Statistical Analysis on icl knockout experiments

Coco Koedooder

12/12/2017

Data and Statistical Analysis regarding ICL mutants

Statistical analysis conducted on the results obtained from both ICL mutants. **icl knockout** measure differences in growth and respiration rates under different levels of iron limitation **icl bioreporter** measure gene expression under different levels of iron limitation

```
exp=read.table('icl_expression.txt', na.strings="NA", sep='\t', header=T, dec=',')
exp1=subset(exp, iron=="replete")
exp1 # inspect data
```

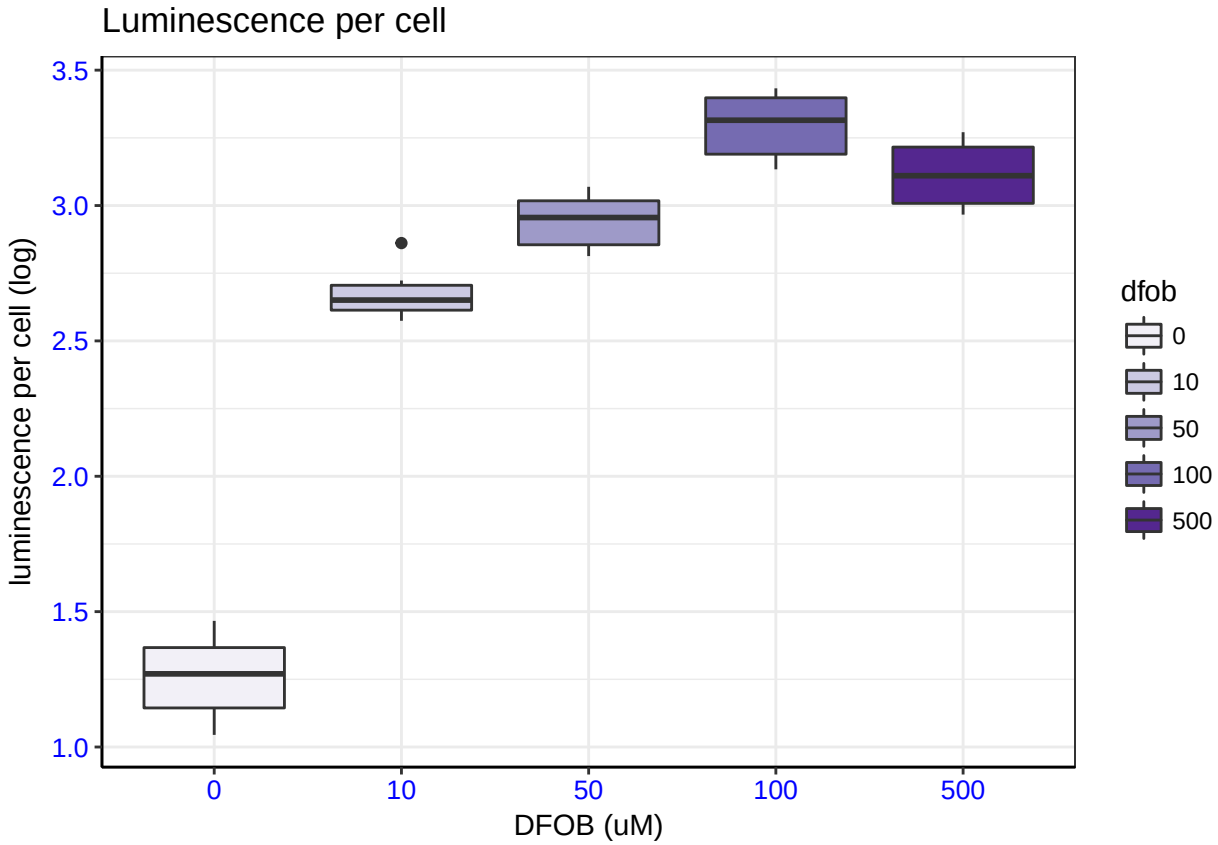
```
##       iron dfob luminescence
## 30 replete    0  2.842889088
## 31 replete    0  4.041059807
## 32 replete    0  4.331404024
## 33 replete   10  13.48810378
## 34 replete   10  14.13670492
## 35 replete   10  15.22671627
## 36 replete   50  16.87912485
## 37 replete   50  19.47637717
## 38 replete   50  18.95493243
## 39 replete  100  30.96405413
## 40 replete  100  30.29318557
## 41 replete  100  23.62707085
## 42 replete  500  23.55657247
## 43 replete  500  19.89361131
## 44 replete  500  21.34877925
## 45 replete    0  3.020356847
## 46 replete    0  3.594393491
## 47 replete    0  3.528756866
## 48 replete   10  13.12134066
## 49 replete   10  17.48387984
## 50 replete   10  14.19237542
## 51 replete   50  20.77272413
## 52 replete   50  21.52469875
## 53 replete   50  16.66442481
## 54 replete  100  28.75253252
## 55 replete  100  26.34827204
## 56 replete  100  22.95847964
## 57 replete  500  25.39999756
## 58 replete  500  26.33389248
## 59 replete  500  19.42245774
```

