

# GSIN-1ug-flow-analysis 10/2 regated

CTB

2024-10-24

```
data <- read.csv("20241002-GSIN-Unmixed_Experiment_016-Unmixed.CSV") %>%
  rename(., "Sample"="X") %>%
  separate(., "Sample", sep = "-(?!mEGFP|Unsigned|Signed)", into=c("Date", "Exp", "Mixed", "Rack", "Gron
  select(Group:GFP..Cells.Min.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.Min.GFP.A = as.integer(GFP..Cells.Min.GFP.A),
         GFP..Cells.Max.GFP.A = as.integer(GFP..Cells.Max.GFP.A)
  )
```

```
## Warning: Expected 7 pieces. Missing pieces filled with 'NA' in 2 rows [1, 2].
```

```
## Warning: There were 3 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 2 remaining warnings.
```

```
head(data)
```

##	Group	Well	Sample	Rep	All.Events.Count
## 1	Reference	Group A1	Unstained (Cells)	NA	10000
## 2	Reference	Group A2	Green GFP (Cells)	NA	10000
## 3	GSIN	A3	CMV-mEGFP-Signed	1	10000
## 4	GSIN	A4	pV18	1	10000
## 5	GSIN	A5	pV19	1	10000
## 6	GSIN	A6	pV20	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	93.27	2	997
## 2	604.23	622610	5879801
## 3	558.97	519580	5863031
## 4	465.12	543127	5840190
## 5	606.43	344771	5862167
## 6	436.30	128178	5907923
##	All.Events.Min.GFP.A	All.Events.Median.GFP.A	Cells.Count Cells...Parent
## 1	-795	2	7300 73.00
## 2	-317	51150	7249 72.49
## 3	-219	20162	7851 78.51
## 4	-270	35537	7088 70.88
## 5	-455	4673	7060 70.60
## 6	-434	577	7440 74.40
##	Cells...Grand.Parent	Cells...Total	Cells.Count.uL Cells.Mean.GFP.A
## 1	73.00	73.00	68.09 -1
## 2	72.49	72.49	438.01 459455
## 3	78.51	78.51	438.85 428885
## 4	70.88	70.88	329.67 403776
## 5	70.60	70.60	428.14 240984
## 6	74.40	74.40	324.61 83777
##	Cells.Max.GFP.A	Cells.Min.GFP.A	Cells.Median.GFP.A GFP..Cells.Count
## 1	599	-795	-1 0
## 2	4973422	-284	21685 4041
## 3	4665630	-219	11462 4150
## 4	4842906	-233	18037 3964
## 5	4747452	-455	1800 3198
## 6	4716101	-434	441 2161
##	GFP..Cells...Parent	GFP..Cells...Grand.Parent	GFP..Cells...Total
## 1	0.00	0.00	0.00
## 2	55.75	40.41	40.41
## 3	52.86	41.50	41.50
## 4	55.93	39.64	39.64
## 5	45.30	31.98	31.98
## 6	29.05	21.61	21.61
##	GFP..Cells.Count.uL	GFP..Cells.Mean.GFP.A	GFP..Cells.Max.GFP.A
## 1	0.00	NA	NA
## 2	244.17	786394	4161831
## 3	231.97	781278	4103897
## 4	184.37	679026	4186844
## 5	193.94	503671	4146679
## 6	94.28	266954	4150323
##	GFP..Cells.Min.GFP.A		
## 1	NA		
## 2	3847		
## 3	3842		
## 4	3902		
## 5	3841		
## 6	3840		

