Acetilation

Lucas Michel Todó

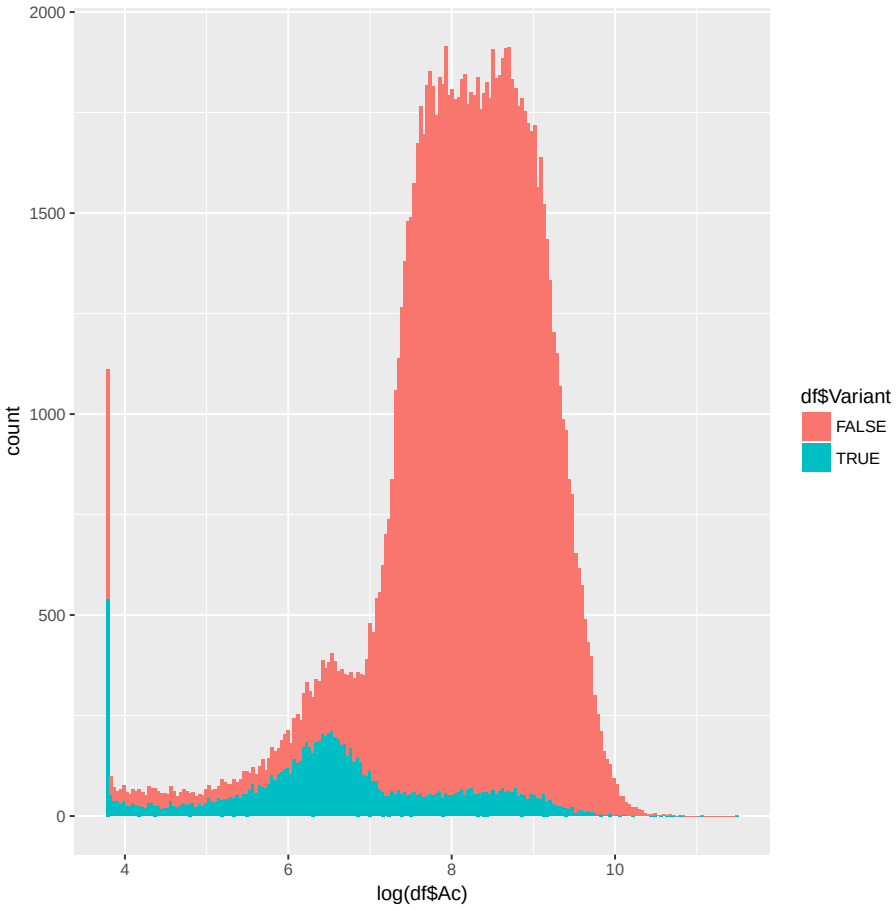
December 28, 2017

Contents

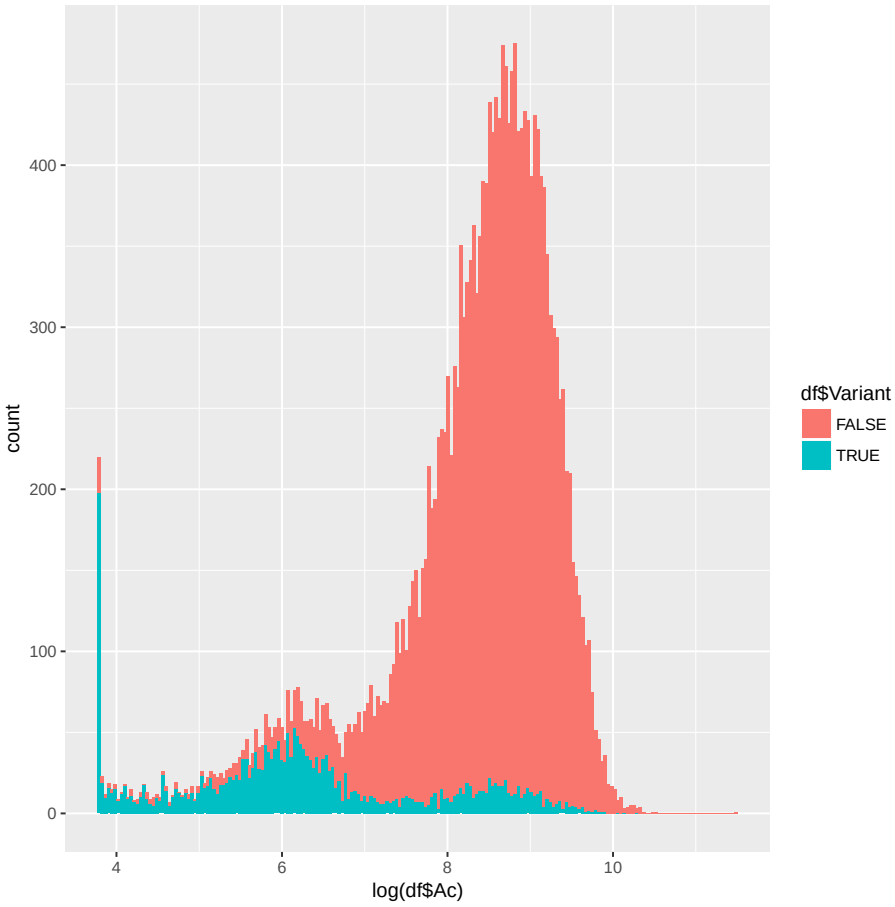
```
##
## Attaching package:  'dplyr'
## The following objects are masked from 'package:stats':
##
##    filter, lag
## The following objects are masked from 'package:base':
##
##    intersect, setdiff, setequal, union
## Classes and Methods for R developed in the
## Political Science Computational Laboratory
## Department of Political Science
## Stanford University
## Simon Jackman
## hurdle and zeroinfl functions by Achim Zeileis
## Loading required package:  XLConnectJars
## XLConnect 0.2-13 by Mirai Solutions GmbH [aut],
##  Martin Studer [cre],
##  The Apache Software Foundation [ctb, cph] (Apache POI),
##  Graph Builder [ctb, cph] (Curvesapi Java library)
## http://www.mirai-solutions.com ,
## http://miraisolutions.wordpress.com
```

