Acetilation

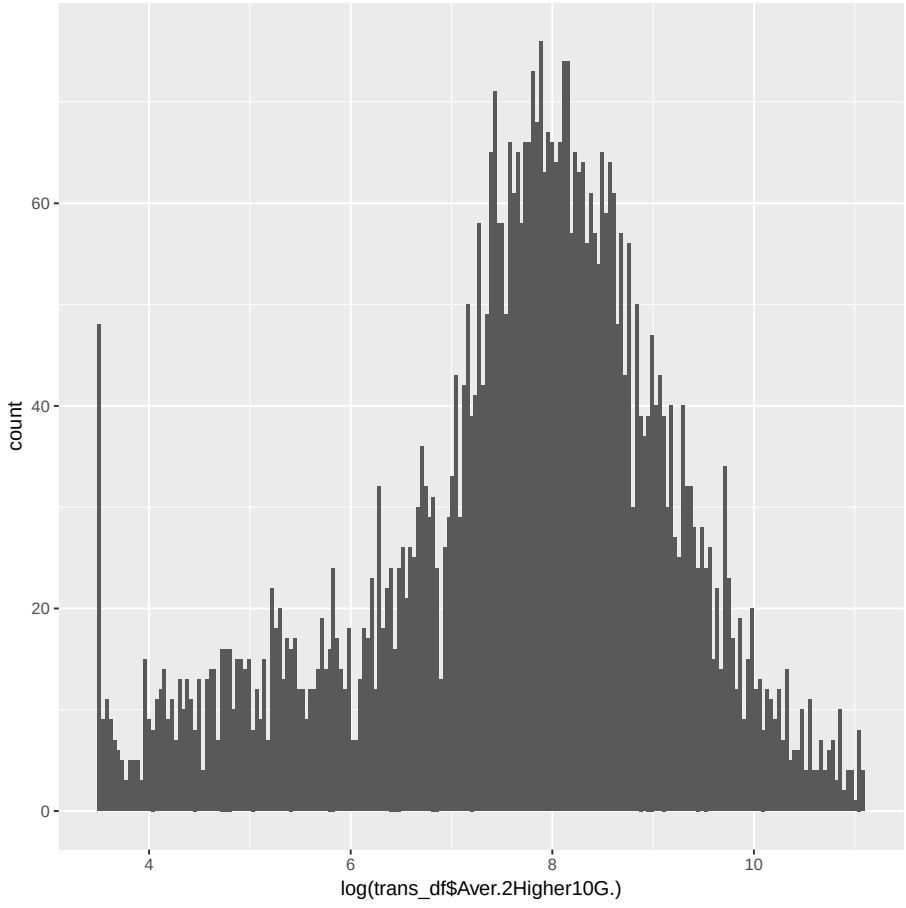
Lucas Michel Todó

January 22, 2018

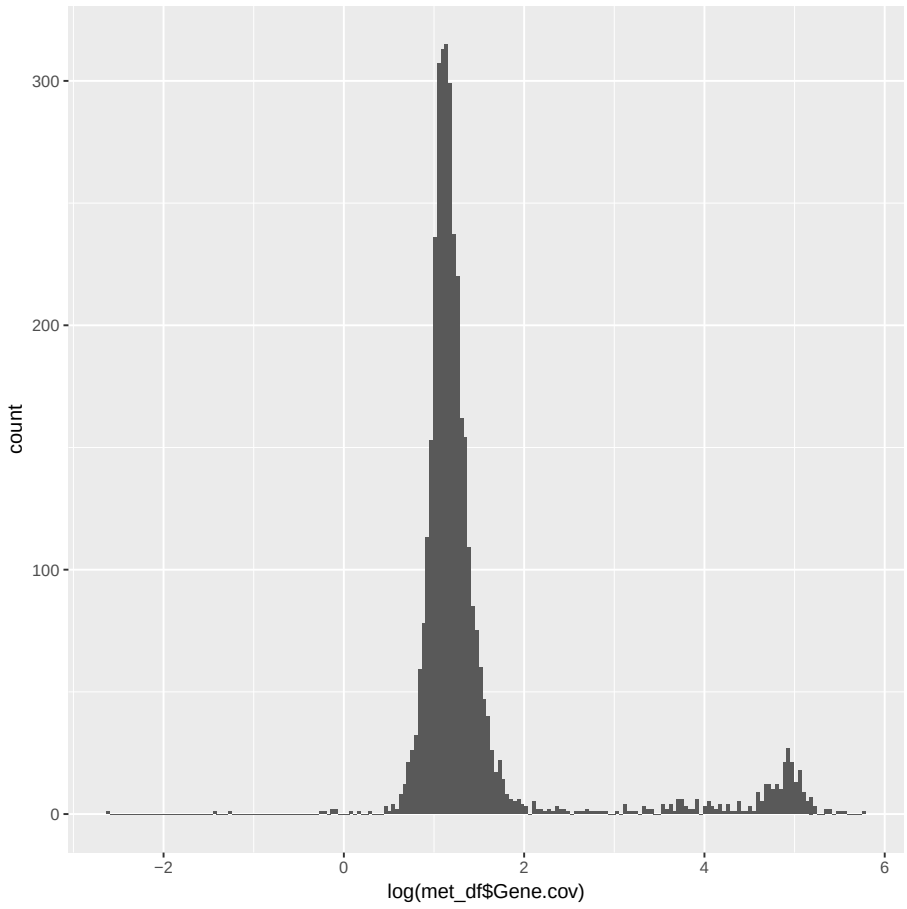
Contents

1 Histograms **3**
- 1.1 log(Ac) All . . . 3
- 1.2 log(Ac) 5' . . . 5
- 1.3 log(Ac) 3' . . . 6
- 1.4 log(Ac) ORF . . . 7
- 1.5 log(Met) All . . . 8
- 1.6 log(Met) 5' . . . 9
- 1.7 log(Met) ORF . . . 10
- 1.8 log(Ac) 3' . . . 11