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

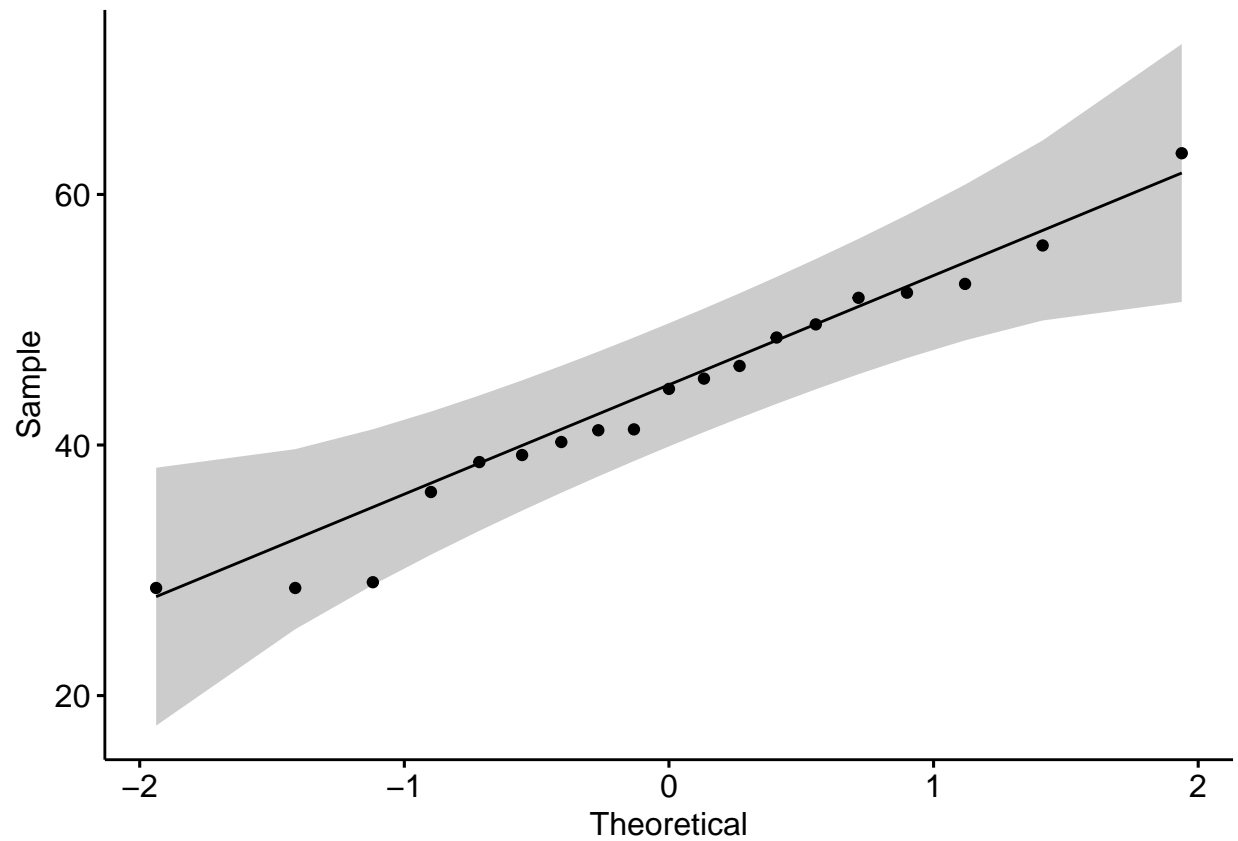
```
green_samples_only <- sample_data_only %>% filter(Sample != "Unstained")
```

```
green_samples_only %>% group_by(Sample) %>% summarise(
  n = n(),
  mean_int = mean(GFP..Cells.Mean.GFP.A),
  sd_mean_int = sd(GFP..Cells.Mean.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: 5 x 10
##   Sample      n mean_int sd_mean_int min_int max_int mean_perc sd_perc min_perc
##   <fct>   <int>   <dbl>       <dbl>   <int>   <int>       <dbl>   <dbl>   <dbl>
## 1 CMV-mEG~    3  703270.    63214.    3838  4180302    46.4     5.25    41.2
## 2 CMV-mEG~    4  766754.    29945.    3842  4191918    54.2     6.33    48.6
## 3 pV18        4  559704.    92895.    3837  4186844    46.1     8.15    38.6
## 4 pV19        4  464873.    27664.    3841  4193757    42.5     2.85    39.2
## 5 pV20        4  293360.    44568.    3838  4187061    30.6     3.76    28.6
## # i 1 more variable: max_perc <dbl>
```