ICL bioreporter

Luminescence Plot the luminescence per 1000 cells of bioreporter bacteria grown under different levels of iron limitation. Data is log transformed due to an improved homoskedasticity. This is because of an increased variance as calculations are done with much lower cell counts as opposed to in the absence of iron limitation.

```
library(ggplot2)
library(car)
library(knitr)
exp1$dfob=as.factor(exp1$dfob)
exp1$luminescence=as.numeric(as.character(exp1$luminescence))
exp1$loglumi=log(exp1$luminescence)
exp1$loglumi=as.numeric(as.character(exp1$loglumi))

lum=ggplot(exp1, aes(x=factor(dfob), y=as.numeric(as.character(loglumi)), fill=dfob))+geom_boxplot()
lum=lum+theme_bw()
lum=lum+scale_x_discrete("DFOB (uM)")+scale_y_continuous("luminescence per cell (log)")
lum=lum+ggtitle("Luminescence per cell")
lum=lum+scale_fill_brewer(palette = 12)
lum=lum+theme(axis.line = element_line(colour = "black",
                                       size = 0.5, linetype = "solid"))
lum=lum+theme(axis.text = element_text(color="blue",size=10))
lum
```




```
# ANOVA one way
lumi=lm(loglumi ~ dfob,
            data = exp1)

ll <- Anova(lumi, type = 3)
kable(ll, digits = 10 ,results = 'asis')
```

	Sum Sq	Df	F value	Pr(>F)
(Intercept)	9.5077553	1	590.7227	0

	Sum Sq	Df	F value	Pr(>F)
dfob	15.9142706	4	247.1909	0
Residuals	0.4023781	25	NA	NA

```
shapiro.test(residuals(lumi))  #ok
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(lumi)
## W = 0.96025, p-value = 0.3143
```

```
leveneTest(lumi) #ok
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  4  0.5933 0.6707
##       25
```

```
# Summary of the analysis
lll <- aov(loglumi ~ dfob, data = exp1)
TukeyHSD(lll)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = loglumi ~ dfob, data = exp1)
##
## $dfob
##                diff         lwr        upr     p adj
## 10-0     1.4181572  1.20304179 1.63327252 0.0000000
## 50-0     1.6834245  1.46830908 1.89853982 0.0000000
## 100-0    2.0361739  1.82105852 2.25128926 0.0000000
## 500-0    1.8549864  1.63987105 2.07010179 0.0000000
## 50-10    0.2652673  0.05015193 0.48038267 0.0103848
## 100-10   0.6180167  0.40290137 0.83313211 0.0000001
## 500-10   0.4368293  0.22171389 0.65194463 0.0000291
## 100-50   0.3527494  0.13763407 0.56786481 0.0005298
## 500-50   0.1715620 -0.04355340 0.38667733 0.1650511
## 500-100 -0.1811875 -0.39630284 0.03392789 0.1290819
```

Data is highly significant between the luminescence and different levels of iron limitation ($p > 0.000001$). Data was shown to be normally distributed using a shapiro wilk test ($p = 0.3143$). Homoskedasticity was also met ($p = 0.6707$) using a Levene Test. The Tukey test reveals that these differences mainly lie between replete iron conditions and strong iron limitation and that at strong iron limitation luminescence does not differ significantly amongst each other. At 500 umol DFOB the luminescence seems to decrease, but this is likely due the fact that growth rate is so severely impacted by iron limitation that the results of luminescence data are unreliable.

ICL knockout

Differences in both growth and respiration rate between the wildtype and gene knockout ofICL subjected to different levels of iron limitation

```
data1=read.table('icl_alldates.txt', na.strings="NA", sep='\t', header=T, dec=',')
data1 # inspect data
```

```
##        date    iron strain conditions substrate respiration growth
## 1  20141015      Fe     WT      wt_fe   glucose        3.71  10.28
## 2  20141013      Fe     WT      wt_fe   glucose        3.26   9.85
## 3  20141013      Fe     WT      wt_fe   glucose        2.92  10.45
## 4  20141013      Fe     WT      wt_fe   glucose        3.17  10.97
## 5  20141013      Fe     WT      wt_fe   glucose        2.89  10.45
## 6  20141020      Fe     KO      ko_fe   glucose        4.79  10.37
## 7  20141022      Fe     KO      ko_fe   glucose        4.62  10.11
## 8  20141022      Fe     KO      ko_fe   glucose        4.51   9.94
## 9  20141022      Fe     KO      ko_fe   glucose        4.84  10.54
## 10 20141105      Fe     KO      ko_fe   glucose        3.67   7.69
## 11 20141105      Fe     KO      ko_fe   glucose        4.12   8.47
## 12 20141105      Fe     KO      ko_fe   glucose        4.05   8.99
## 13 20151011  10DFOB     WT  wt_10dfob   glucose        3.08   6.19
## 14 20151011  10DFOB     WT  wt_10dfob   glucose        1.94   5.99
## 15 20151011  10DFOB     WT  wt_10dfob   glucose        1.86    6.4
## 16 20151011  10DFOB     KO  ko_10dfob   glucose         1.4   4.04
## 17 20151011  10DFOB     KO  ko_10dfob   glucose        2.17   6.73
## 18 20151011  10DFOB     KO  ko_10dfob   glucose        2.31   7.68
## 19 20141020 100DFOB     KO ko_100dfob   glucose        0.78   1.21
## 20 20141020 100DFOB     KO ko_100dfob   glucose        0.97   1.93
## 21 20141020 100DFOB     KO ko_100dfob   glucose         0.9   1.17
## 22 20141025 100DFOB     WT wt_100dfob   glucose        1.41   3.02
## 23 20141025 100DFOB     WT wt_100dfob   glucose        1.63   3.44
## 24 20141025 100DFOB     WT wt_100dfob   glucose        1.66   3.08
## 25 20141025 100DFOB     WT wt_100dfob   glucose        1.02   3.33
## 26 20141119      Fe     WT      wt_fe   glucose        3.88   9.85
## 27 20141119      Fe     WT      wt_fe   glucose        6.62   6.46
## 28 20141119      Fe     WT      wt_fe   glucose        5.25    8.9
```

