Compare StepOne vs QuantaStudio for Ssid samples

AnaPalacio

Jan 26, 2018

General project set-up

```
# Get all libraries and sources required to run the script
        source("STEPoneFunction.R")# R.Cunning steponeR function
        library(plyr)
        library(dplyr)
        library(reshape2)
        library(ggplot2)
        library(lme4)
        library(ggthemes)
```

1. Calculate qPCR RATIOS (Symbiont/Coral)

Get the raw data for Ssid R.Cunning steponeR function:

- Get list of plate files to read
- Calculate the ratios
- Extract the results

```
  Ssid.plates <- list.files(path="Data", pattern=".csv",
                            full.names=TRUE, recursive = TRUE)
  # Ssid.plates

  # Run stepone function to get Ratios

  Ssid.Out <- steponeR(files=Ssid.plates, target.ratios=c("C.Ssid", "D.Ssid"),
                       fluor.norm=list(C=0, D=0, Ssid=0),
                       copy.number=list(C=1, D=1, Ssid=1),
                       ploidy=list(C=1, D=1, Ssid=2),
                       extract=list(C=0.813, D=0.813, Ssid=0.982))

  # Target ratio results
  Ssid<-Ssid.Out$result
```

2. Data CLEANING

```
  # 1. Check and remove NTC wells
    ntc <- Ssid[which(Ssid$Sample.Name=="NTC"), ]
    Ssid <- droplevels(Ssid[!rownames(Ssid) %in% rownames(ntc), ])

  # 2. Check and remove + Control wells
    Positive <- Ssid[which(Ssid$Sample.Name=="+"| Ssid$Sample.Name=="Sample 1"), ]
    Ssid <- droplevels(Ssid[!rownames(Ssid) %in% rownames(Positive), ])
```

```
# 3. Create unique sample ID+FileName to relabel samples
  Ssid$Sample.Plate<-paste(Ssid$Sample.Name,Ssid$File.Name, sep="_")

# 4.If Clade only detected in one technical replicate, set its ratio to NA
  One.C<- Ssid[which(Ssid$C.reps==1),]
  Ssid$C.Ssid[which(Ssid$C.reps==1)] <- NA

  One.D<- Ssid[which(Ssid$D.reps==1),]
  Ssid$D.Ssid[which(Ssid$D.reps==1)] <- NA

# 5. Make NA=0
   # colnames(Ssid)[which(colnames(Ssid) %in% "A.Ssid")] <- "A.SH"
   Ssid$C.Ssid[is.na(Ssid$C.Ssid)] <- 0
   Ssid$D.Ssid[is.na(Ssid$D.Ssid)] <- 0

# 6.If coral detected in one technical replicate, remove the sample
  ReRun.Coral <- Ssid[which(Ssid$Ssid.reps==1), ]
  Ssid <- droplevels(Ssid[!rownames(Ssid) %in% rownames(ReRun.Coral), ])
```

3. Get the cell ratios and log 10 transformations

```
  # Total SH
      Ssid$TotalSH<-(Ssid$C.Ssid + Ssid$D.Ssid)

  # Log 10
      Ssid$logC.SH <- log10(Ssid$C.Ssid)
      Ssid$logD.SH <- log10(Ssid$D.Ssid)
      Ssid$logSH<-log10(Ssid$TotalSH)
```

Data summary

```
summary(Ssid)
```

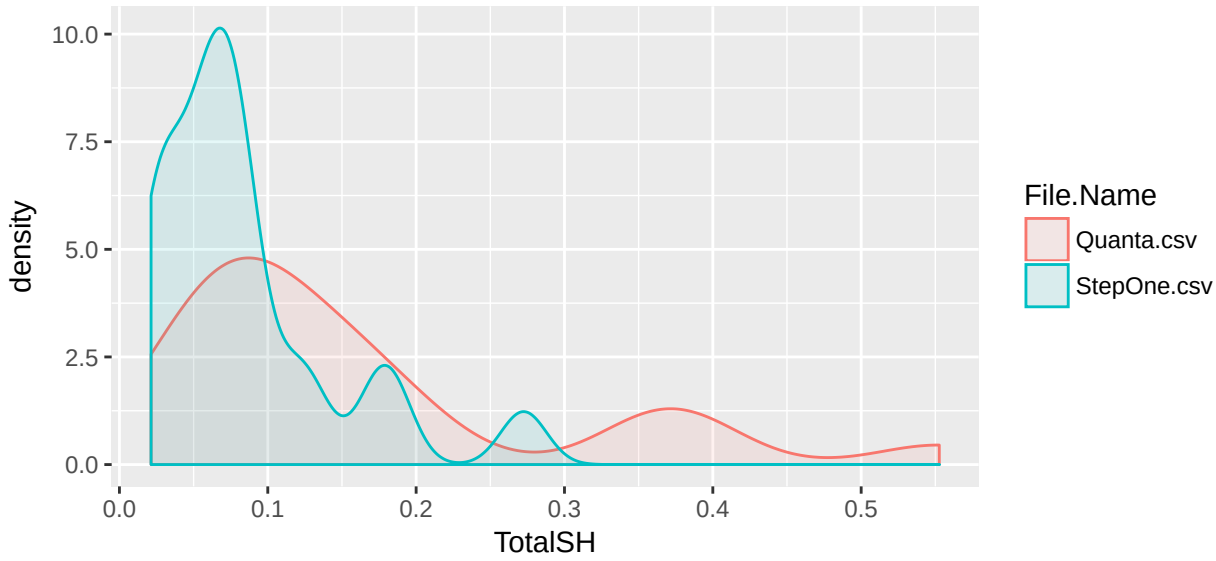
```
##  Sample.Name         File.Name           C.CT.mean       D.CT.mean
##  Length:42          Length:42          Min.   :25.44   Min.   :25.39
##  Class :character   Class :character   1st Qu.:27.11   1st Qu.:26.25
##  Mode  :character   Mode  :character   Median :29.37   Median :27.02
##                                        Mean   :30.76   Mean   :29.07
##                                        3rd Qu.:33.61   3rd Qu.:31.77
##                                        Max.   :38.82   Max.   :38.54
##                                        NA's   :9       NA's   :4
##   Ssid.CT.mean      C.CT.sd            D.CT.sd          Ssid.CT.sd
##  Min.   :20.52   Min.   :0.009192   Min.   :0.02051   Min.   :0.003654
##  1st Qu.:21.52   1st Qu.:0.144849   1st Qu.:0.23266   1st Qu.:0.058336
##  Median :22.08   Median :0.286067   Median :0.52856   Median :0.127543
##  Mean   :22.04   Mean   :0.472246   Mean   :0.56831   Mean   :0.185319
##  3rd Qu.:22.58   3rd Qu.:0.796368   3rd Qu.:0.72933   3rd Qu.:0.284979
##  Max.   :23.52   Max.   :1.834235   Max.   :1.87484   Max.   :1.101672
##                  NA's   :19         NA's   :8
##      C.reps          D.reps        Ssid.reps     C.Ssid
##  Min.   :0.000   Min.   :0.000   Min.   :2   Min.   :0.0000000
##  1st Qu.:1.000   1st Qu.:2.000   1st Qu.:2   1st Qu.:0.0000000
```

```
##  Median :2.000   Median :2.000   Median :2   Median :0.0006247
##  Mean   :1.333   Mean   :1.714   Mean   :2   Mean   :0.0293431
##  3rd Qu.:2.000   3rd Qu.:2.000   3rd Qu.:2   3rd Qu.:0.0568050
##  Max.   :2.000   Max.   :2.000   Max.   :2   Max.   :0.2048966
##
##      D.Ssid         Sample.Plate          TotalSH           logC.SH
##  Min.   :0.000000   Length:42          Min.   :0.02131   Min.   :   -Inf
##  1st Qu.:0.000396   Class :character   1st Qu.:0.05965   1st Qu.:   -Inf
##  Median :0.054300   Mode  :character   Median :0.08363   Median :-3.2043
##  Mean   :0.094108                      Mean   :0.12345   Mean   :   -Inf
##  3rd Qu.:0.141714                      3rd Qu.:0.15043   3rd Qu.:-1.2464
##  Max.   :0.552661                      Max.   :0.55266   Max.   :-0.6885
##
##     logD.SH            logSH
##  Min.   :   -Inf   Min.   :-1.6713
##  1st Qu.:-3.4106   1st Qu.:-1.2246
##  Median :-1.2700   Median :-1.0778
##  Mean   :   -Inf   Mean   :-1.0490
##  3rd Qu.:-0.8491   3rd Qu.:-0.8227
##  Max.   :-0.2575   Max.   :-0.2575
##
```