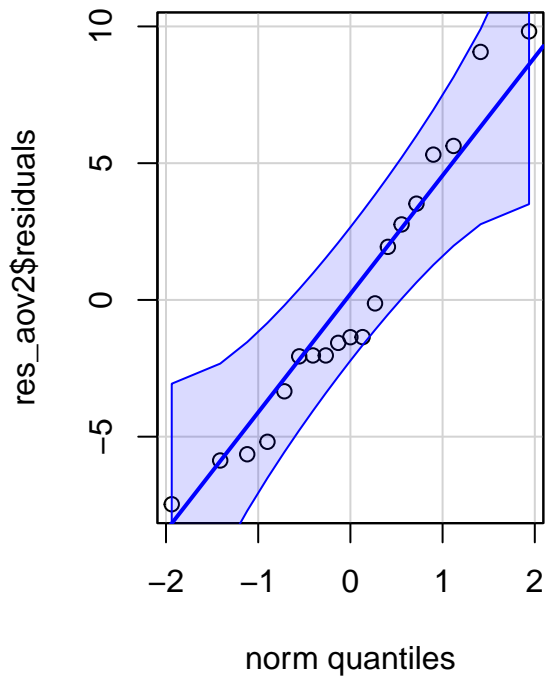
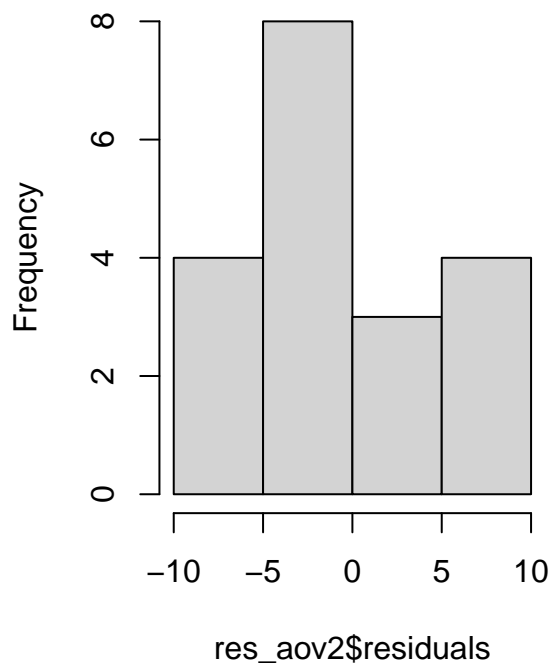
Using only green samples for stats - Dunnett's test compares to control (unsigned) Samples with values 0 (unstained) make ANOVA calculations impossible

```
#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

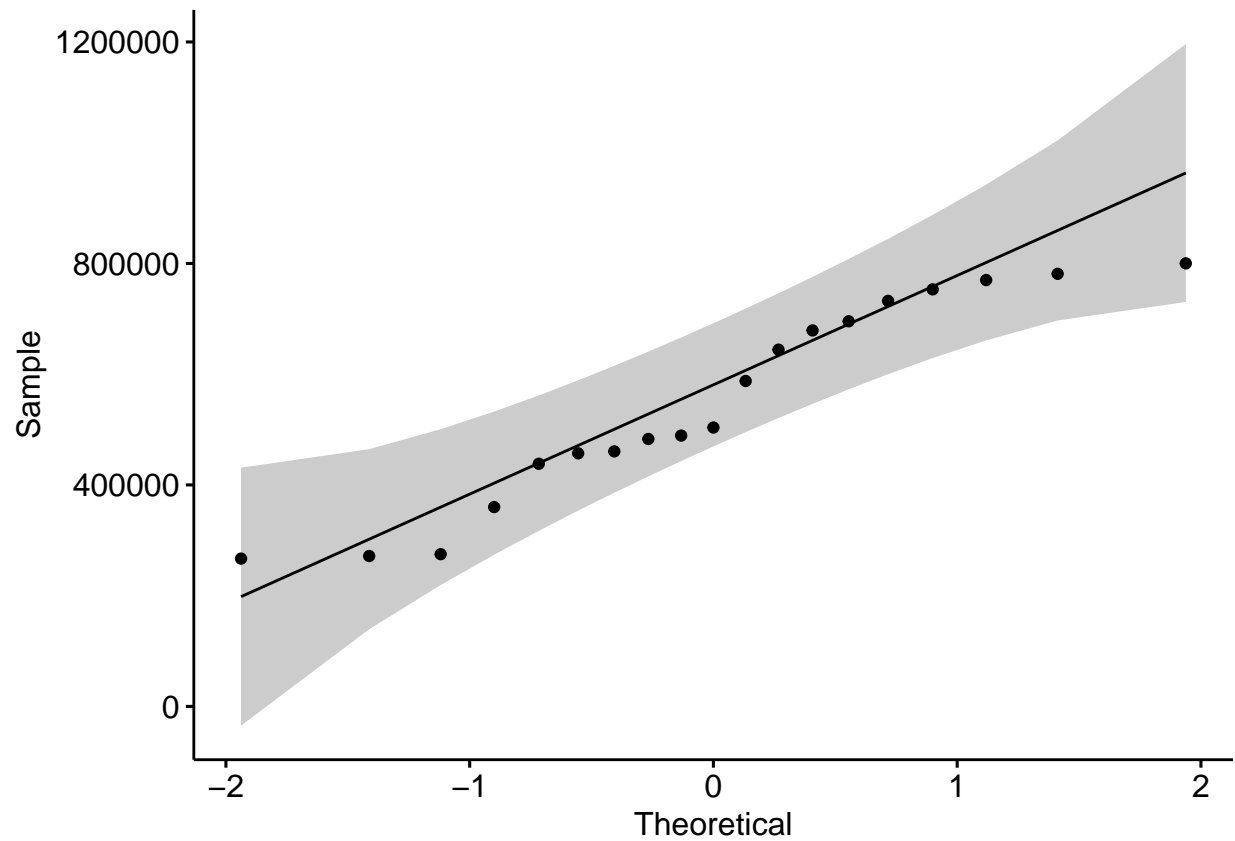