2 Metilation/Acetilation plots **11**


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

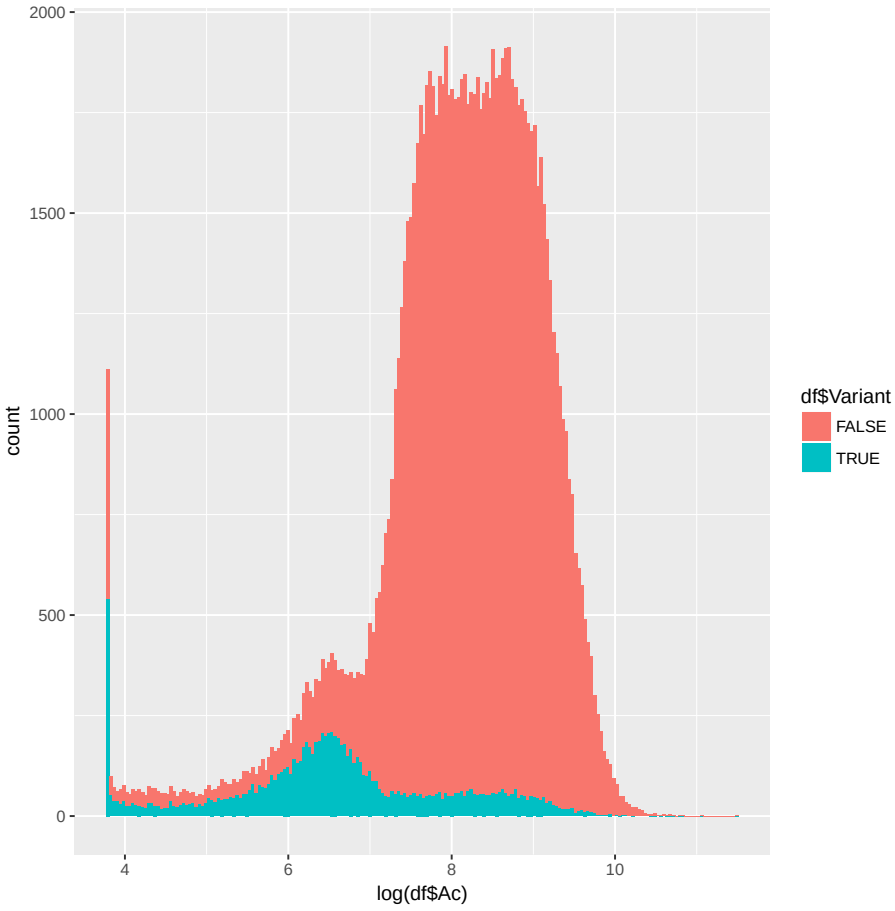
@



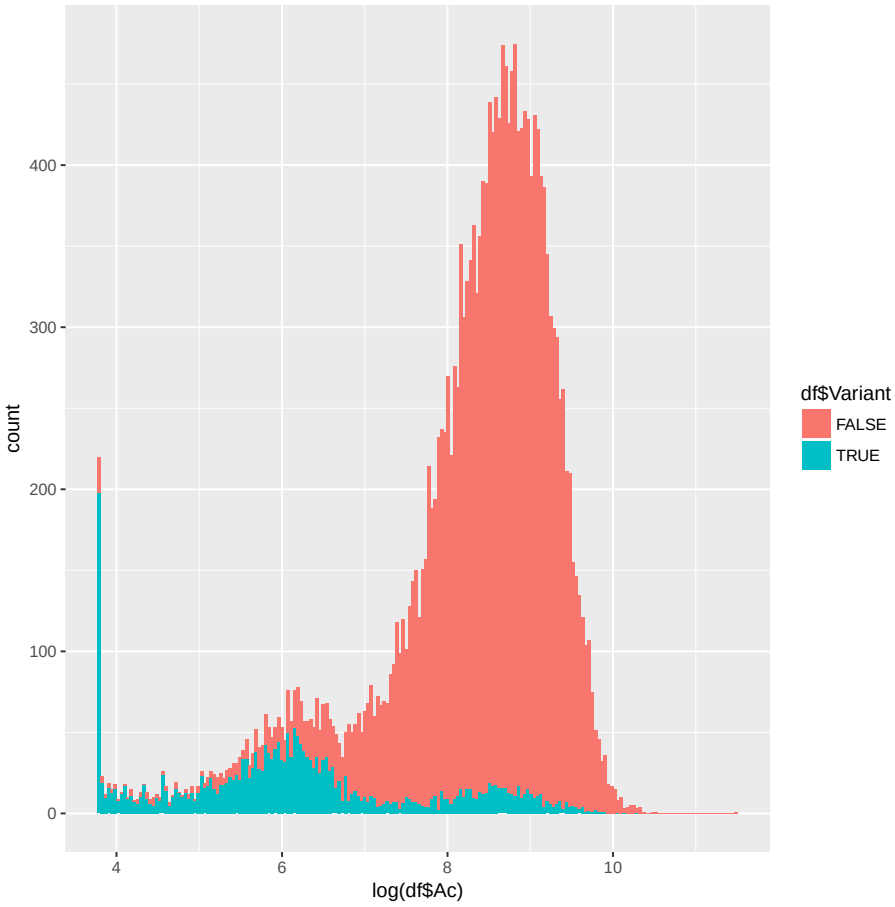

```
##
## FALSE  TRUE
##   214    74
##
## FALSE  TRUE
##    61    74
```