1 Density Plots

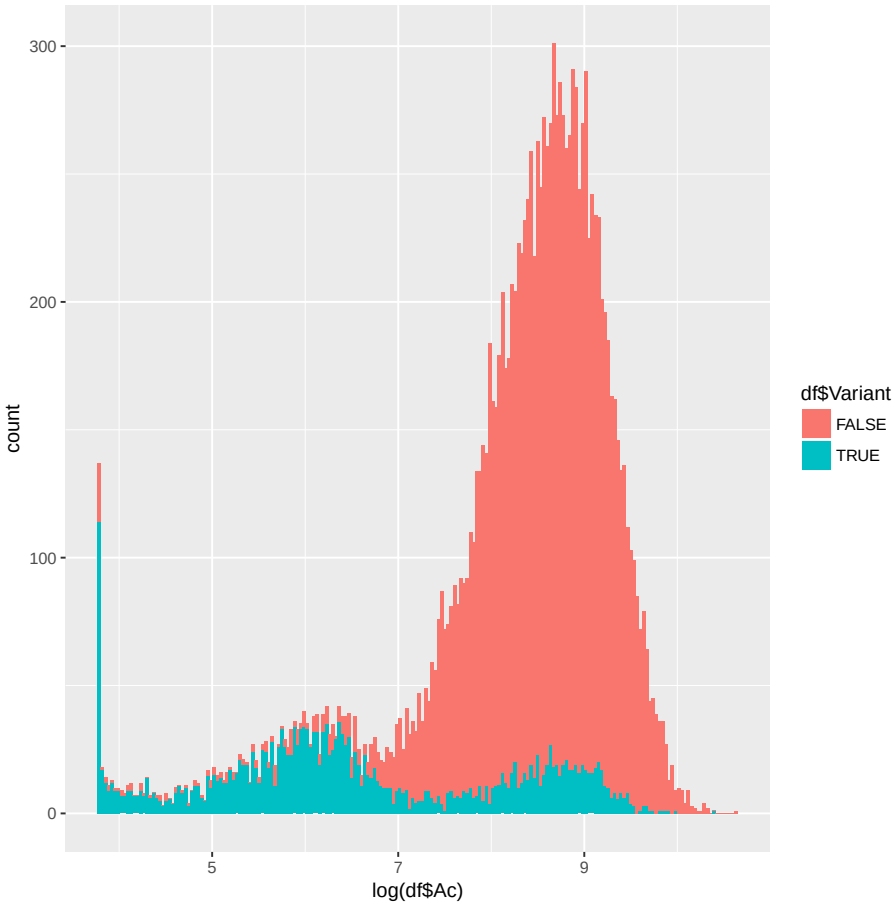
1.1 log(Ac) All


2000
1500
1000
500
0
count
4
6
8
10
log(df$Ac)
df$Variant