Luminescence for Different Carbon Concentrations

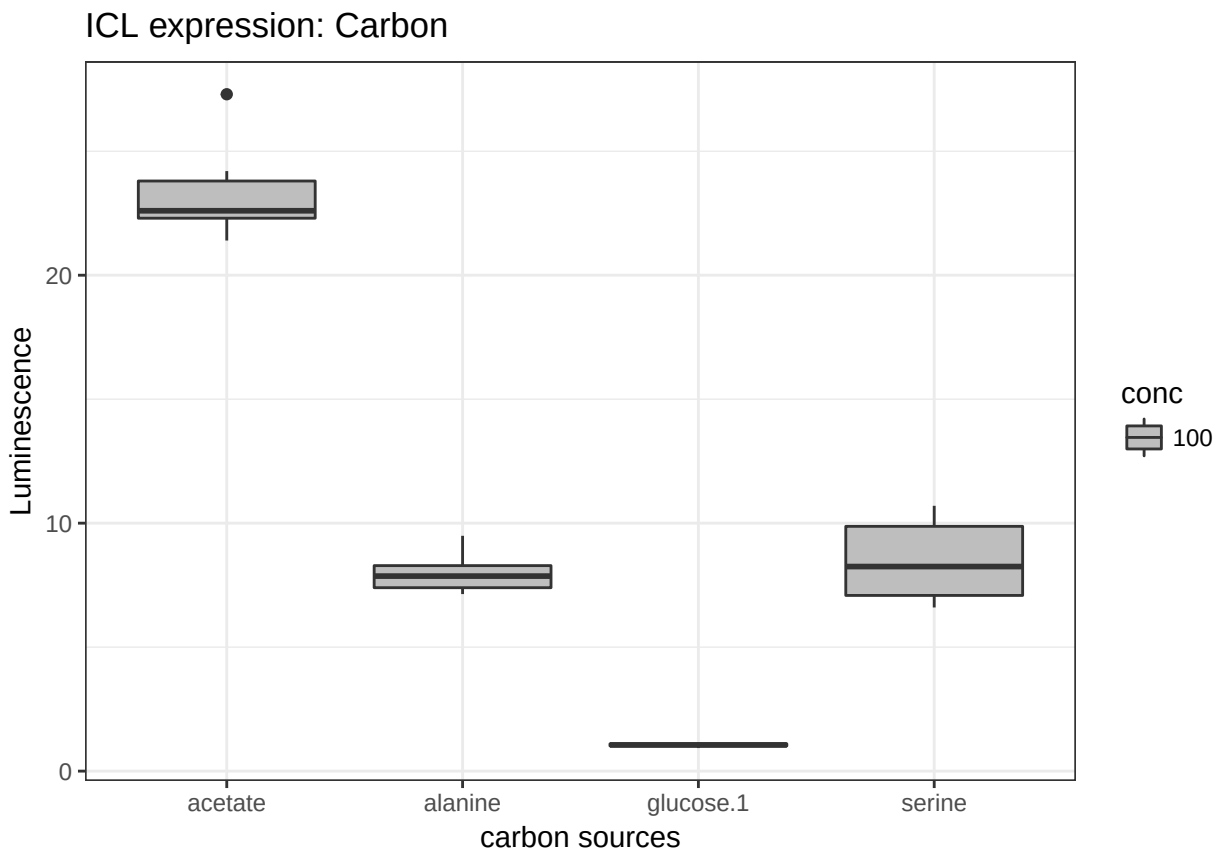
```
carb=read.table('PUBLICATION_carbonconc.txt', na.strings="NA", sep='\t', header=T, dec=',')
carb1=subset(carb, carbon!="glucose.6")
carb1 # inspect data
```

```
##       iron conc    carbon  counts lumi
## 1  replete  100 glucose.1 2174.42 1.11
## 2  replete  100 glucose.1 2190.52 1.01
## 3  replete  100 glucose.1 1802.27 1.15
## 4  replete  100 glucose.1 2331.95 0.93
## 5  replete  100 glucose.1 2110.26 1.08
## 6  replete  100 glucose.1 2180.65 1.03
## 13 replete  100   acetate 2567.66 24.2
## 14 replete  100   acetate 2634.09 22.6
## 15 replete  100   acetate 2154.61 27.3
## 16 replete  100   acetate 2574.61 22.2
## 17 replete  100   acetate 2615.06 21.4
## 18 replete  100   acetate 2521.43 22.6
## 19 replete  100   alanine 1827.21 9.49
```

```
## 20 replete  100   alanine 1742.34 8.34
## 21 replete  100   alanine 1586.88 8.13
## 22 replete  100   alanine 1755.45 7.14
## 23 replete  100   alanine 1625.58 7.33
## 24 replete  100   alanine 1621.62  7.6
## 25 replete  100    serine 2638.18 10.7
## 26 replete  100    serine 2293.12 10.2
## 27 replete  100    serine 1833.44 7.61
## 28 replete  100    serine 1966.49  6.6
## 29 replete  100    serine 2037.86 6.91
## 30 replete  100    serine 1828.77 8.89
```

```
carb1$conc=as.character(carb1$conc)
carb1$lumi=as.numeric(as.character(carb1$lumi))


c=ggplot(carb1, aes(x=carbon, y=lumi, fill=conc))+geom_boxplot()
c=c+theme_bw()
c=c+scale_x_discrete("carbon sources")+scale_y_continuous("Luminescence")
c=c+ggtitle("ICL expression: Carbon")
c=c+scale_fill_manual(values=c("grey"))
c
```




```
##anova - untransformed data
cc=lm(lumi ~ carbon,
        data = carb1)
ccc <- Anova(cc, type = 3)
kable(ccc, digits = 10,results = 'asis')
```