1 Histograms

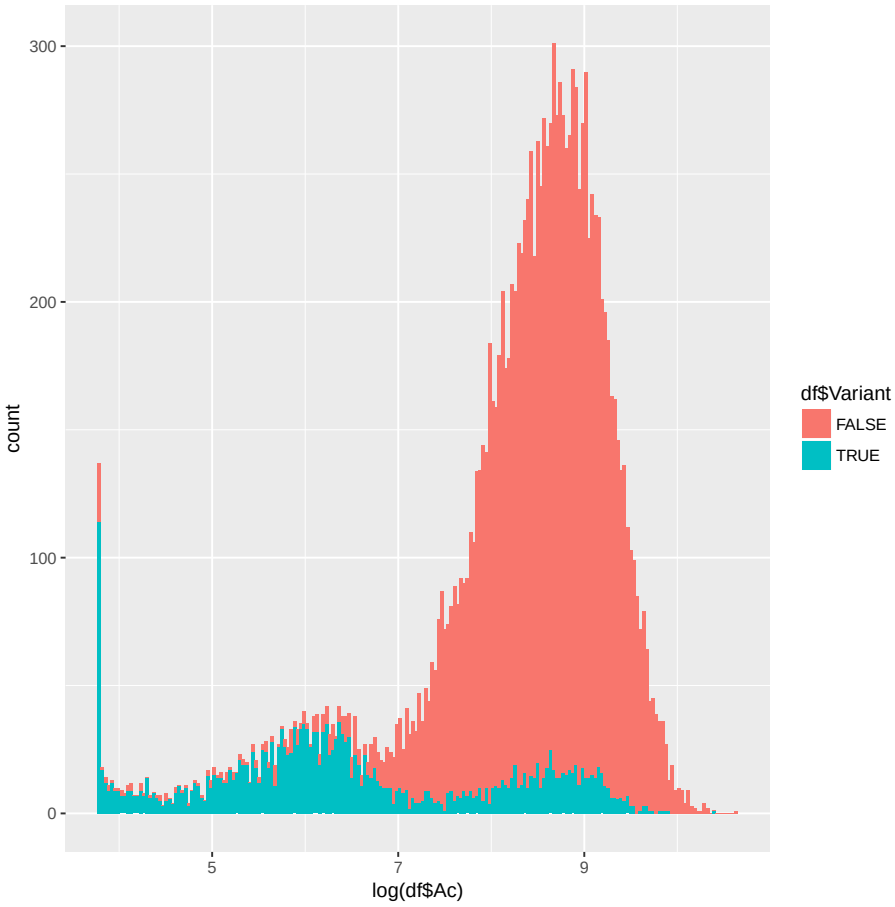
1.1 log(Ac) All


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

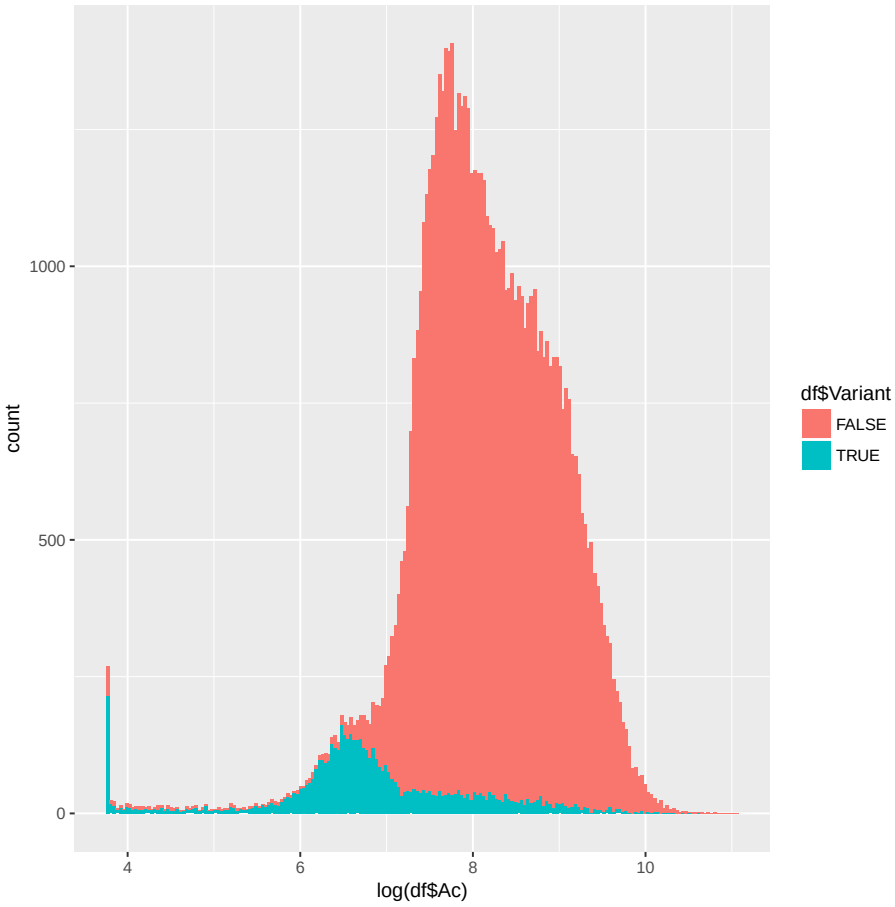
1.2 log(Ac) 5'