```
shapiro.test(green_samples_only$GFP..Cells...Parent)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: green_samples_only$GFP..Cells...Parent  
## W = 0.96768, p-value = 0.7292
```

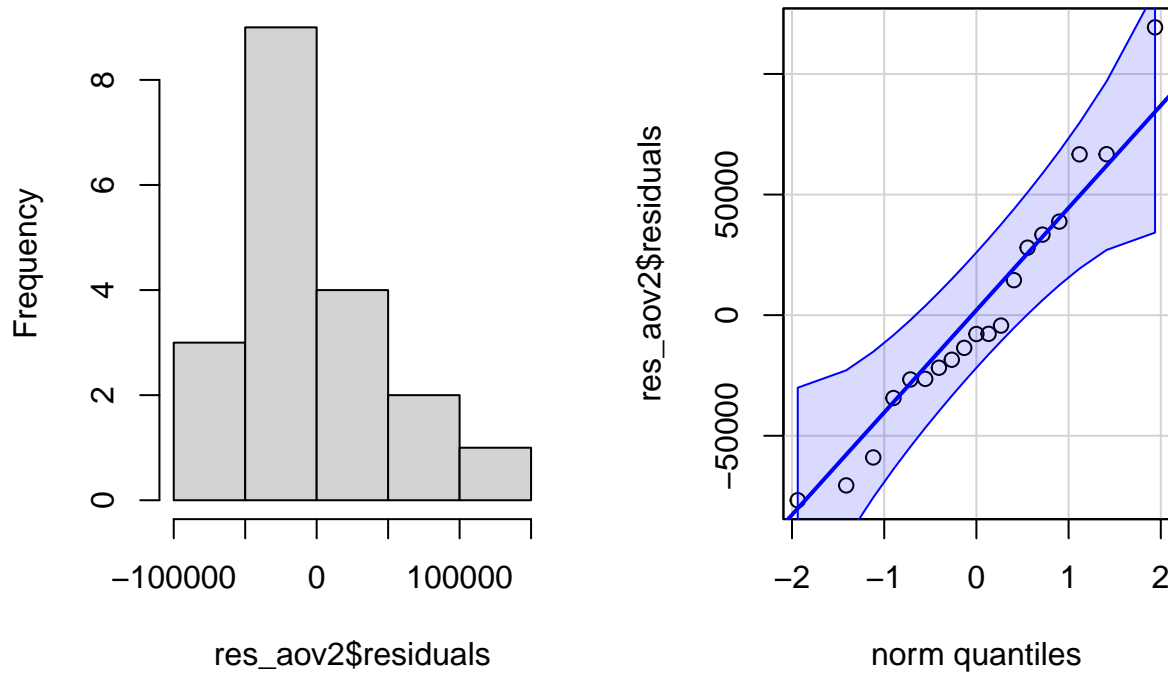
```
#Normally distributed!
```

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