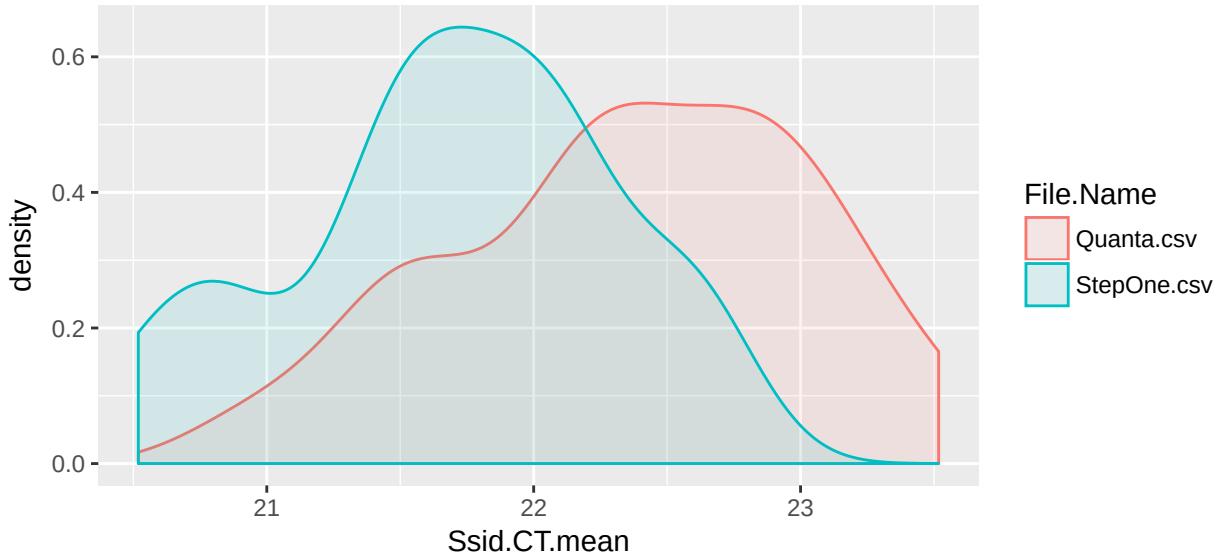
```
# StepOne and Quanta mean CT values and mean SD

Ssid %>%
  group_by(File.Name) %>%
  summarise(meanC = mean(C.CT.mean, na.rm=TRUE),
            meanD = mean(D.CT.mean, na.rm=TRUE),
            meanSsid = mean(Ssid.CT.mean),
            SD_C = mean(C.CT.sd, na.rm=TRUE),
            SD_D = mean(D.CT.sd, na.rm=TRUE),
            Sd_Ssid = mean(Ssid.CT.sd), n = n())
```

```
## # A tibble: 2 x 8
##   File.Name   meanC meanD meanSsid  SD_C  SD_D Sd_Ssid     n
##   <chr>       <dbl> <dbl>    <dbl> <dbl> <dbl>   <dbl> <int>
## 1 Quanta.csv   30.8  28.9     22.4 0.513 0.655   0.248    21
## 2 StepOne.csv  30.8  29.3     21.7 0.435 0.470   0.123    21
```

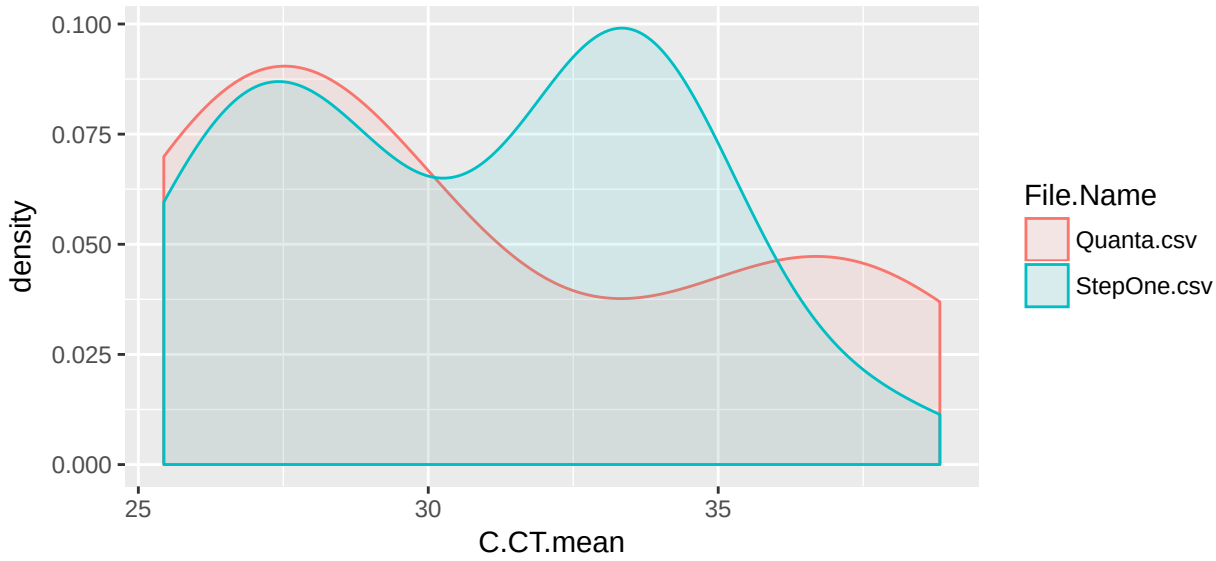
```
 # density Plots

   DenSH<- ggplot(Ssid, aes(TotalSH, fill = File.Name , colour = File.Name)) +
           geom_density(alpha = 0.1)
   DenSH
```



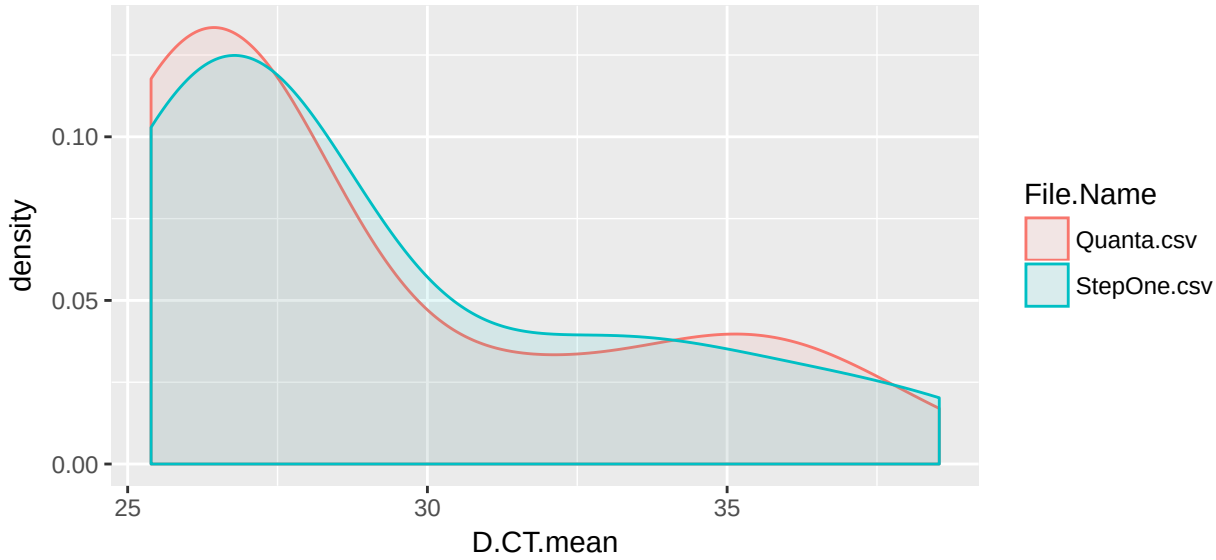

```
DenSidCT<-ggplot(Ssid, aes(Ssid.CT.mean, fill = File.Name , colour = File.Name)) +
        geom_density(alpha = 0.1)
DenSidCT
```




```
DenC_CT<-ggplot(Ssid, aes(C.CT.mean, fill = File.Name , colour = File.Name)) +
        geom_density(alpha = 0.1)
DenC_CT
```




```
DenD_CT<-ggplot(Ssid, aes(D.CT.mean, fill = File.Name , colour = File.Name)) +
        geom_density(alpha = 0.1)
DenD_CT
```