FALSE
TRUE

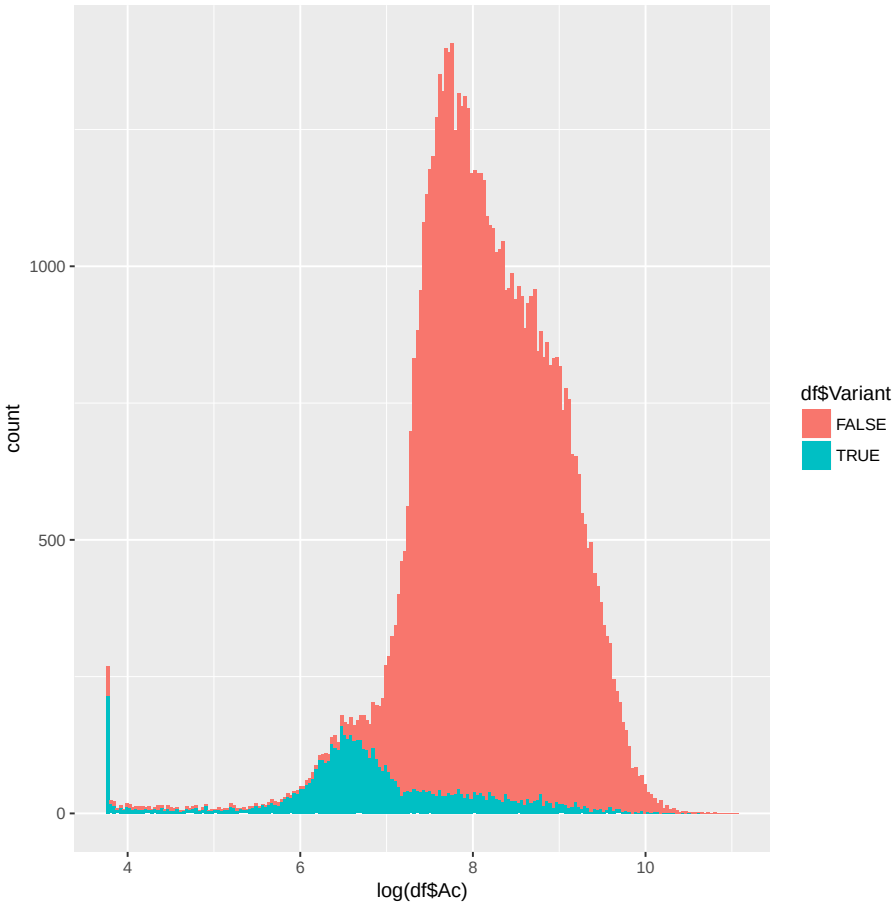
1.2 log(Ac) 5'



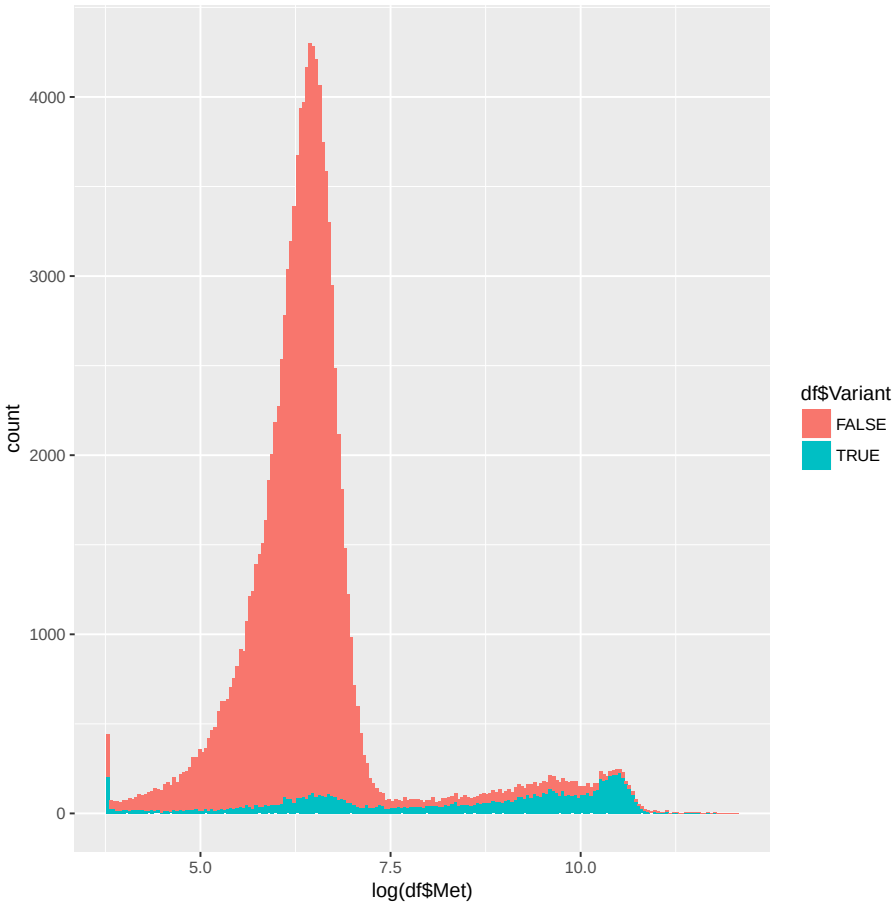
1.3 log(Ac) 3'