	Sum Sq	Df	F value	Pr(>F)
(Intercept)	3280.68167	1	1596.0071	0
carbon	1591.48575	3	258.0787	0
Residuals	41.11112	20	NA	NA

```
shapiro.test(residuals(cc))  #ok
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(cc)
## W = 0.92028, p-value = 0.05926
```

```
leveneTest(cc) #ok
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value  Pr(>F)
## group  3  2.4944 0.08934 .
##       20
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
cccc <- aov(lumi ~ carbon, data = carb1)
TukeyHSD(cccc)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = lumi ~ carbon, data = carb1)
##
## $carbon
##                          diff        lwr        upr     p adj
## alanine-acetate   -15.378333 -17.695178 -13.061488 0.0000000
## glucose.1-acetate -22.331667 -24.648512 -20.014822 0.0000000
## serine-acetate    -14.898333 -17.215178 -12.581488 0.0000000
## glucose.1-alanine  -6.953333  -9.270178  -4.636488 0.0000003
## serine-alanine      0.480000  -1.836845   2.796845 0.9369716
## serine-glucose.1    7.433333   5.116488   9.750178 0.0000001
```

Normality was met (p = 0.05926) and homoskedasticity (p = 0.08934) Significant differences (p<0.0005) in the luminescence between different carbon sources. A tukey test further revealed that there was no significant difference between alanine and serine (p = 0.9369716).