4. Tests for "teatment" = machine effects

```
# Total (log10) SH
    # T test
        PLate_SH <-t.test(logSH ~ File.Name, data = Ssid)
        PLate_SH
```

```
##
##  Welch Two Sample t-test
##
## data:  logSH by File.Name
## t = 2.5377, df = 39.354, p-value = 0.01523
```

```
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.05208421 0.46064770
## sample estimates:
##  mean in group Quanta.csv mean in group StepOne.csv
##                -0.9208339                -1.1771998
```

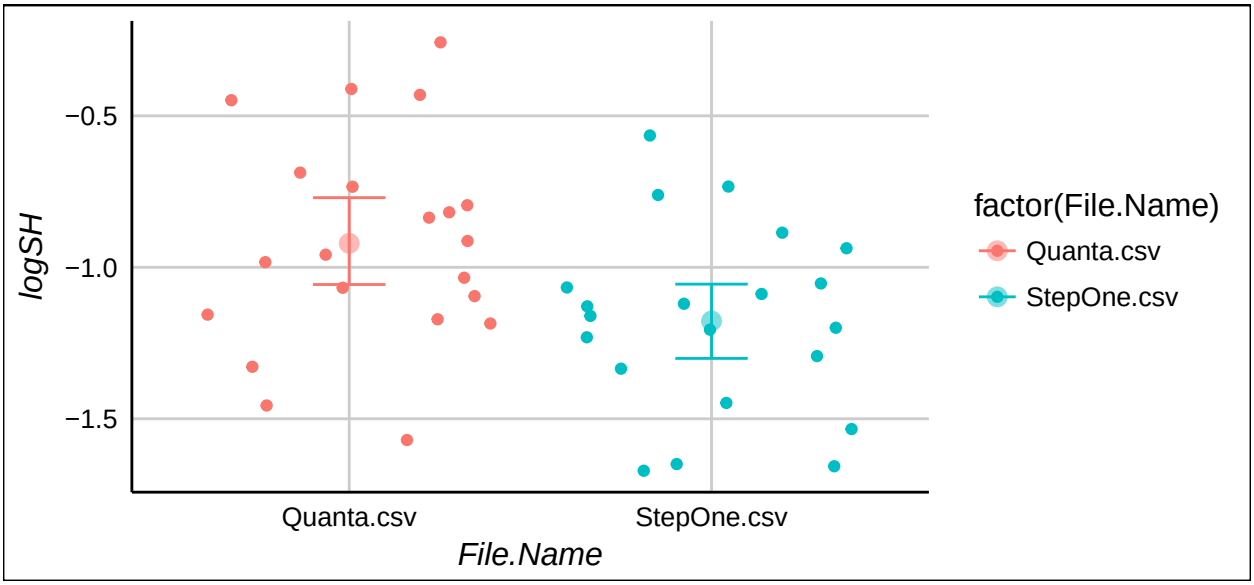
```
    # or (lm with Sample nested?)
        library(nlme)

        SH_LM<-lme(logSH ~ File.Name, random=~1|Sample.Name,
                   data=Ssid,
                   method="REML")
        anova.lme(SH_LM)
```

```
##             numDF denDF   F-value p-value
## (Intercept)     1    20 224.36730  <.0001
## File.Name       1    20  82.85623  <.0001
```

```
        summary(SH_LM)
```

```
## Linear mixed-effects model fit by REML
##  Data: Ssid
##         AIC      BIC   logLik
##   0.2453585 7.000876 3.877321
##
## Random effects:
##  Formula: ~1 | Sample.Name
##         (Intercept)   Residual
## StdDev:   0.3143765 0.09126251
##
## Fixed effects: logSH ~ File.Name
##                            Value  Std.Error DF   t-value p-value
## (Intercept)          -0.9208339 0.07143477 20 -12.89056       0
## File.NameStepOne.csv -0.2563660 0.02816422 20  -9.10254       0
##  Correlation:
##                      (Intr)
## File.NameStepOne.csv -0.197
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -1.57919063 -0.50965011 -0.05081367  0.59887780  1.25600280
##
## Number of Observations: 42
## Number of Groups: 21
```

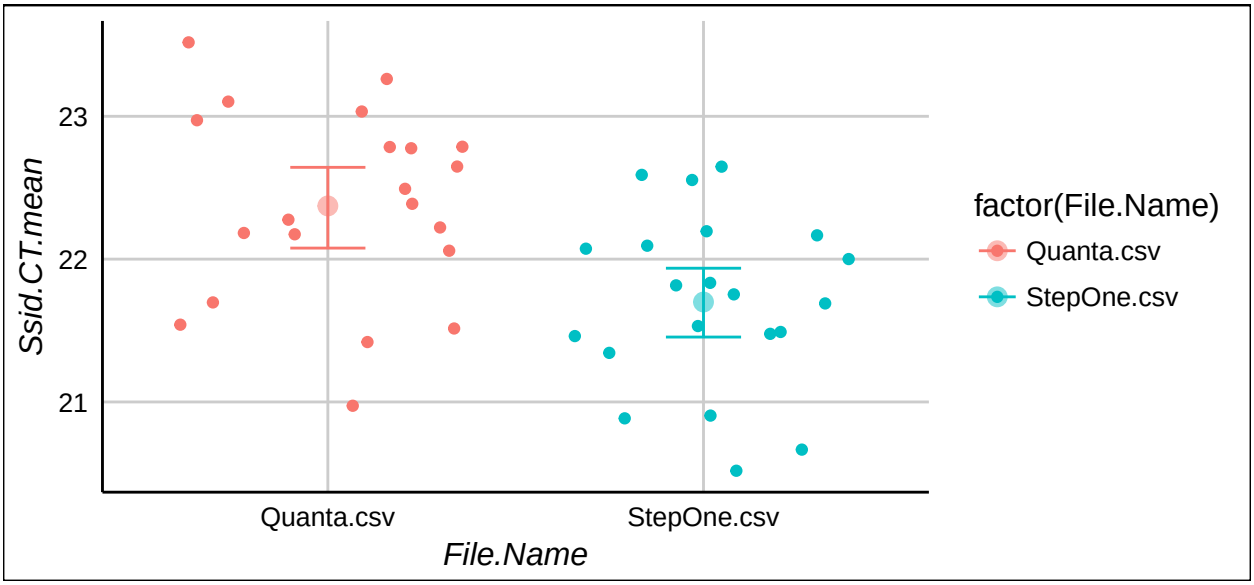
```
       # coef(PLate_SH)
       # layout(matrix(1:4,2,2))
       # plot(PLate_SH)

    logSH <- ggplot(Ssid, aes (File.Name, logSH, colour=factor(File.Name))) +
      stat_summary(fun.data = "mean_cl_boot",geom = "errorbar", width = 0.2)+
      stat_summary(fun.y=mean, geom="point", size =3, alpha=0.5) + theme_gdocs() +
      geom_jitter()
    logSH
```




```
# Ssid (SYBR) CT means

    # T test
        PLate_Ssid <-t.test(Ssid.CT.mean ~ File.Name, data = Ssid)
        PLate_Ssid
```

```
##
##  Welch Two Sample t-test
##
## data:  Ssid.CT.mean by File.Name
## t = 3.4097, df = 39.612, p-value = 0.001508
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.2738781 1.0717220
## sample estimates:
##  mean in group Quanta.csv mean in group StepOne.csv
##                  22.37252                  21.69972
```

