP..Cells.Mean.GFP.A, green_samples_only$Sample, control="CMV_mEGFP_Uns
```

```
##
## Dunnett's test for comparing several treatments with a control :
## 95% family-wise confidence level
##
## $CMV_mEGFP_Unsigned
##
##          diff      lwr.ci      upr.ci      pval
## CMV_mEGFP_Signed-CMV_mEGFP_Unsigned  75221 -41649.47 192091.469 0.2648
## pV18-CMV_mEGFP_Unsigned             -112798 -229668.47  4072.469 0.0594 .
## pV19-CMV_mEGFP_Unsigned             -190991 -307861.47 -74120.531 0.0020 **
## pV20-CMV_mEGFP_Unsigned             -343660 -460530.47 -226789.531 9.2e-07 ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Performing ANOVA for all pairwise comparisons
```

```
#Doesn't work with unstained cells where y=0
```

```
int_aov <- aov(GFP..Cells.Mean.GFP.A ~ Sample, data = green_data_only)
```

```
summary(int_aov)
```

```
##          Df      Sum Sq  Mean Sq F value    Pr(>F)
## Sample      5 4.618e+11 9.237e+10   29.29 6.09e-07 ***
## Residuals   14 4.416e+10 3.154e+09
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 1 observation deleted due to missingness
```

```
tukey <- TukeyHSD(int_aov)
```

```
# Extract p-values from Tukey's HSD
```

```
tukey_pvalues <- tukey$Sample[, "p adj"]
```

```
# Generate groupings (a, b, c, etc.) based on Tukey HSD
```

```
group_letters <- multcompLetters(tukey_pvalues)
```

```
print(group_letters$Letters)
```

```
## CMV_mEGFP_Unsigned  Green GFP (Cells)          pV18          pV19
##          "ab"          "a"          "bc"          "c"
##          pV20    CMV_mEGFP_Signed
##          "d"          "a"
```

```

#Checking if median intensity is stat diff across groups
#Comparing all sample groups only to unsigned control group
DunnettTest(green_samples_only$GFP..Cells.Median.GFP.A, green_samples_only$Sample, control="CMV_mEGFP_U

```

```

##
##   Dunnett's test for comparing several treatments with a control :
##     95% family-wise confidence level
##
## $CMV_mEGFP_Unsigned
##               diff      lwr.ci      upr.ci      pval
## CMV_mEGFP_Signed-CMV_mEGFP_Unsigned  77763.25 -27802.32  183328.82  0.1768
## pV18-CMV_mEGFP_Unsigned             -131330.75 -236896.32 -25765.18  0.0140
## pV19-CMV_mEGFP_Unsigned             -226595.00 -332160.57 -121029.43  9.4e-05
## pV20-CMV_mEGFP_Unsigned             -335000.75 -440566.32 -229435.18  2.1e-07
##
## CMV_mEGFP_Signed-CMV_mEGFP_Unsigned
## pV18-CMV_mEGFP_Unsigned              *
## pV19-CMV_mEGFP_Unsigned             ***
## pV20-CMV_mEGFP_Unsigned             ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

#Performing ANOVA for all pairwise comparisons
#Doesn't work with unstained cells where y=0
int_aov <- aov(GFP..Cells.Median.GFP.A ~ Sample, data = green_data_only)

summary(int_aov)

```

```

##              Df      Sum Sq   Mean Sq F value    Pr(>F)
## Sample         5 4.916e+11  9.831e+10   38.2 1.12e-07 ***
## Residuals     14 3.603e+10  2.573e+09
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 1 observation deleted due to missingness

```

```

tukey <- TukeyHSD(int_aov)

# Extract p-values from Tukey's HSD
tukey_pvalues <- tukey$Sample[, "p adj"]

# Generate groupings (a, b, c, etc.) based on Tukey HSD
group_letters <- multcompLetters(tukey_pvalues)
print(group_letters$Letters)

```

```

## CMV_mEGFP_Unsigned  Green GFP (Cells)          pV18          pV19
##                   "a"                   "a"          "b"          "bc"
##                   pV20  CMV_mEGFP_Signed
##                   "c"                   "a"

```