Different Glucose Concentrations

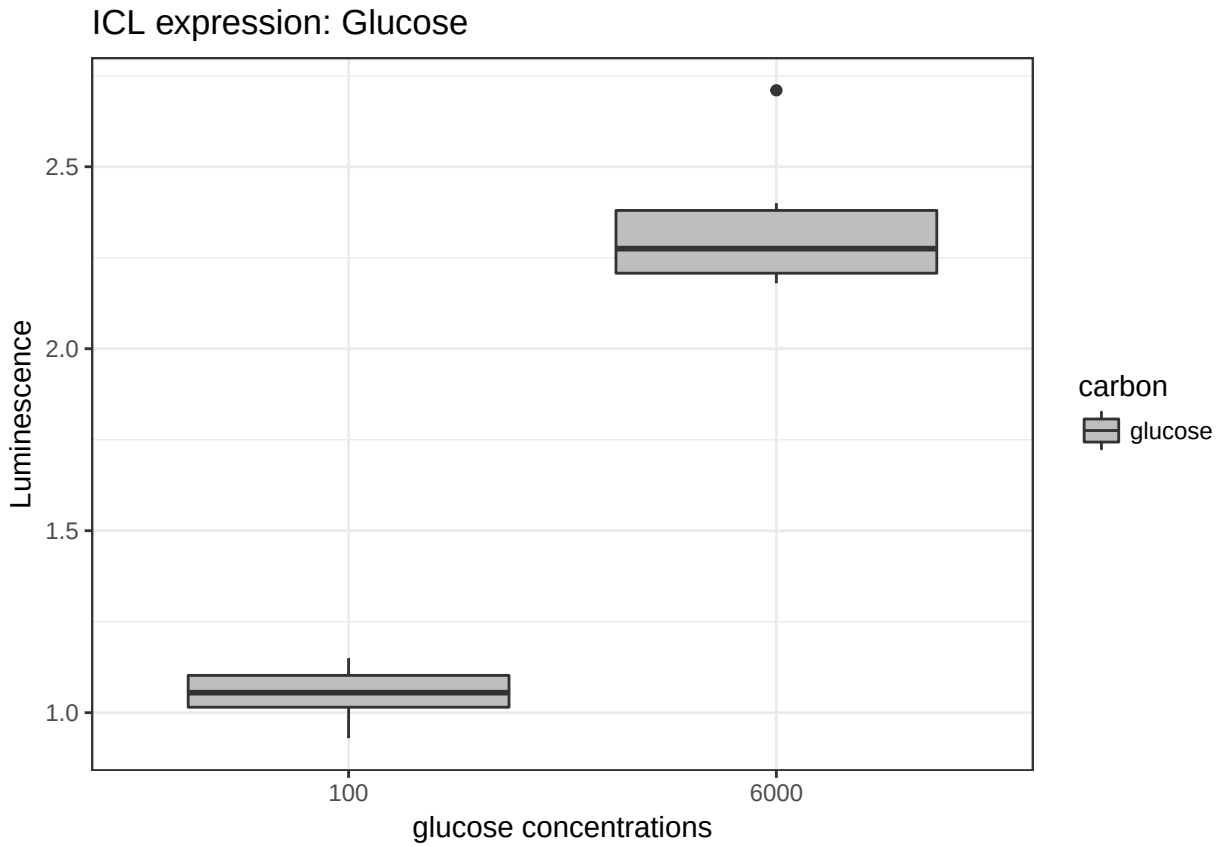
```
glu=read.table('glucose.txt', na.strings="NA", sep='\t', header=T, dec=',')
glu=subset(glu, carbon=="glucose")
glu # inspect data
```

```
##       iron conc  carbon  counts lumi
## 1  replete  100 glucose 2174.42 1.11
## 2  replete  100 glucose 2190.52 1.01
## 3  replete  100 glucose 1802.27 1.15
## 4  replete  100 glucose 2331.95 0.93
```

```
## 5  replete  100 glucose 2110.26 1.08
## 6  replete  100 glucose 2180.65 1.03
## 7  replete 6000 glucose 2605.52 2.23
## 8  replete 6000 glucose 2760.13  2.2
## 9  replete 6000 glucose 2445.97  2.4
## 10 replete 6000 glucose 2694.87 2.18
## 11 replete 6000 glucose 2616.04 2.32
## 12 replete 6000 glucose 2088.18 2.71
```

```
glu$conc=as.character(glu$conc)
glu$lumi=as.numeric(as.character(glu$lumi))

gluc=ggplot(glu, aes(x=conc, y=lumi, fill=carbon))+geom_boxplot()
gluc=gluc+theme_bw()
gluc=gluc+scale_x_discrete("glucose concentrations")+scale_y_continuous("Luminescence")
gluc=gluc+ggtitle("ICL expression: Glucose")
gluc=gluc+scale_fill_manual(values=c("grey"))
gluc
```




```
##anova - untransformed data
gluco=lm(lumi ~ conc,
        data = glu)
glucose <- Anova(gluco, type = 3)
kable(glucose, digits = 10,results = 'asis')
```