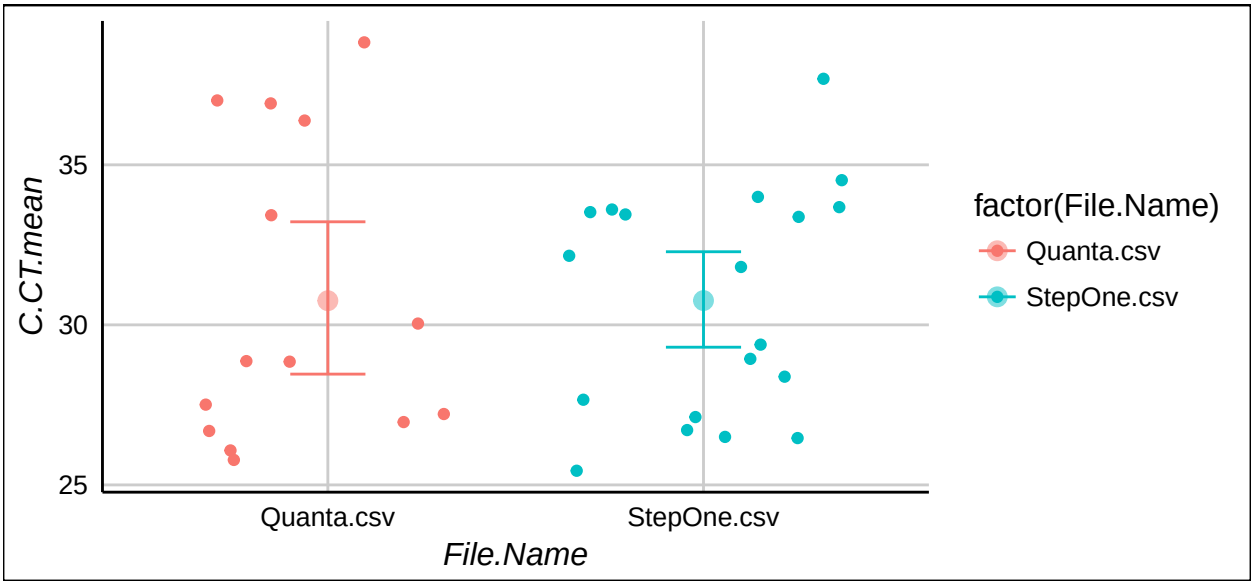
```
      Ssid_SYBER <- ggplot(Ssid, aes (File.Name, Ssid.CT.mean, colour=factor(File.Name))) +
        stat_summary(fun.data = "mean_cl_boot",geom = "errorbar", width = 0.2)+
        stat_summary(fun.y=mean, geom="point", size =3, alpha=0.5) + theme_gdocs() +
        geom_jitter()
      Ssid_SYBER
```




```
# C (VIC) CT means

    # T test
      PLate_C <-t.test(C.CT.mean ~ File.Name, data = Ssid)
      PLate_C
```

```
##
##  Welch Two Sample t-test
##
## data:  C.CT.mean by File.Name
## t = -0.0030188, df = 23.282, p-value = 0.9976
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.112696  3.103619
## sample estimates:
##  mean in group Quanta.csv mean in group StepOne.csv
##                  30.75525                  30.75979
```

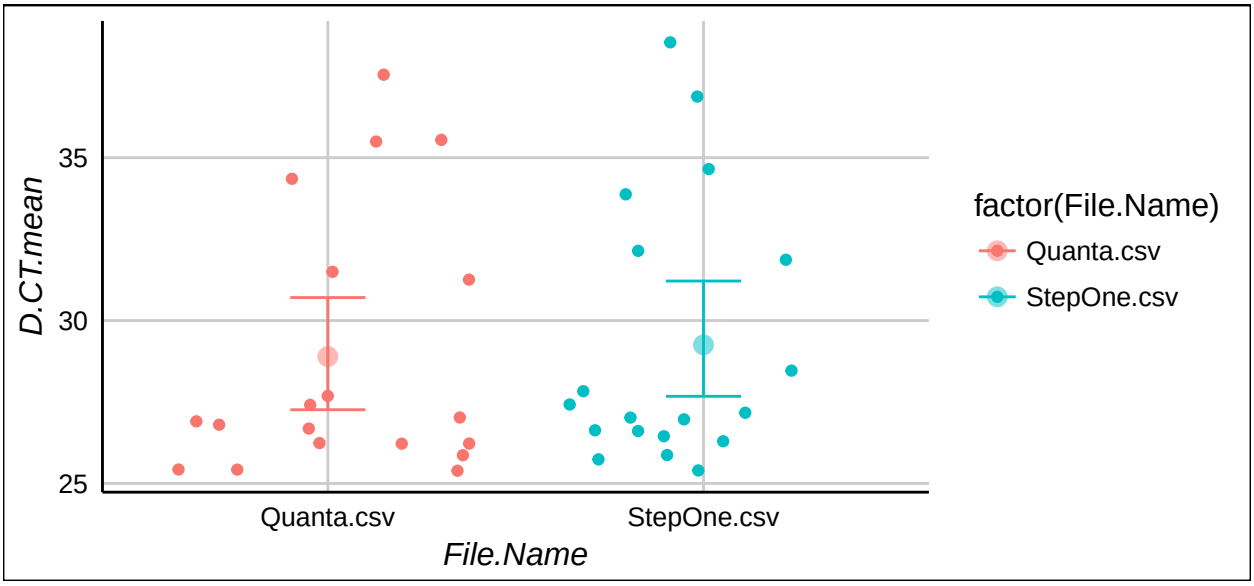
```
      C_VIC <- ggplot(Ssid, aes (File.Name, C.CT.mean, colour=factor(File.Name))) +
        stat_summary(fun.data = "mean_cl_boot",geom = "errorbar", width = 0.2)+
        stat_summary(fun.y=mean, geom="point", size =3, alpha=0.5) + theme_gdocs() +
        geom_jitter()
      C_VIC
```




```
# D (FAM) CT means

   # T test

     PLate_C <-t.test(D.CT.mean ~ File.Name, data = Ssid)
     PLate_C
```

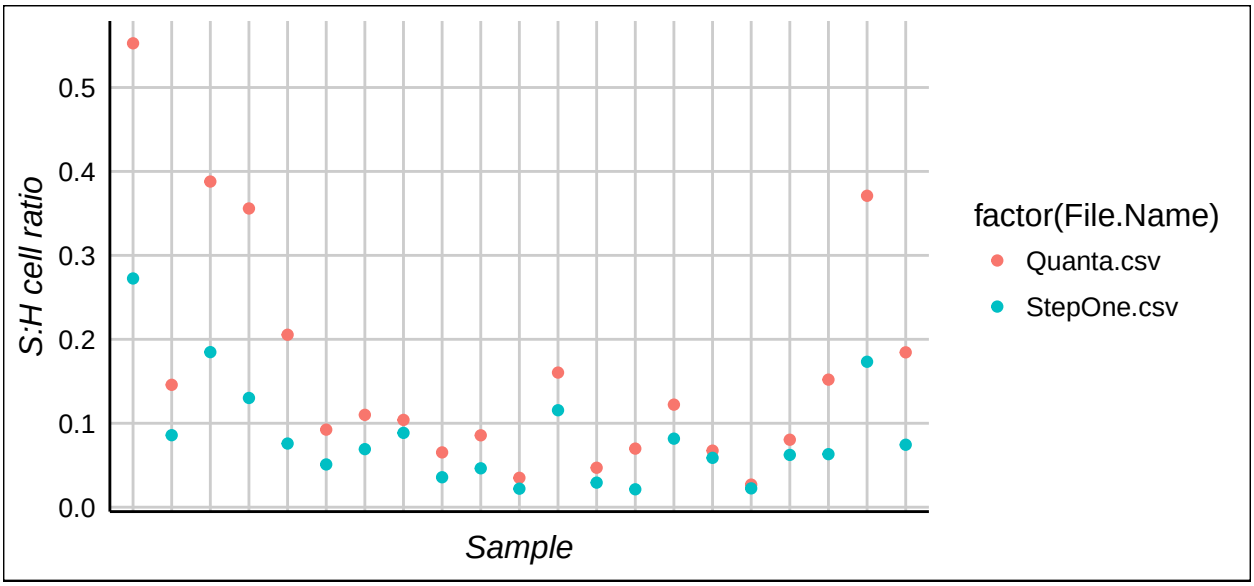
```
##
##  Welch Two Sample t-test
##
## data:  D.CT.mean by File.Name
## t = -0.27281, df = 35.994, p-value = 0.7866
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.026461  2.308783
## sample estimates:
##  mean in group Quanta.csv mean in group StepOne.csv
##                  28.89463                  29.25347
```

```
     D_VIC <- ggplot(Ssid, aes (File.Name, D.CT.mean, colour=factor(File.Name))) +
       stat_summary(fun.data = "mean_cl_boot",geom = "errorbar", width = 0.2)+
       stat_summary(fun.y=mean, geom="point", size =3, alpha=0.5) + theme_gdocs() +
       geom_jitter()
     D_VIC
```