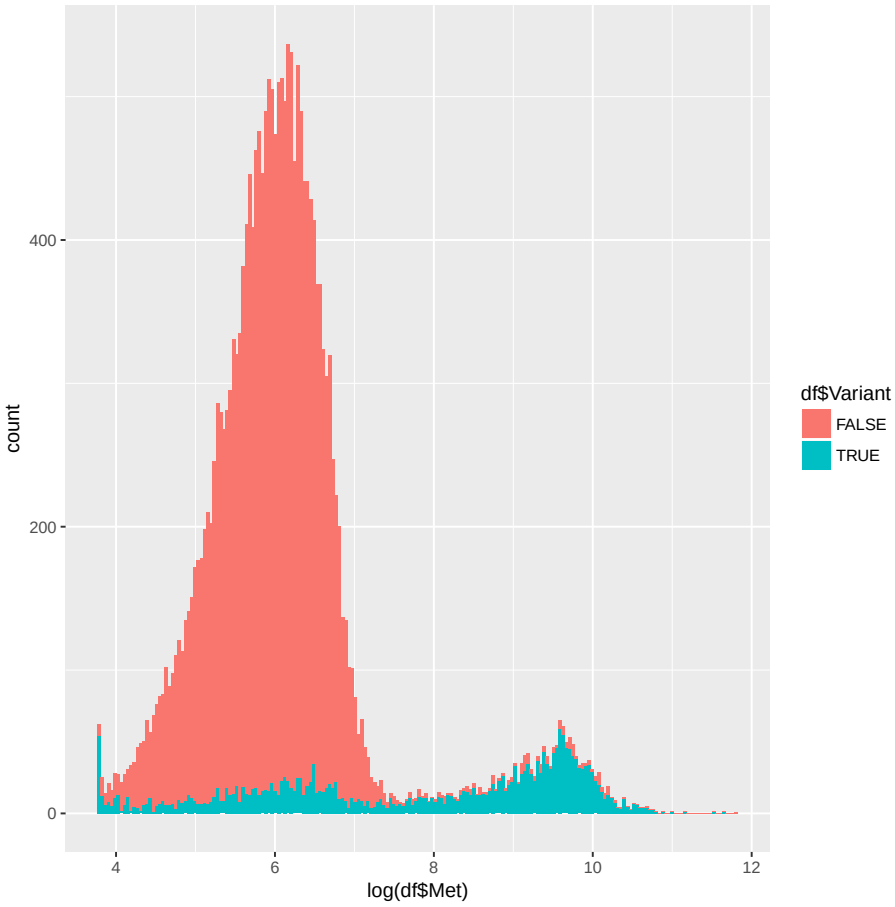
1.4 log(Ac) ORF



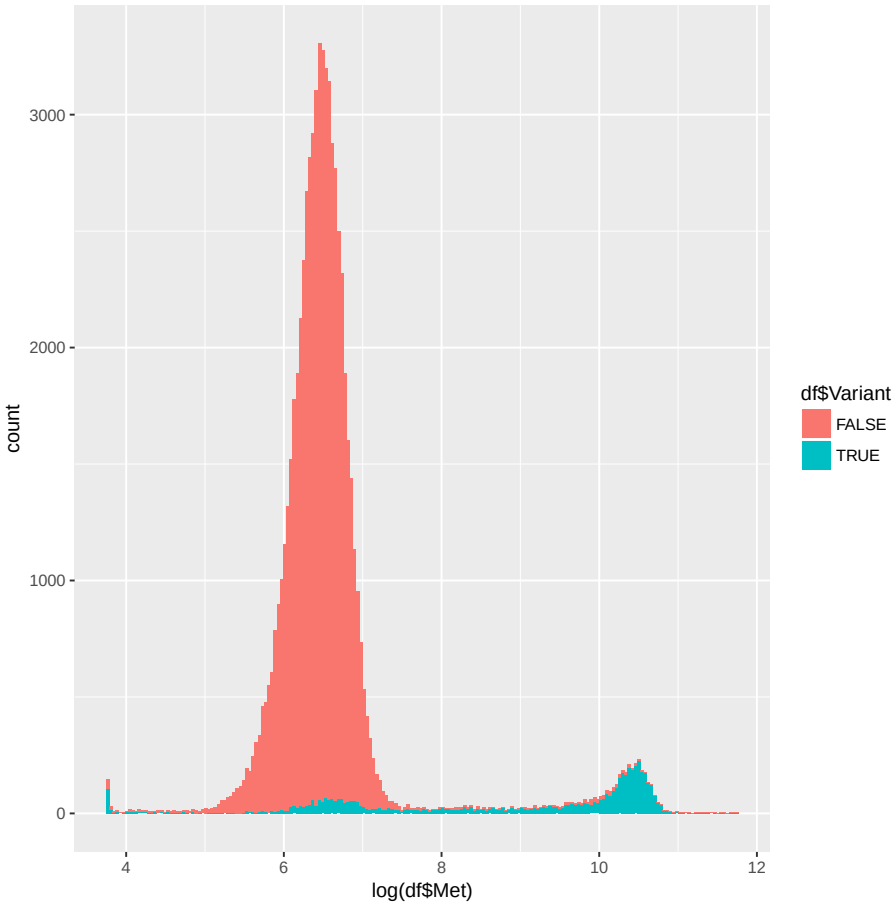