	Sum Sq	Df	F value	Pr(>F)
(Intercept)	6.6360167	1	289.6770	1.03e-08

	Sum Sq	Df	F value	Pr(>F)
conc	4.9794083	1	217.3623	4.13e-08
Residuals	0.2290833	10	NA	NA

```
shapiro.test(residuals(gluco))  #ok
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(gluco)
## W = 0.86549, p-value = 0.0573
```

```
leveneTest(gluco) #ok
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

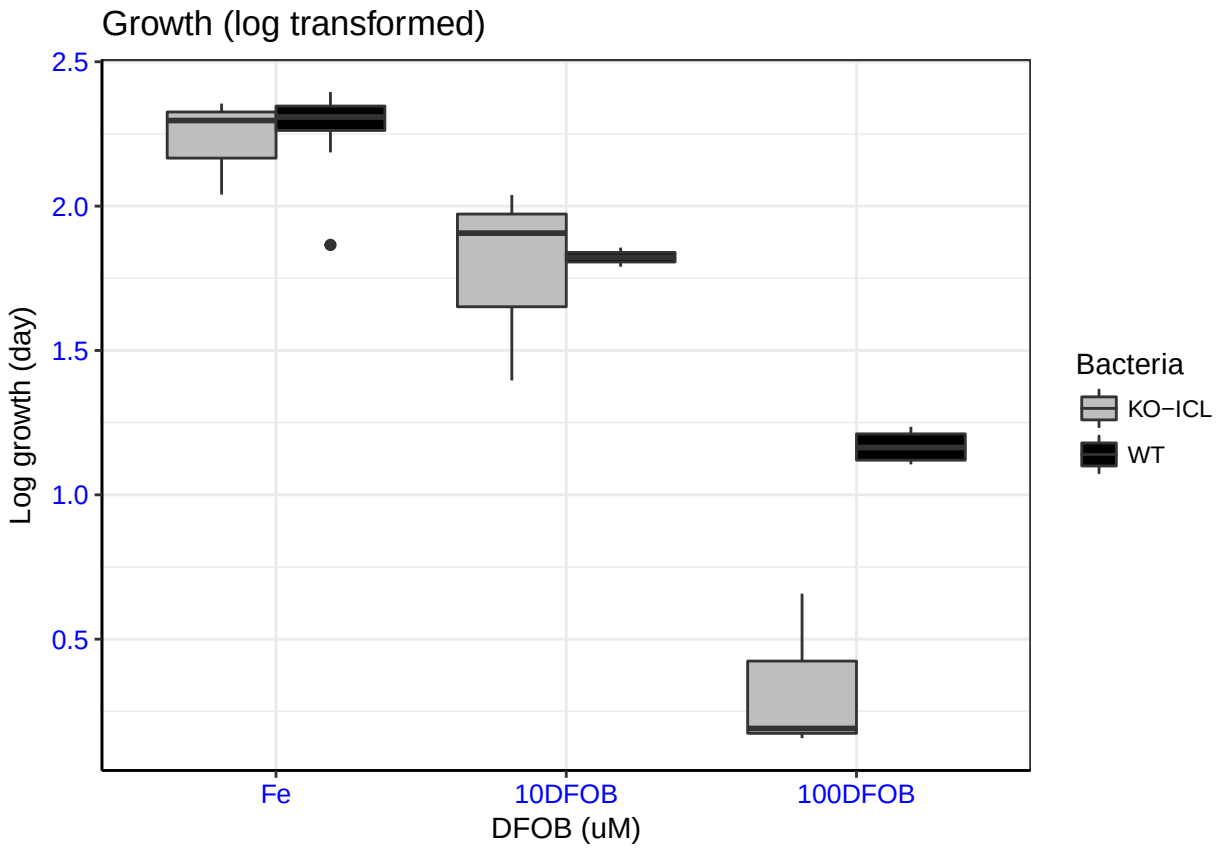
```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  1  1.4108 0.2624
##       10
```

Growth Rate

```
data1$strain=as.factor(data1$strain)
data1$iron=as.factor(data1$iron)
data1$growth=as.numeric(as.character(data1$growth))
data1$respiration=as.numeric(as.character(data1$respiration))

### graph ###
data1$loggrowth=log(data1$growth)

graph=ggplot(data1, aes(x=iron, y=loggrowth, fill=strain))+geom_boxplot()
graph=graph+theme_bw()
graph=graph+scale_x_discrete("DFOB (uM)", limits=c("Fe", "10DFOB", "100DFOB"))+scale_y_continuous("Log g
graph=graph+ggtitle("Growth (log transformed)")
graph=graph+scale_fill_manual(values=c("grey","black"), name="Bacteria", labels=c("KO-ICL", "WT"))
graph=graph+theme(axis.line = element_line(colour = "black",
                                           size = 0.5, linetype = "solid"))
graph=graph+theme(axis.text = element_text(color="blue",size=10))
graph
```




```
##anova - untransformed data
orig=lm(growth ~ iron * strain,
            data = data1,
            contrasts = list(iron = "contr.sum", strain = "contr.poly"))
gg <- Anova(orig, type = 3)

##anova - log transformed data
model=lm(loggrowth ~ iron * strain,
            data = data1,
            contrasts = list(iron = "contr.sum", strain = "contr.poly"))

gg <- Anova(model, type = 3)
kable(gg, digits = 10,results = 'asis')
```

	Sum Sq	Df	F value	Pr(>F)
(Intercept)	60.7295529	1	1949.45240	0.0000000000
iron	10.5269653	2	168.96072	0.0000000000
strain	0.5228526	1	16.78386	0.0004762854
iron:strain	0.8446201	2	13.55639	0.0001457171
Residuals	0.6853464	22	NA	NA

```
#check normality
shapiro.test(residuals(model))  #ok
```

```
##
##  Shapiro-Wilk normality test
```

```
##
## data:  residuals(model)
## W = 0.93603, p-value = 0.08762
```

```
leveneTest(model) #ok
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  5  0.7502 0.5947
##       22
```