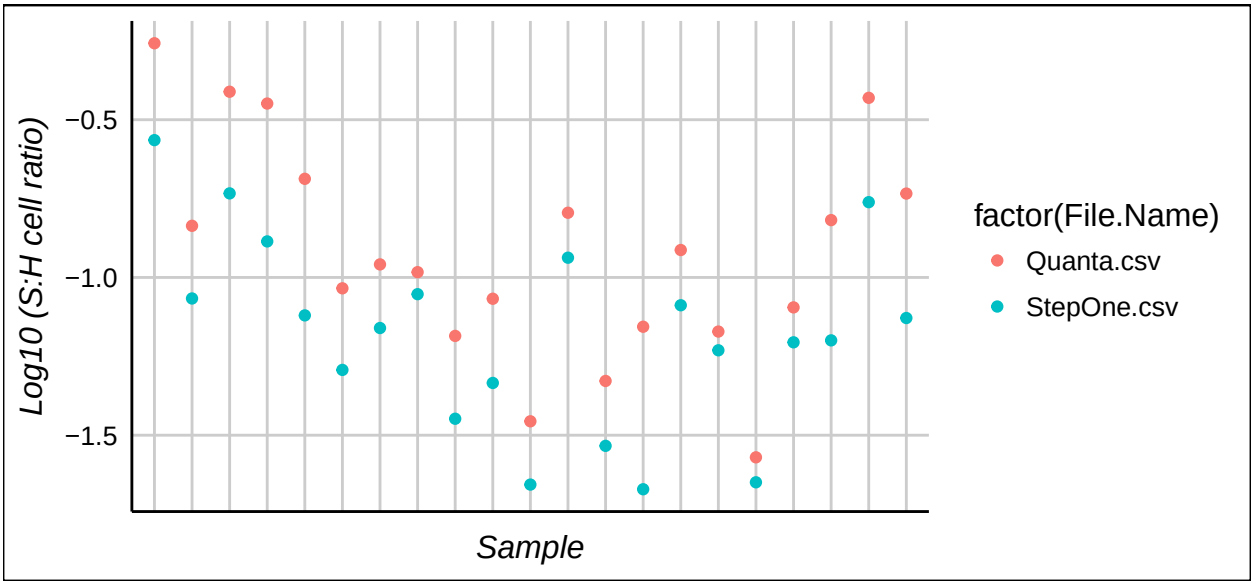


5. Exploratory graphs - Comparision bw rxn in each sample:

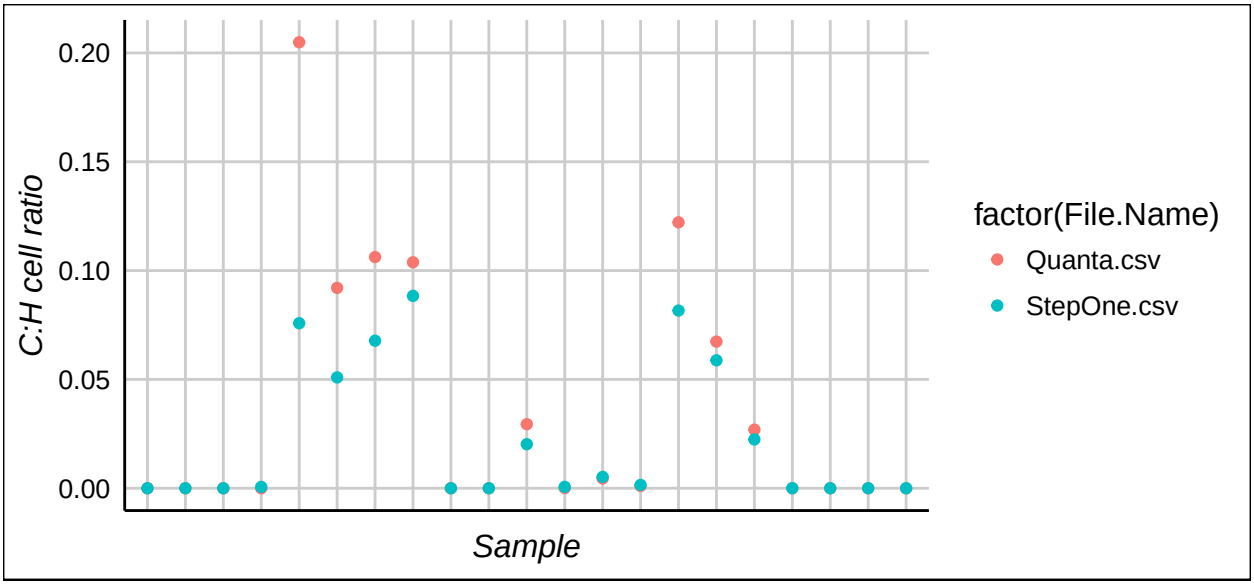
```
library("ggthemes")
library("scales")

Tot_SH <- ggplot(Ssid, aes(Sample.Name , TotalSH)) +
  geom_point(aes(colour=factor(File.Name)))  + theme_gdocs() +
  xlab("Sample") + ylab("S:H cell ratio") + theme(axis.text.x = element_blank())
Tot_SH
```



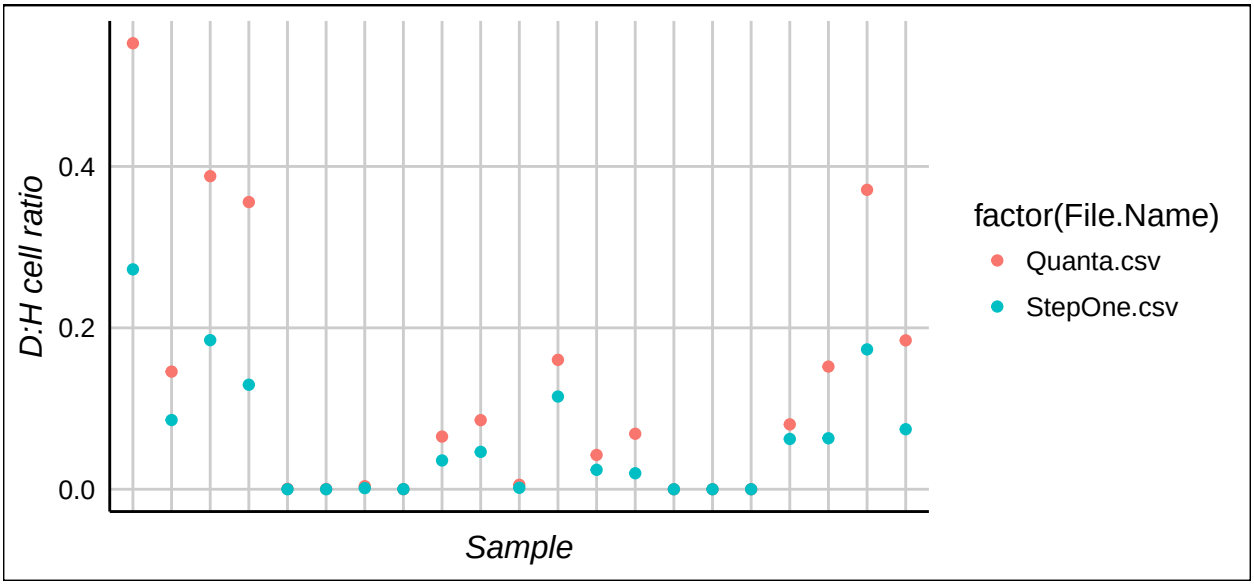

```
LogTot_SH <- ggplot(Ssid, aes(Sample.Name , logSH )) +
  geom_point(aes(colour=factor(File.Name)))  + theme_gdocs() +
  xlab("Sample") + ylab("Log10 (S:H cell ratio)") + theme(axis.text.x = element_blank())
LogTot_SH
```