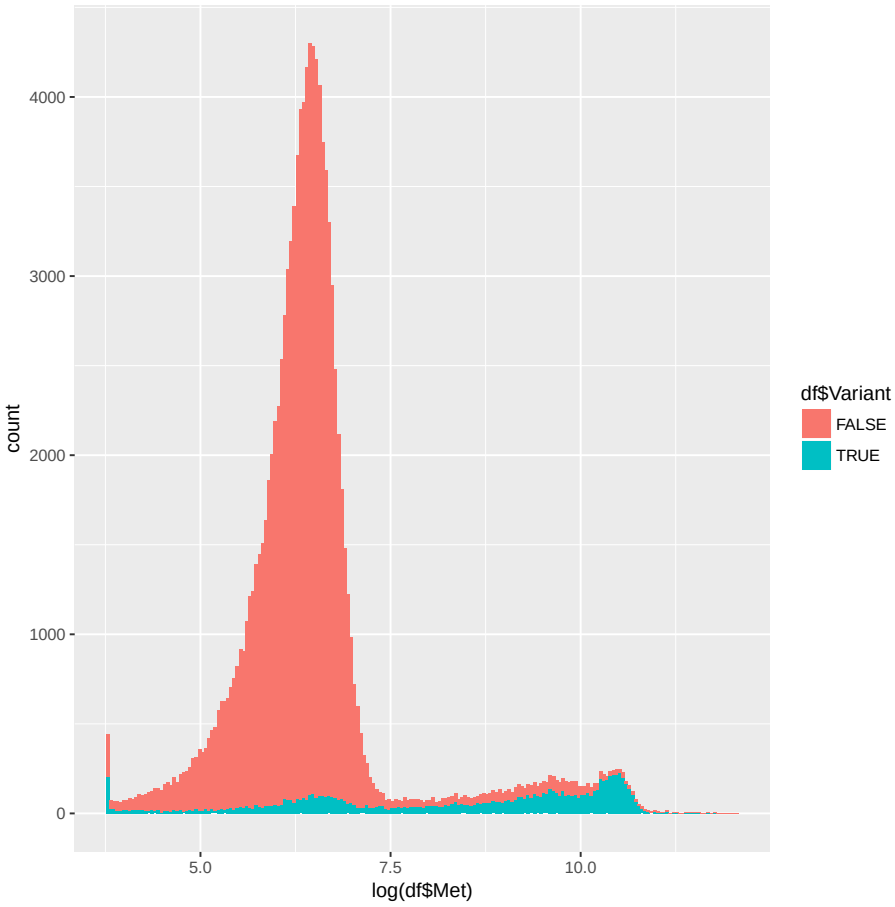
1.3 log(Ac) 3'