1.5 log(Met) All



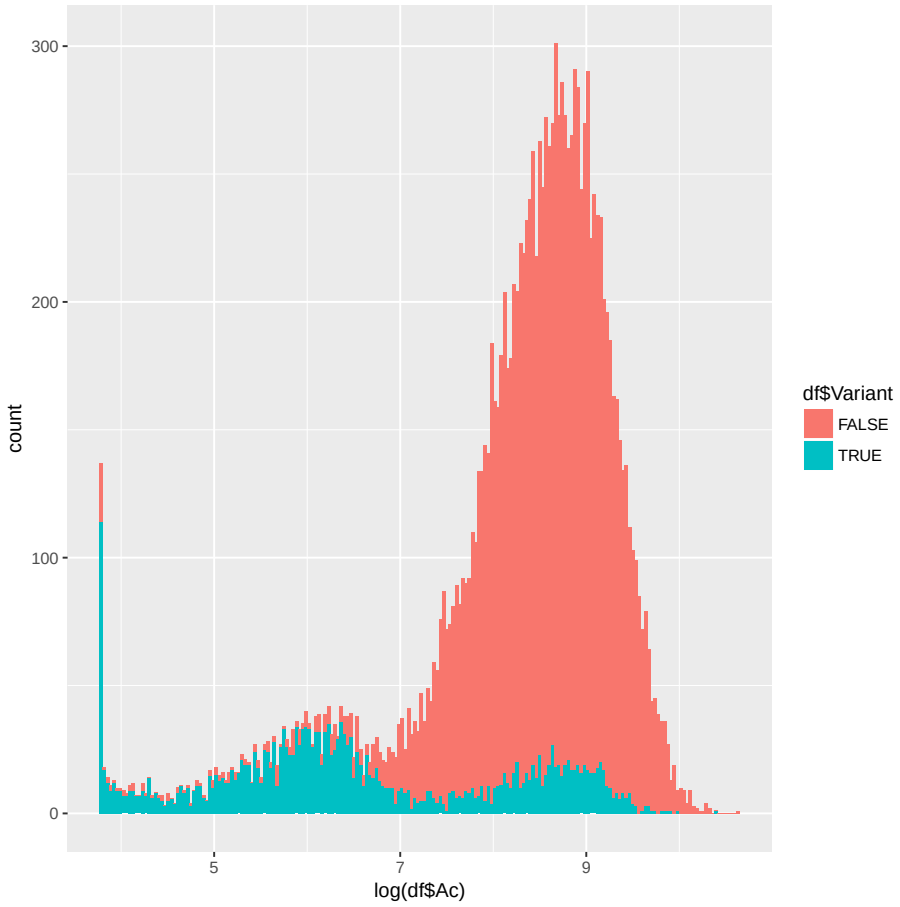
1.6 log(Met) 5'