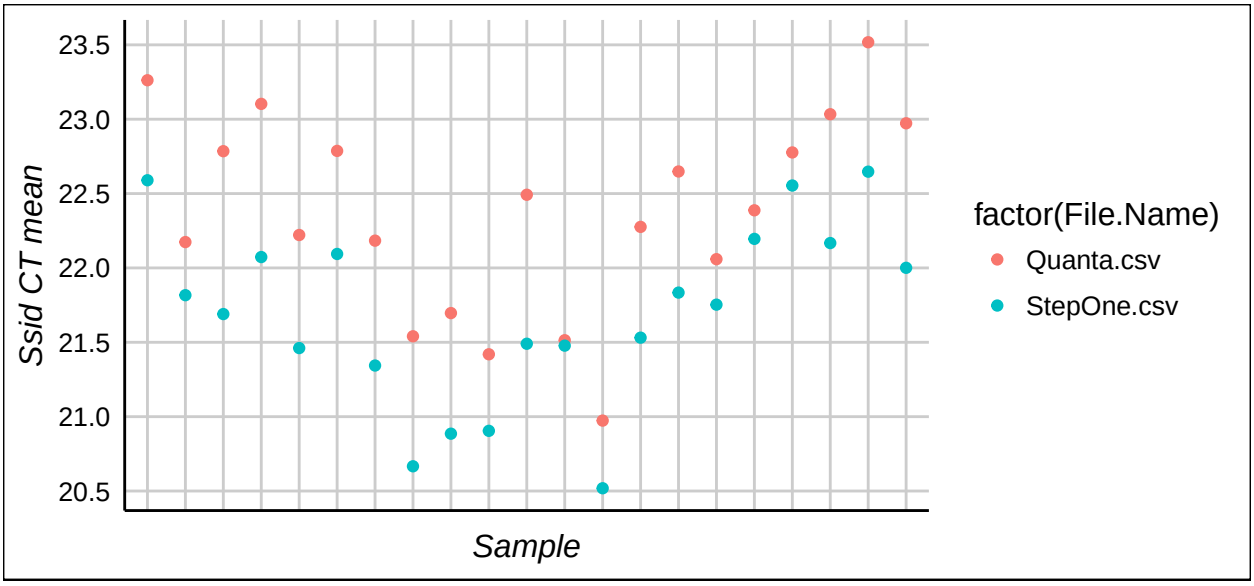

```
C_SH <- ggplot(Ssid, aes(Sample.Name , C.Ssid)) +
   geom_point(aes(colour=factor(File.Name)))  + theme_gdocs() +
   xlab("Sample") + ylab("C:H cell ratio") + theme(axis.text.x = element_blank())
C_SH
```



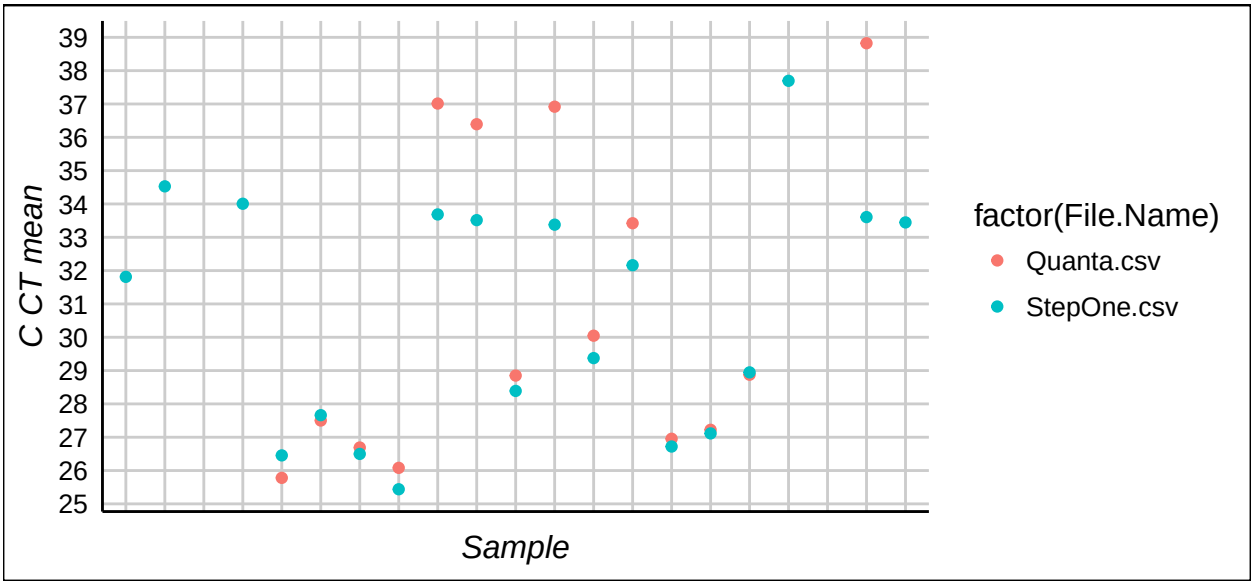

```
D_SH <- ggplot(Ssid, aes(Sample.Name , D.Ssid)) +
   geom_point(aes(colour=factor(File.Name)))  + theme_gdocs()+
  xlab("Sample") + ylab("D:H cell ratio") + theme(axis.text.x = element_blank())
D_SH
```



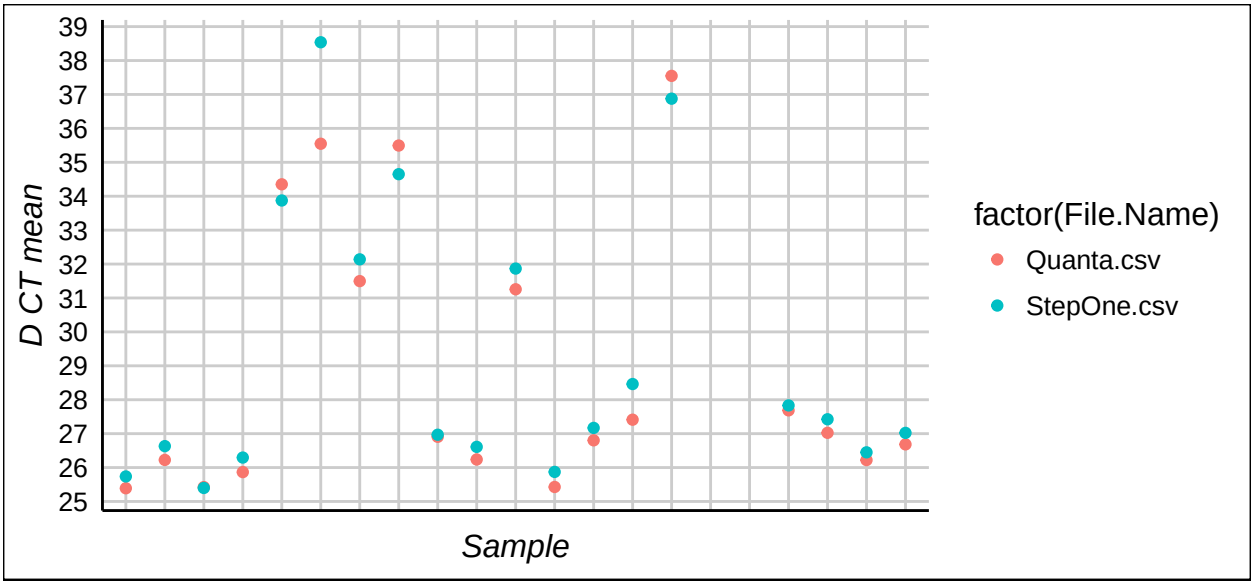

```
Ssid_CT <- ggplot(Ssid, aes(Sample.Name , Ssid$Ssid.CT.mean)) +
   geom_point(aes(colour=factor(File.Name)))  + theme_gdocs()  +
   xlab("Sample") + theme(axis.text.x = element_blank()) + scale_y_continuous(name="Ssid CT mean", breal
Ssid_CT
```




```
C_CT <- ggplot(Ssid, aes(Sample.Name , Ssid$C.CT.mean)) +
   geom_point(aes(colour=factor(File.Name)))  + theme_gdocs()  +
   xlab("Sample") + theme(axis.text.x = element_blank()) + scale_y_continuous(name="C CT mean", breaks=s
C_CT
```




```
D_CT <- ggplot(Ssid, aes(Sample.Name , Ssid$D.CT.mean)) +
   geom_point(aes(colour=factor(File.Name)))  + theme_gdocs() +
   xlab("Sample")+ theme(axis.text.x = element_blank()) + scale_y_continuous(name="D CT mean", breaks=s
D_CT
```