```
res_aov2 <- aov(GFP..Cells.Mean.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.Mean.GFP.A)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  green_samples_only$GFP..Cells.Mean.GFP.A
## W = 0.92015, p-value = 0.1139
```

```
#Normally distributed!
```

```
#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  7.7858333 -3.896242 19.467909 0.2406
## pV18-CMV-mEGFP-Unsigned             -0.3266667 -12.008742 11.355409 1.0000
## pV19-CMV-mEGFP-Unsigned             -3.8966667 -15.578742  7.785409 0.7542
```

```
## pV20-CMV-mEGFP-Unsigned          -15.8166667 -27.498742 -4.134591 0.0077 **
##
## ---
## 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  63483.83 -54268.81  181236.48 0.40052
## pV18-CMV-mEGFP-Unsigned             -143565.17 -261317.81 -25812.52 0.01620
## pV19-CMV-mEGFP-Unsigned             -238396.92 -356149.56 -120644.27 0.00019
## pV20-CMV-mEGFP-Unsigned             -409909.92 -527662.56 -292157.27 2.6e-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
```

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

```
##              Df      Sum Sq   Mean Sq F value    Pr(>F)
## Sample         4 5.512e+11  1.378e+11   43.04 1.02e-07 ***
## Residuals     14 4.483e+10  3.202e+09
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
TukeyHSD(int_aov)
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = GFP..Cells.Mean.GFP.A ~ Sample, data = green_samples_only)
##
## $Sample
##              diff      lwr      upr      p adj
## CMV-mEGFP-Signed-CMV-mEGFP-Unsigned  63483.83 -71178.5  198146.168 0.5969914
## pV18-CMV-mEGFP-Unsigned             -143565.17 -278227.5   -8902.832 0.0343061
## pV19-CMV-mEGFP-Unsigned             -238396.92 -373059.3 -103734.582 0.0006127
## pV20-CMV-mEGFP-Unsigned             -409909.92 -544572.3 -275247.582 0.0000015
```