```
rrr <- aov(loggrowth ~ iron*strain, data = data1)
TukeyHSD(rrr)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = loggrowth ~ iron * strain, data = data1)
##
## $iron
##                      diff       lwr       upr    p adj
## 10DFOB-100DFOB 0.9912548 0.7445820 1.2379276 0.00e+00
## Fe-100DFOB     1.4375903 1.2346394 1.6405412 0.00e+00
## Fe-10DFOB      0.4463355 0.2321635 0.6605075 8.56e-05
##
## $strain
##            diff        lwr       upr     p adj
## WT-KO 0.2226462 0.08394267 0.3613498 0.0030452
##
## $`iron:strain`
##                                diff         lwr         upr     p adj
## 10DFOB:KO-100DFOB:KO   1.44543176  0.99650524  1.89435827 0.0000000
## Fe:KO-100DFOB:KO       1.90448779  1.52507563  2.28389994 0.0000000
## 100DFOB:WT-100DFOB:KO  0.83210951  0.41217721  1.25204182 0.0000433
## 10DFOB:WT-100DFOB:KO   1.48806013  1.03913361  1.93698664 0.0000000
## Fe:WT-100DFOB:KO       1.92060096  1.54837076  2.29283116 0.0000000
## Fe:KO-10DFOB:KO        0.45905603  0.07964388  0.83846818 0.0118421
## 100DFOB:WT-10DFOB:KO  -0.61332224 -1.03325454 -0.19338994 0.0019212
## 10DFOB:WT-10DFOB:KO    0.04262837 -0.40629814  0.49155488 0.9996490
## Fe:WT-10DFOB:KO        0.47516921  0.10293900  0.84739941 0.0073495
## 100DFOB:WT-Fe:KO      -1.07237827 -1.41699647 -0.72776007 0.0000000
## 10DFOB:WT-Fe:KO       -0.41642766 -0.79583981 -0.03701551 0.0259753
## Fe:WT-Fe:KO            0.01611318 -0.26844594  0.30067229 0.9999725
## 10DFOB:WT-100DFOB:WT   0.65595061  0.23601831  1.07588291 0.0009104
## Fe:WT-100DFOB:WT       1.08849145  0.75179656  1.42518633 0.0000000
## Fe:WT-10DFOB:WT        0.43254084  0.06031064  0.80477104 0.0166068
```

Summary Growth Rate Data Analysis

Data reveals a skewed data set where differences between the KO and WT are shown particularly in 100 DFOB conditions, while differences do not seem to be present for 10 DFOB and Fe conditions. In this particular case, it is appropriate to conduct a log transformation in order to discern a pattern. This is highlighted by the fact that a log transformation gives a normal distribution of the data ($p = 0.08762$) using a shapiro wilk test. Homoskedasticity was also met ($p = 0.5947$).

The first ANOVA table shows the data untransformed. While the second ANOVA table shows data after transformation. This is followed by tests to check whether data meet the assumptions of the ANOVA (**note that assumptions of normality were not met for untransformed data)

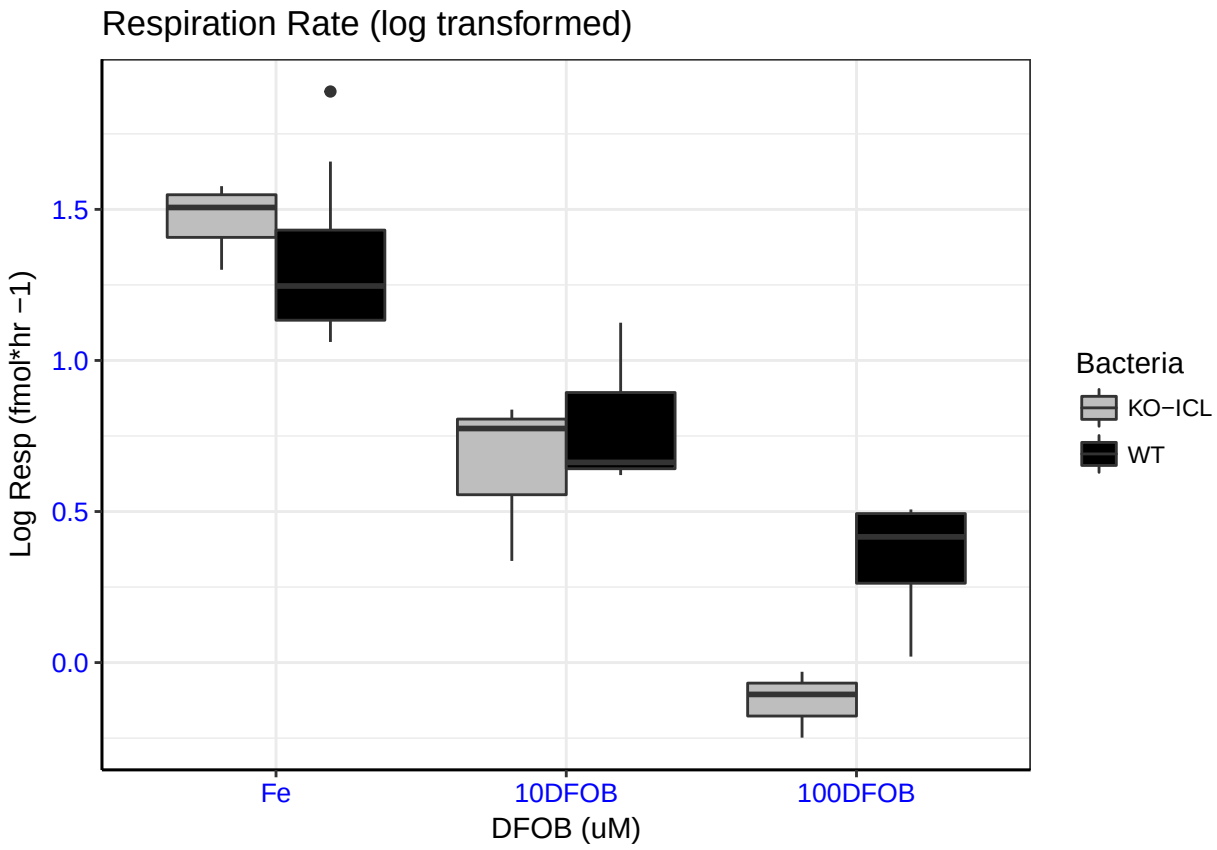
A tukey test reveals a significant differences betweeen WT and KO grown in strong iron limitation of 100uM DFOB ($p = 0.0000433$).

Respiration Rate