Differences bw the CTs in each machine

```
Data.Wide<- reshape(Ssid, idvar = "Sample.Name", timevar = "File.Name", direction = "wide")

# CTs Differences

    Data.Wide$Ssid_Diff<-(Data.Wide$Ssid.CT.mean.Quanta.csv-Data.Wide$Ssid.CT.mean.StepOne.csv)
      summary(Data.Wide$Ssid_Diff)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.03651 0.45486 0.76033 0.67280 0.87031 1.09523
```

```
    Data.Wide$C_Diff<-(Data.Wide$C.CT.mean.Quanta.csv-Data.Wide$C.CT.mean.StepOne.csv)
```

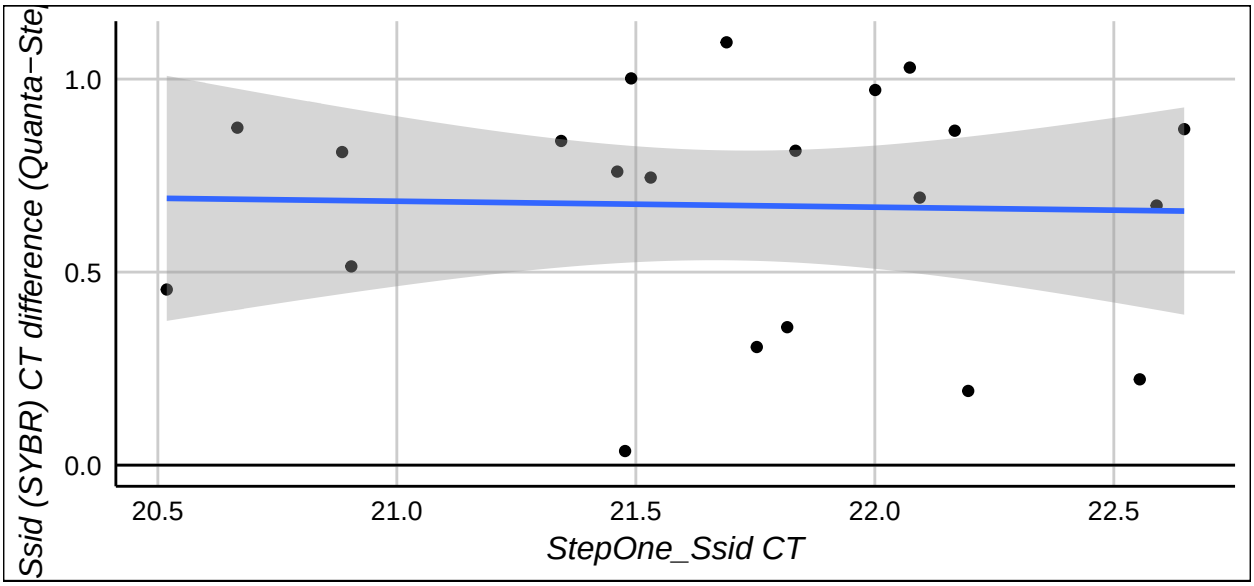
```
    summary(Data.Wide$C_Diff)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
## -0.6739  0.1294  0.5514  1.2591  2.4741  5.2172       7
```

```
    Data.Wide$D_Diff<-(Data.Wide$D.CT.mean.Quanta.csv-Data.Wide$D.CT.mean.StepOne.csv)
    summary(Data.Wide$D_Diff)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
## -2.9925 -0.4365 -0.3668 -0.3588 -0.1034  0.8444       2
```

```
  # Graphs

    Ssid_CT_Diff_A <- ggplot(Data.Wide, aes(Ssid.CT.mean.StepOne.csv , Data.Wide$Ssid_Diff)) +
          geom_point()  + theme_gdocs() + geom_hline(yintercept =0) +
          xlab("StepOne_Ssid CT") +  geom_smooth(method="lm") +
          scale_y_continuous(name=" Ssid (SYBR) CT difference (Quanta-StepOne)",breaks=seq(-2,3,0.5))
    Ssid_CT_Diff_A
```

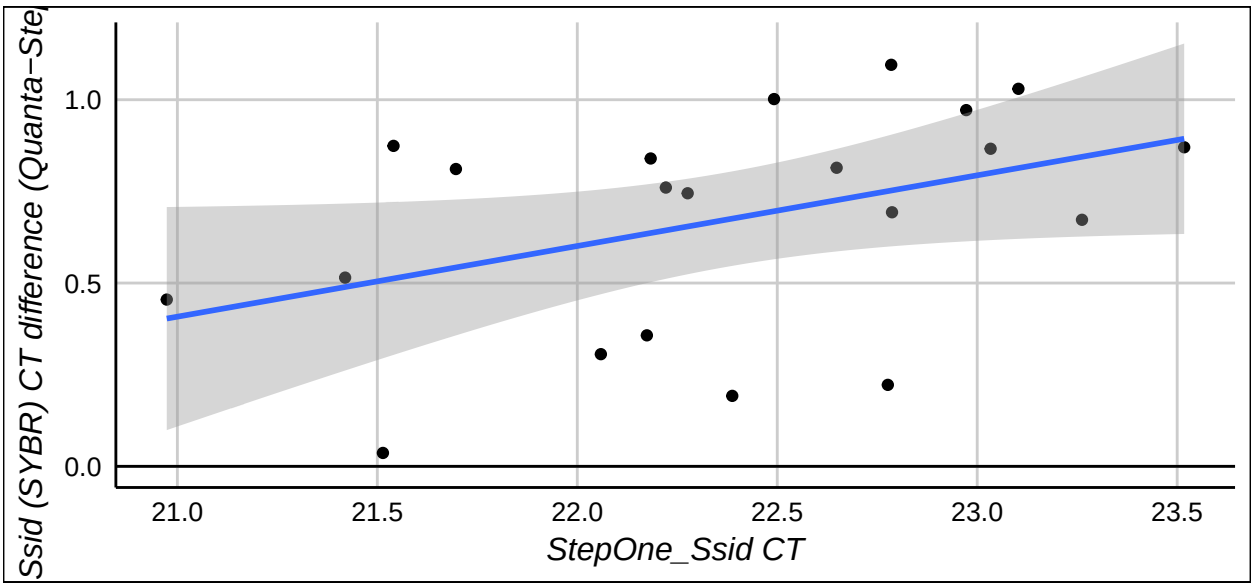



```
    SYBR_CT<-lm (Ssid_Diff ~ Ssid.CT.mean.StepOne.csv, data=Data.Wide)
    summary(SYBR_CT)
```