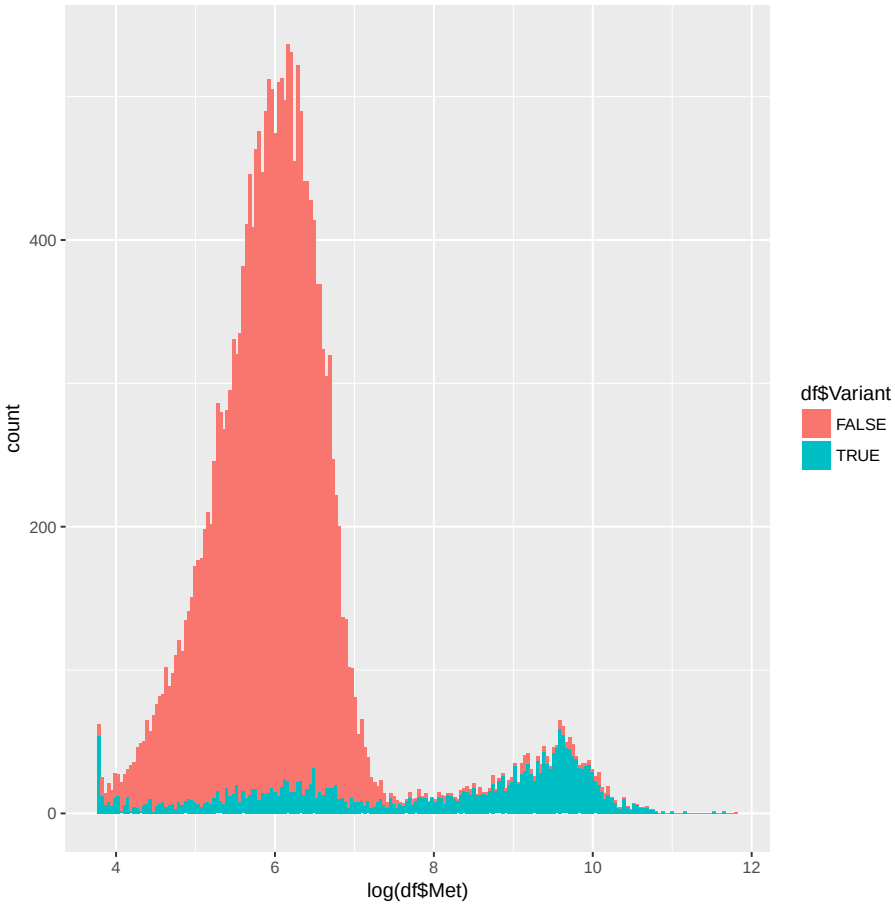
1.4 log(Ac) ORF



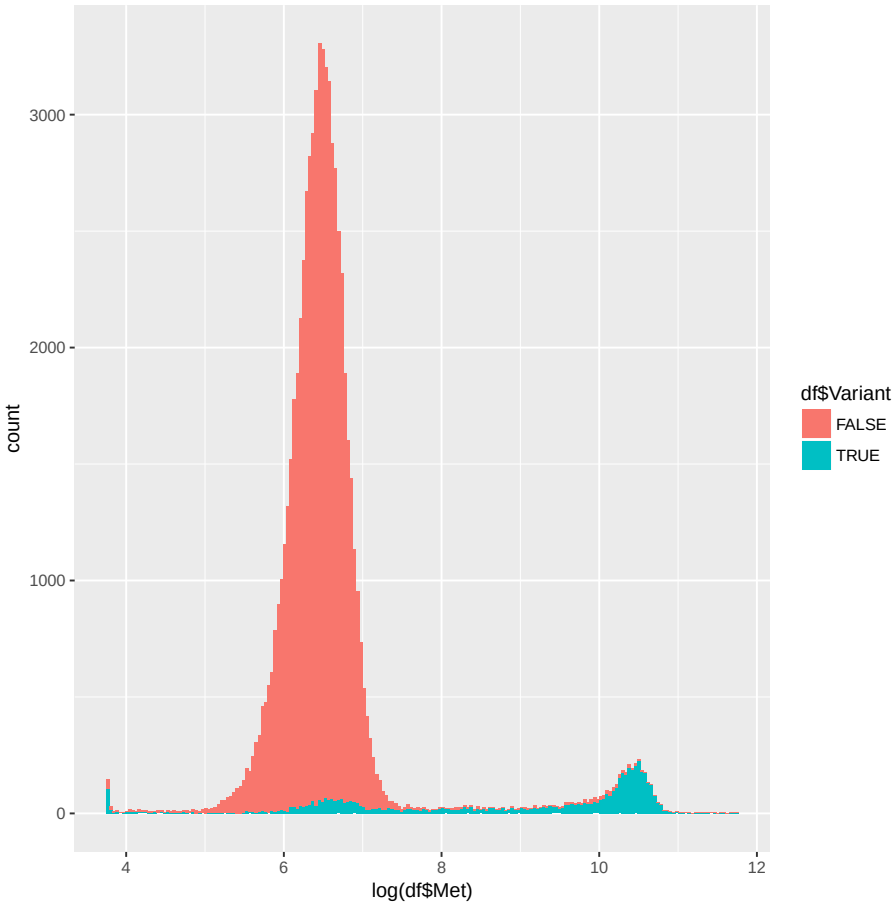
1.5 log(Met) All



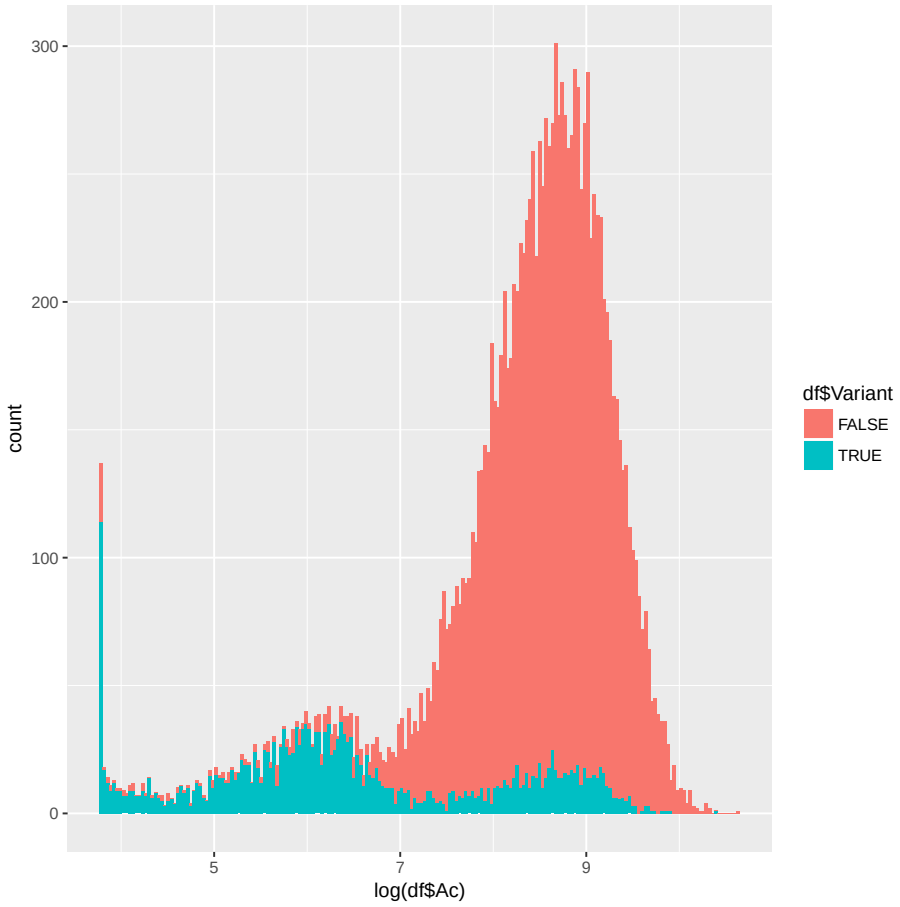
1.6 log(Met) 5'



1.7 log(Met) ORF



1.8 log(Ac) 3'



2 Metilation/Acetilation plots

Genes have been classified as variant or not on the basis of a previous study.