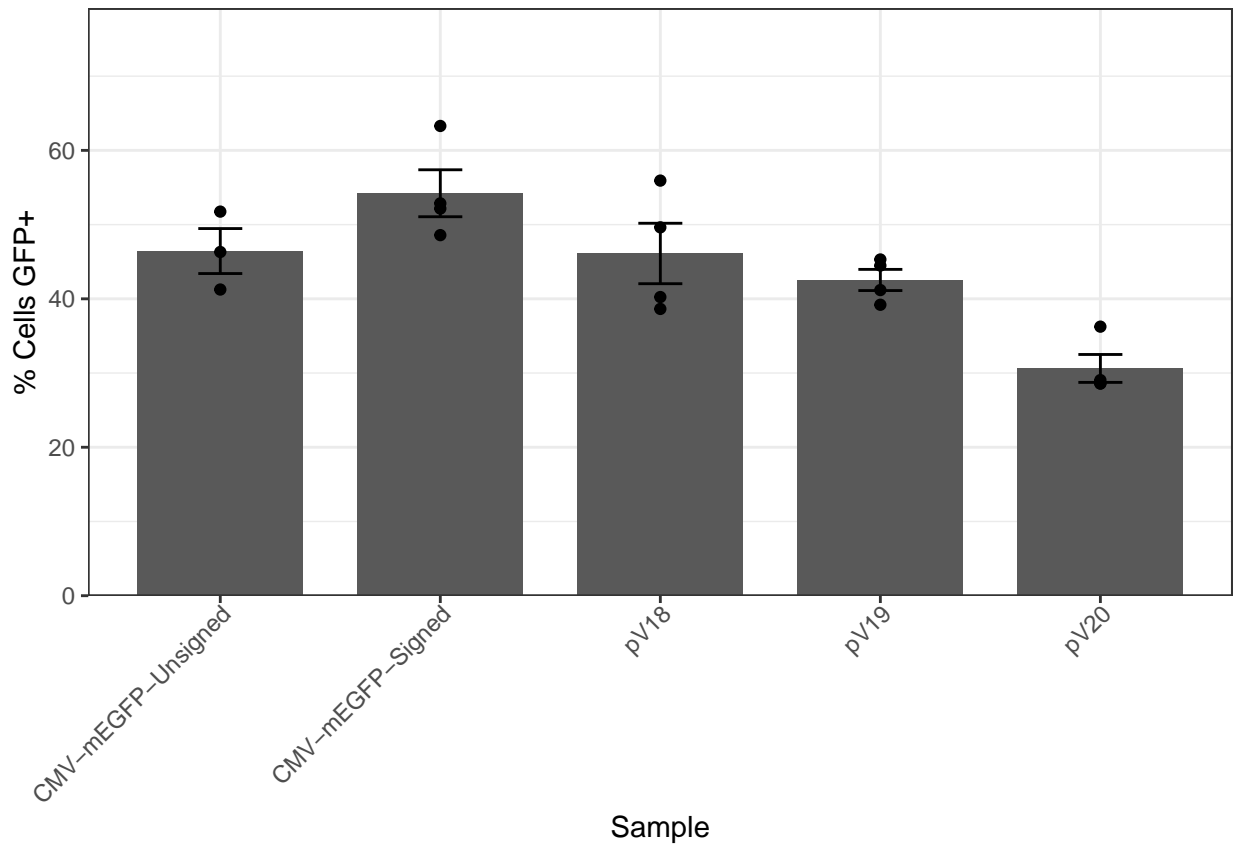
```
## pV18-CMV-mEGFP-Signed          -207049.00 -331722.1  -82375.904 0.0011209
## pV19-CMV-mEGFP-Signed          -301880.75 -426553.8  -177207.654 0.0000227
## pV20-CMV-mEGFP-Signed          -473393.75 -598066.8  -348720.654 0.0000001
## pV19-pV18                       -94831.75 -219504.8   29841.346 0.1806572
## pV20-pV18                      -266344.75 -391017.8  -141671.654 0.0000903
## pV20-pV19                      -171513.00 -296186.1  -46839.904 0.0056843
```

```
perc_plot <- ggplot(data = green_samples_only, mapping = aes(x = Sample, y = GFP..Cells...Parent)) +
  geom_bar(position = "dodge", stat = "summary", fun.y = "mean", width=0.75) +
  stat_summary(geom="errorbar", width=0.2) +
  geom_point() +
  scale_y_continuous(name="% Cells GFP+", expand = expansion(mult = c(0, 0.25))) +
  theme_bw() %+replace%
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1))
```

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

```
perc_plot
```

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



```
ggsave("perc_plot.png", plot=perc_plot, height=4, width=5)
```

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

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

```
ggsave("perc_plot.svg", plot=perc_plot, height=4, width=5)
```

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

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