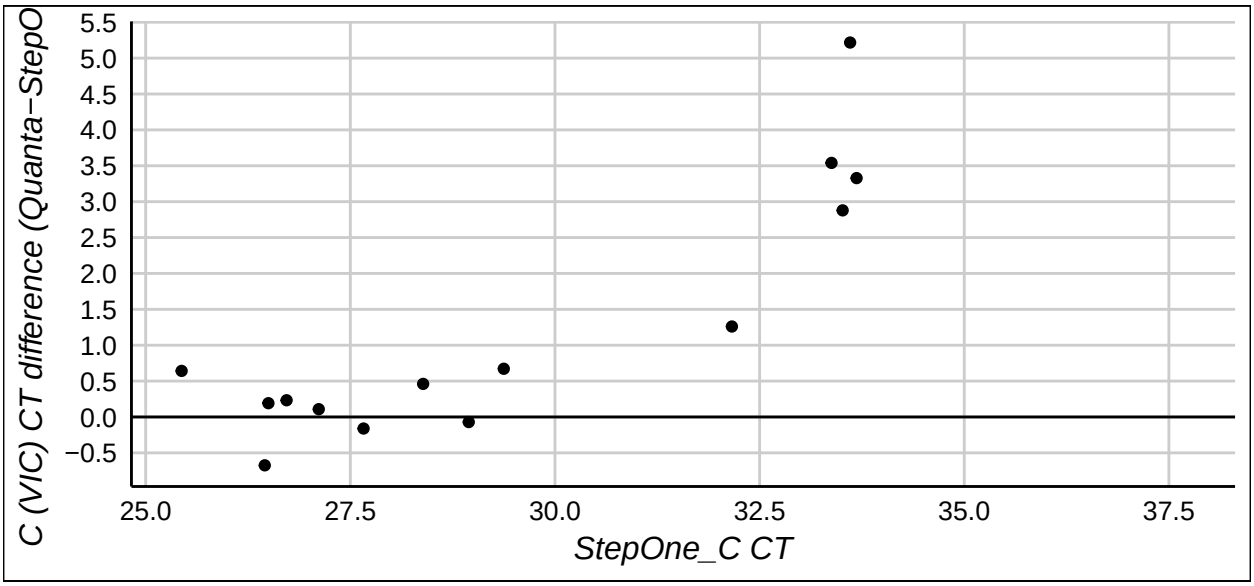
```
##
## Call:
## lm(formula = Ssid_Diff ~ Ssid.CT.mean.StepOne.csv, data = Data.Wide)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.63972 -0.23618  0.08385  0.20071  0.42228
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)                1.00784    2.49185   0.404    0.690
## Ssid.CT.mean.StepOne.csv -0.01544    0.11479  -0.135    0.894
##
## Residual standard error: 0.3116 on 19 degrees of freedom
## Multiple R-squared:  0.0009513,  Adjusted R-squared:  -0.05163
```

```
## F-statistic: 0.01809 on 1 and 19 DF,  p-value: 0.8944
```

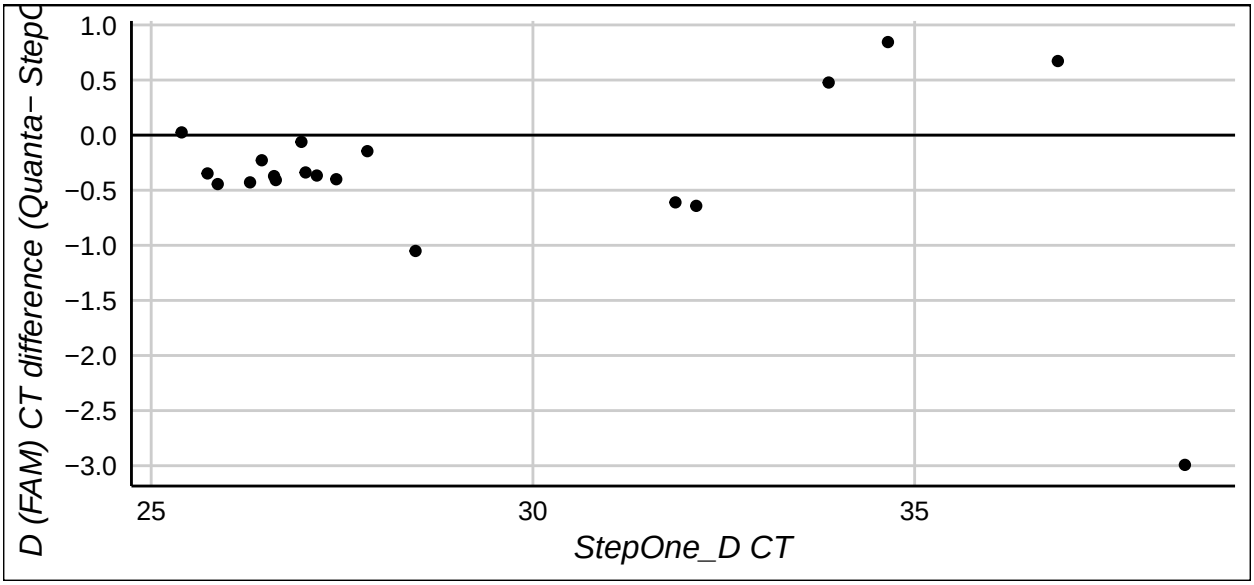
```
    Ssid_CT_Diff <- ggplot(Data.Wide, aes(Ssid.CT.mean.Quanta.csv , Data.Wide$Ssid_Diff)) +
          geom_point()  + theme_gdocs() + geom_hline(yintercept =0) +
          xlab("StepOne_Ssid CT") +  geom_smooth(method="lm") +
          scale_y_continuous(name=" Ssid (SYBR) CT difference (Quanta-StepOne)",breaks=seq(-2,3,0.5))
    Ssid_CT_Diff
```




```
    C_CT_Diff <- ggplot(Data.Wide, aes(C.CT.mean.StepOne.csv , Data.Wide$C_Diff)) +
          geom_point()  + theme_gdocs() + geom_hline(yintercept =0) + xlab("StepOne_C CT") +
          scale_y_continuous(name="C (VIC) CT difference (Quanta-StepOne)", breaks=seq(-2,6,0.5))
    C_CT_Diff
```




```
    D_CT_Diff <- ggplot(Data.Wide, aes(D.CT.mean.StepOne.csv , Data.Wide$D_Diff)) +
          geom_point()  + theme_gdocs()  + geom_hline(yintercept =0) + xlab("StepOne_D CT") +
          scale_y_continuous(name="D (FAM) CT difference (Quanta- StepOne)", breaks=seq(-3,1,0.5))
    D_CT_Diff
```




```
# SH Differences

# Data.Wide$C_Host_Diff<-(Data.Wide$C.Ssid.Quanta.csv-Data.Wide$C.Ssid.StepOne.csv)
# Data.Wide$D_Host_Diff<-(Data.Wide$D.Ssid.Quanta.csv-Data.Wide$D.Ssid.StepOne.csv)
# Data.Wide$TotalSH_Diff<-(Data.Wide$Ssid.CT.mean.Quanta.csv-Data.Wide$Ssid.CT.mean.StepOne.csv)
#
# SH_Diff <- ggplot(Data.Wide, aes(Sample.Name, TotalSH_Diff)) +
#        geom_point()  + theme_gdocs() + geom_hline(yintercept =0) +
#       xlab("Sample") + theme(axis.text.x = element_blank()) +
#       scale_y_continuous(name="SH cell ratio difference (Quanta-StepOne)",   breaks=seq(-0.5,1,0.2))
# SH_Diff
#
# CH_Diff <- ggplot(Data.Wide, aes(Sample.Name,  C_Host_Diff)) +
#       geom_point() + xlab("Sample") +  theme(axis.text.x = element_blank()) +
#       scale_y_continuous(name="CH cell ratio difference (Quanta-StepOne)", breaks=seq(-0.5,1,0.2)) +
#       theme_gdocs() + geom_hline(yintercept =0)
# CH_Diff
```