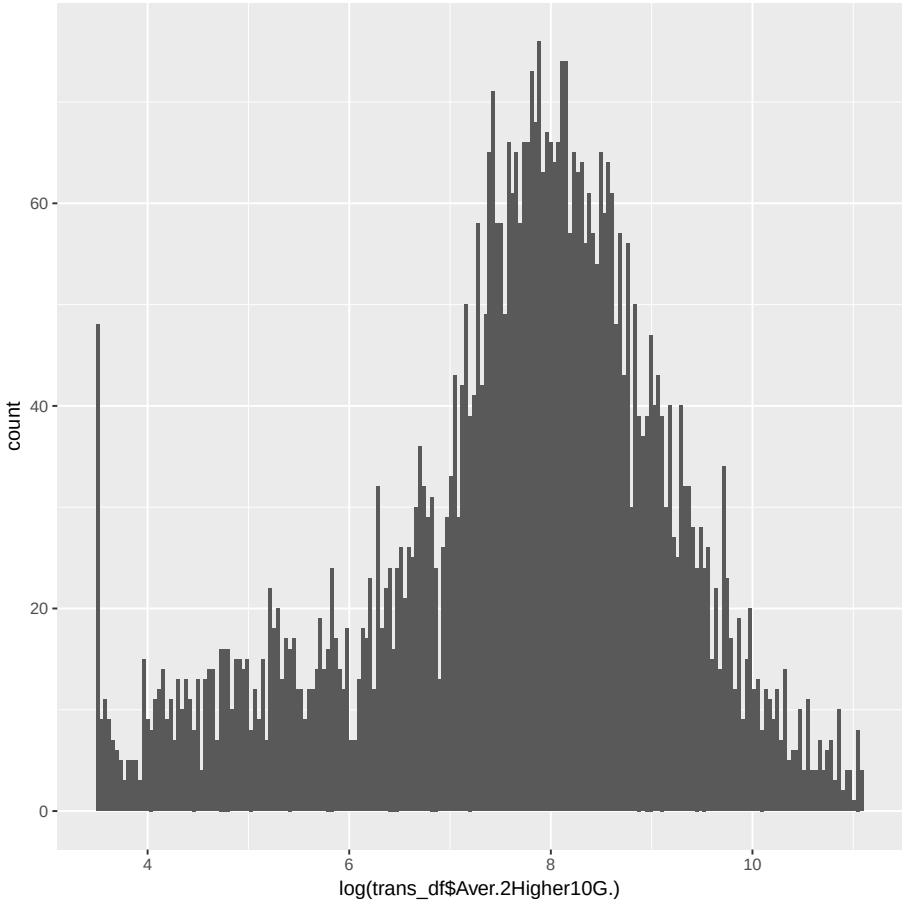
1.7 log(Met) ORF

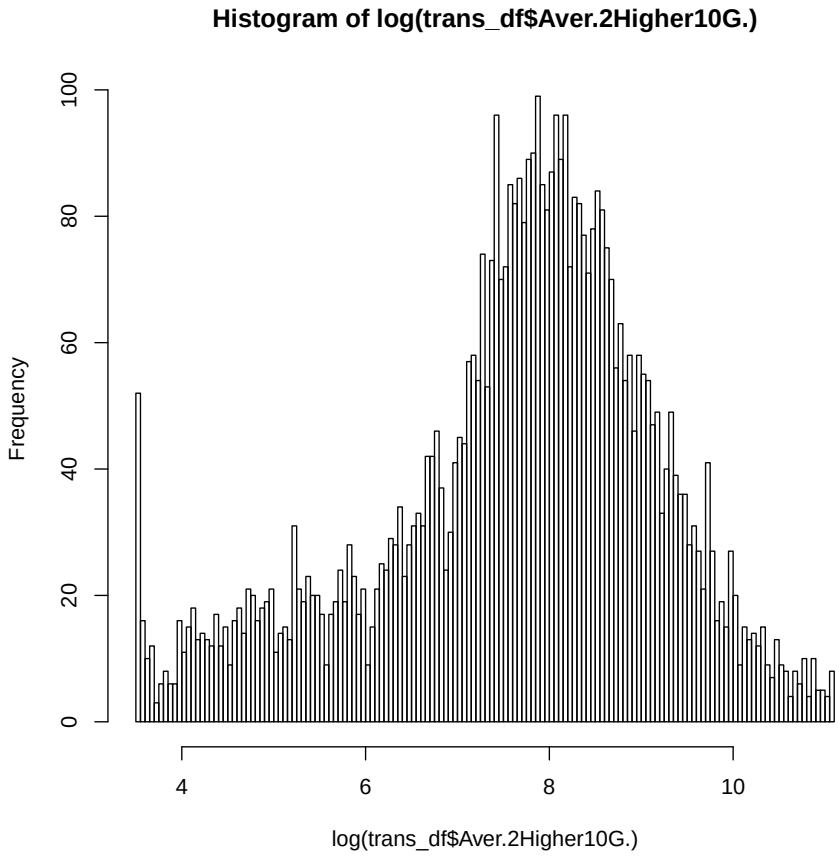


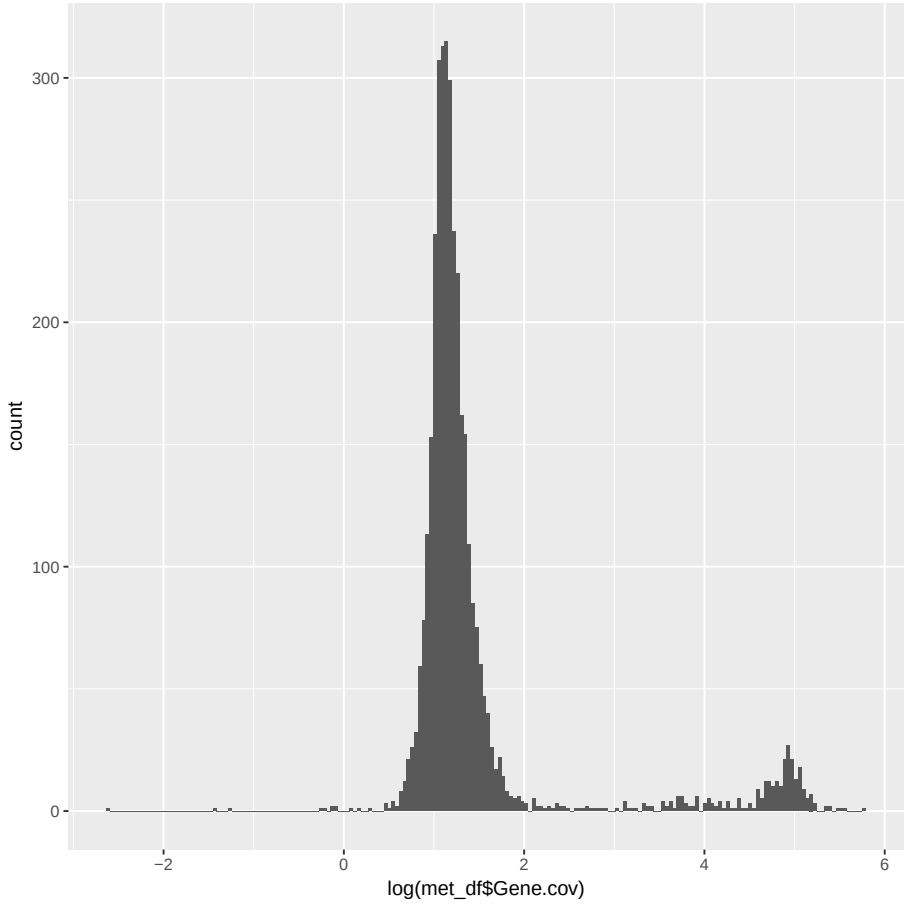
1.8 log(Ac) 3'




```
##          Chom Start Stop      Met       Ac  Type Variant
## 1 Pf3D7_01_v3     0  200 1658.627 395.7799 other   FALSE
## 2 Pf3D7_01_v3   200  400 3100.228 885.5960 other   FALSE
## 3 Pf3D7_01_v3   400  600 6365.753 654.0001 other   FALSE
## 4 Pf3D7_01_v3   600  800 6512.168 667.2569 other   FALSE
## 5 Pf3D7_01_v3   800 1000 4864.506 515.7165 other   FALSE
## 6 Pf3D7_01_v3  1000 1200 6566.868 620.1471 other   FALSE
##          Col1 Aver.2Higher10G.            ID
## 1   PF14_0010         37.90000 PF3D7_1401000
## 2   PF07_0048        450.92917 PF3D7_0711700
## 3 MAL13P1.490         71.21667 PF3D7_1372500
## 4    PFC0110w      53468.68333 PF3D7_0302200
## 5    PFF1580c        310.15208 PF3D7_0632500
## 6    PFD0090c        969.11667 PF3D7_0402000
```


count
60
40
20
0
4
6
8
10
log(trans_df$Aver.2Higher10G.)


Histogram of log(trans_df$Aver.2Higher10G.)
Frequency
0
20
40
60
80
100
4
6
8
10
log(trans_df$Aver.2Higher10G.)