```
intensity_plot <- ggplot(data = green_samples_only, mapping = aes(x = Sample, y = GFP..Cells.Mean.GFP.A  
  geom_bar(position = "dodge", stat = "summary", fun.y = "mean", width=0.75) +  
  stat_summary(geom="errorbar", width=0.2) +  
  geom_point() +  
  scale_y_continuous(name="Mean GFP Intensity",  
    expand = expansion(mult = c(0, 0.25)),  
    labels = scales::scientific) +  
  theme_bw() %+replace%  
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1))
```

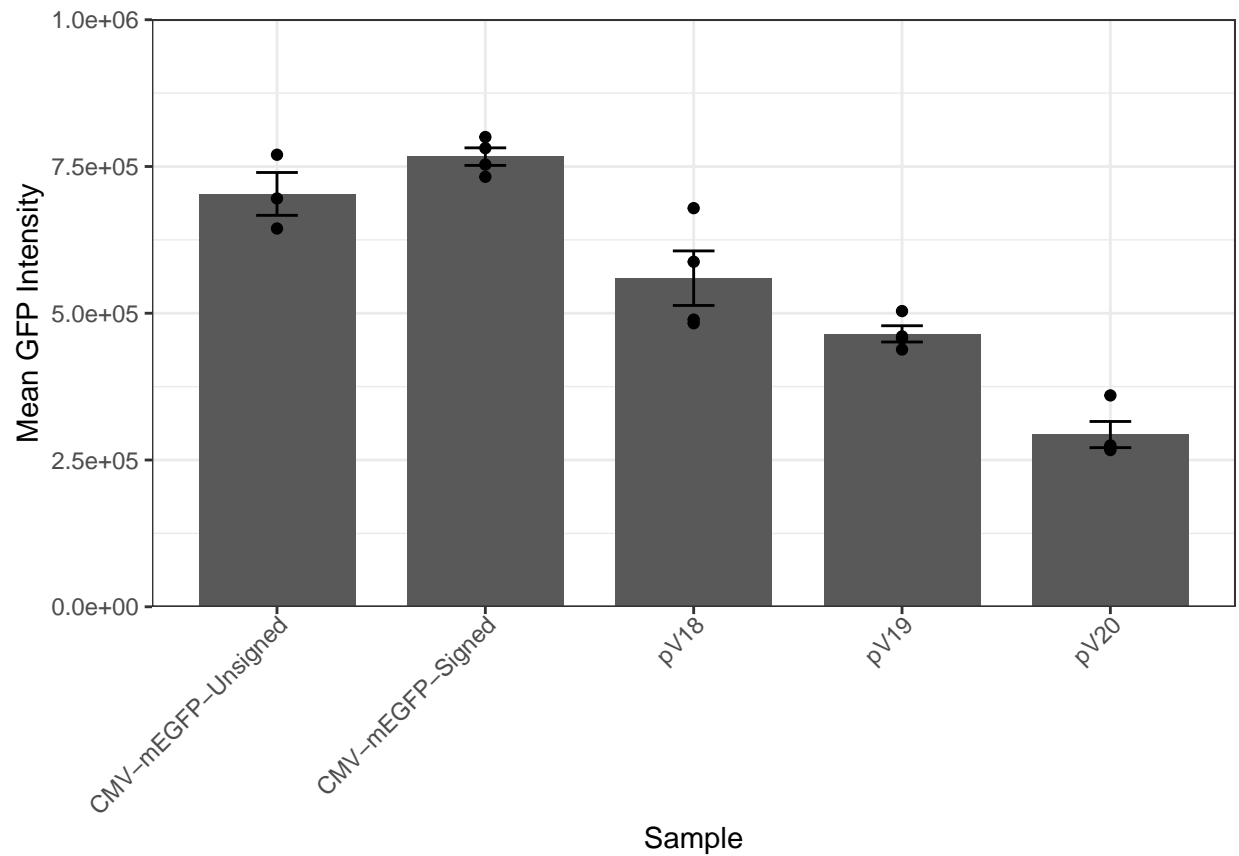
```
## Warning in geom_bar(position = "dodge", stat = "summary", fun.y = "mean", :
```

```
## Ignoring unknown parameters: 'fun.y'
```

```
intensity_plot
```

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

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



```
ggsave("intensity_plot.png", plot=intensity_plot, height=4, width=5)
```

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

```
ggsave("intensity_plot.svg", plot=intensity_plot, height=4, width=5)
```

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