```
data2=read.table('icl_alldates.txt', na.strings="NA", sep='\t', header=T, dec=',')

data1$strain=as.factor(data1$strain)
data1$iron=as.factor(data1$iron)
data1$respiration=as.numeric(as.character(data1$respiration))

data1$logresp=log(data1$respiration)
graph2=ggplot(data1, aes(x=iron, y=logresp, fill=strain))+geom_boxplot()
graph2=graph2+scale_fill_manual(values=c("grey","black"), name="Bacteria", labels=c("KO-ICL", "WT"))
graph2=graph2+theme_bw()
graph2=graph2+scale_x_discrete("DFOB (uM)", limits=c("Fe", "10DFOB", "100DFOB"))+scale_y_continuous("Log
graph2=graph2+ggtitle("Respiration Rate (log transformed)")
graph2=graph2+theme(axis.line = element_line(colour = "black",
                                             size = 0.5, linetype = "solid"))
graph2=graph2+theme(axis.text = element_text(color="blue",size=10))
graph2
```



```
data1$logresp=as.numeric(as.character(data1$logresp))
r=lm(logresp ~ iron * strain,
            data = data1,
            contrasts = list(iron = "contr.sum", strain = "contr.poly"))

rr <- Anova(r, type = 3)
kable(rr, digits = 10,results = 'asis')
```

	Sum Sq	Df	F value	Pr(>F)
(Intercept)	13.1636128	1	249.794791	0.0000000000
iron	8.2481974	2	78.259546	0.0000000001
strain	0.1554684	1	2.950194	0.0999130426
iron:strain	0.4387217	2	4.162626	0.0292954996
Residuals	1.1593496	22	NA	NA

```
#check normality
shapiro.test(residuals(r))  #ok
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(r)
## W = 0.96411, p-value = 0.4343
```

```
leveneTest(r) #ok
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  5  0.6193 0.6865
##       22
```

```
rrr <- aov(logresp ~ iron*strain, data = data1)
TukeyHSD(rrr)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = logresp ~ iron * strain, data = data1)
##
## $iron
##                      diff       lwr       upr     p adj
## 10DFOB-100DFOB 0.5868913 0.2660624 0.9077202 0.0003991
## Fe-100DFOB     1.2593437 0.9953806 1.5233067 0.0000000
## Fe-10DFOB      0.6724524 0.3938949 0.9510098 0.0000121
##
## $strain
##             diff        lwr       upr     p adj
## WT-KO 0.07564552 -0.1047558 0.2560468 0.3939015
##
## $`iron:strain`
##                               diff         lwr        upr     p adj
## 10DFOB:KO-100DFOB:KO   0.7775760  0.19369090  1.3614611 0.0049290
## Fe:KO-100DFOB:KO       1.5987935  1.10532053  2.0922665 0.0000000
## 100DFOB:WT-100DFOB:KO  0.4677912 -0.07838332  1.0139657 0.1225341
```

```
## 10DFOB:WT-100DFOB:KO   0.9308250  0.34693995  1.5147101 0.0007186
## Fe:WT-100DFOB:KO       1.4635299  0.97939792  1.9476618 0.0000000
## Fe:KO-10DFOB:KO        0.8212175  0.32774453  1.3146905 0.0004298
## 100DFOB:WT-10DFOB:KO  -0.3097848 -0.85595932  0.2363897 0.5055363
## 10DFOB:WT-10DFOB:KO    0.1532490 -0.43063606  0.7371341 0.9611117
## Fe:WT-10DFOB:KO        0.6859539  0.20182191  1.1700858 0.0026468
## 100DFOB:WT-Fe:KO      -1.1310023 -1.57922142 -0.6827832 0.0000011
## 10DFOB:WT-Fe:KO       -0.6679685 -1.16144144 -0.1744955 0.0042034
## Fe:WT-Fe:KO           -0.1352636 -0.50536837  0.2348411 0.8600070
## 10DFOB:WT-100DFOB:WT   0.4630339 -0.08314063  1.0092084 0.1289673
## Fe:WT-100DFOB:WT       0.9957387  0.55782486  1.4336525 0.0000057
## Fe:WT-10DFOB:WT        0.5327048  0.04857287  1.0168368 0.0254862
```

Summary Respiration Rate Data Analysis

Data again was log transformed due to skewedness of data. Assumptions of the ANOVA were met through a shapiro test (p = 0.4343) on the log transformed residuals to test for normality while a levene test (p = 0.6865)was conducted to confirm homoskedastiicty.

Data is significant for interaction effects between strain and iron conditions but a tukey test reveals that these differences are not statistically significant between WT and KO grown in 100DFOB conditions (p = 0.1225341). It is therefore tricky to discern how ICL affects the respiration rate under iron limitation.