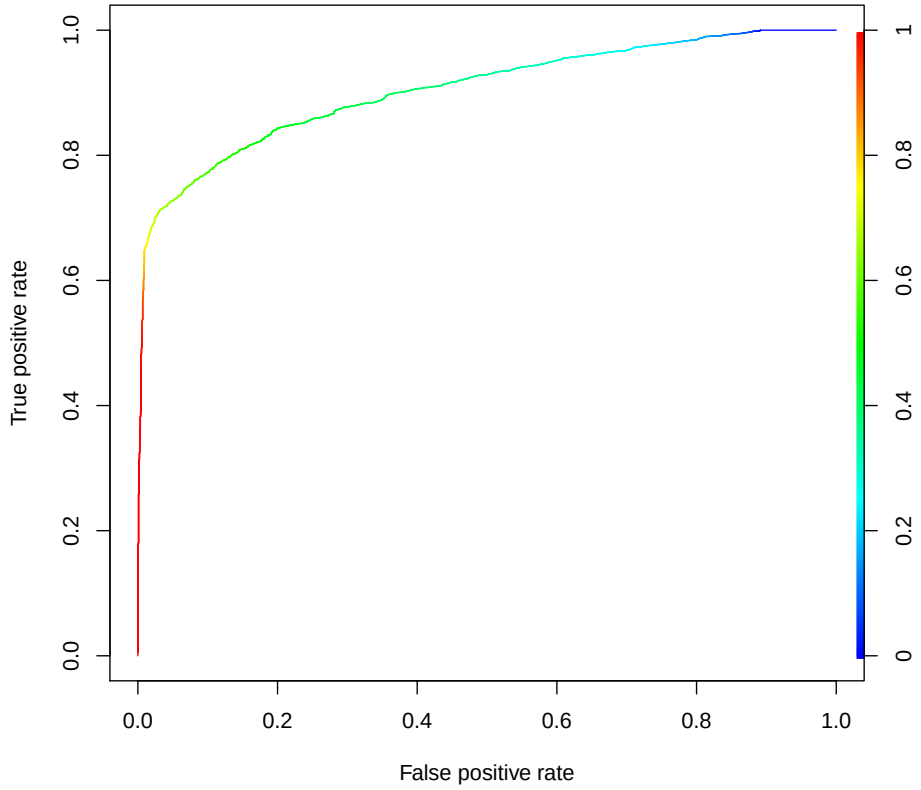



```
##
## FALSE  TRUE
##   214    74
##
## FALSE  TRUE
##    61    74
```

```
##
##  FALSE   TRUE
## 105691  10966
## Analysis of Deviance Table
##
## Model 1: Variant ~ Ac + Met + Type + Start + Stop
## Model 2: Variant ~ Ac + Met
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
```

```
## 1      8479     6267.7
## 2      8483     7801.9 -4  -1534.2 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Call:
## glm(formula = Variant ~ Ac + Met + Type + Start + Stop, family = binomial(link = "logit")
##     data = train_df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -5.7070  -0.7818   0.0012   0.5811   3.1900
##
## Coefficients: (1 not defined because of singularities)
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.340e+00  9.252e-02  14.480  < 2e-16 ***
## Ac          -1.613e-04  8.978e-06 -17.971  < 2e-16 ***
## Met          5.312e-04  3.074e-05  17.281  < 2e-16 ***
## Type5prima  -5.331e-01  9.249e-02  -5.764  8.2e-09 ***
## TypeORF     -1.404e+00  8.137e-02 -17.261  < 2e-16 ***
## Typeother   -1.458e+02  3.667e+06   0.000        1
## Start       -3.313e-07  4.023e-08  -8.234  < 2e-16 ***
## Stop                NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 11134.0  on 8485  degrees of freedom
## Residual deviance:  6267.7  on 8479  degrees of freedom
## AIC: 6281.7
##
## Number of Fisher Scoring iterations: 10
##
##         FALSE TRUE
##   FALSE  2713  558
##   TRUE   1051 4523
## [1] "Accuracy 0.818089315997739"
## [1] "Accuracy of null model 0.369813453928773"

## Loading required package:  gplots
##
## Attaching package:  'gplots'
## The following object is masked from 'package:stats':
##
##     lowess
```




```
##
## 3prima 5prima    ORF
##    273    220     65
##
## 3prima 5prima    ORF
##     98    174    779
##           llh       llhNull            G2      McFadden          r2ML
## -3133.8490369 -5566.9799695  4866.2618652     0.4370648     0.4364199
##          r2CU
##     0.5972388
```