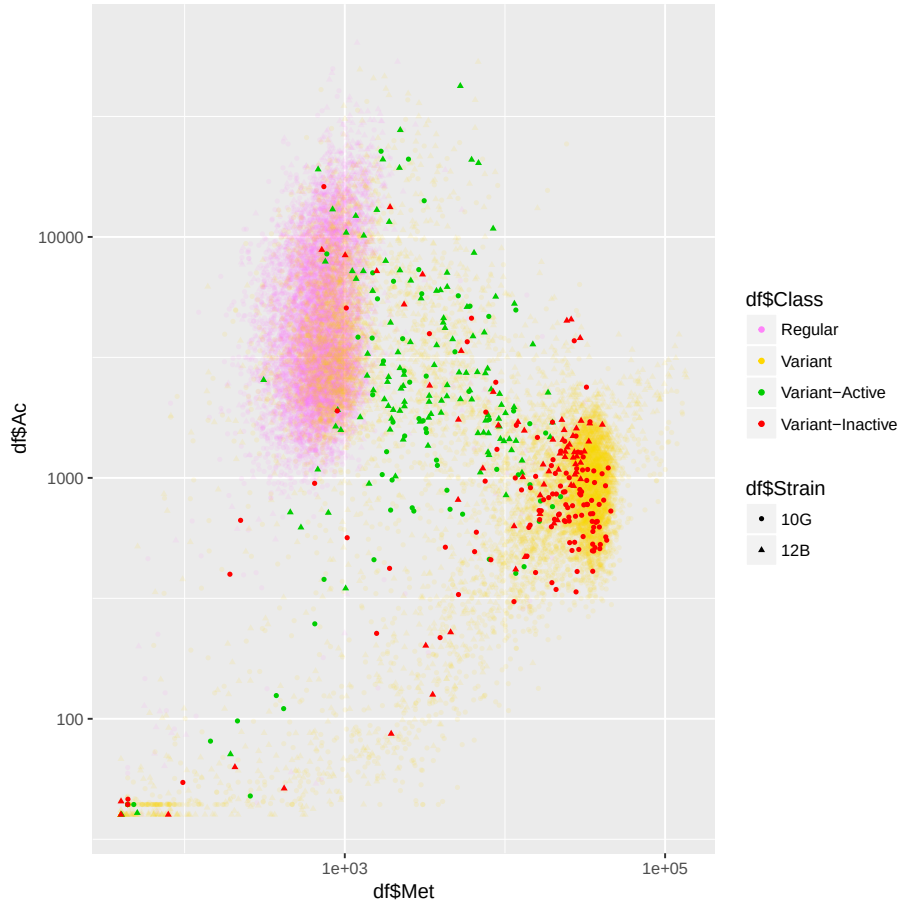
Fragments have been set to "noexprs" if transcription value is under threshold (¡4). Fragments have been set to "silenced" if methilation value is over threshold(¿3).

```
##
##  FALSE   TRUE
## 114212   2445
```

10000
1000
100
df$Ac
100
1000
10000
df$Met
df$noexprs
FALSE
TRUE
df$silenced
FALSE
TRUE

Here we have highlited those genes that are "variant" (as shown in previous studies) and differentially expressed between 10G and 1.2B. Variant genes which are overexpressed have been set to "variant active" and viceversa.

10G and 1.2B fragments have been pooled toghether for the grafic. Blurred point correspond to fragments representing the rest of genes (those that are not diferentially expressed).




```
##
##  FALSE   TRUE
## 105873  10784
## Analysis of Deviance Table
##
## Model 1: Variant ~ Ac + Met + Type + Start + Stop + silenced + noexprs
## Model 2: Variant ~ Ac + Met + Type + Start + Stop
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1      8477     5498.4
## 2      8479     6141.6 -2  -643.16 < 2.2e-16 ***
## ---
```

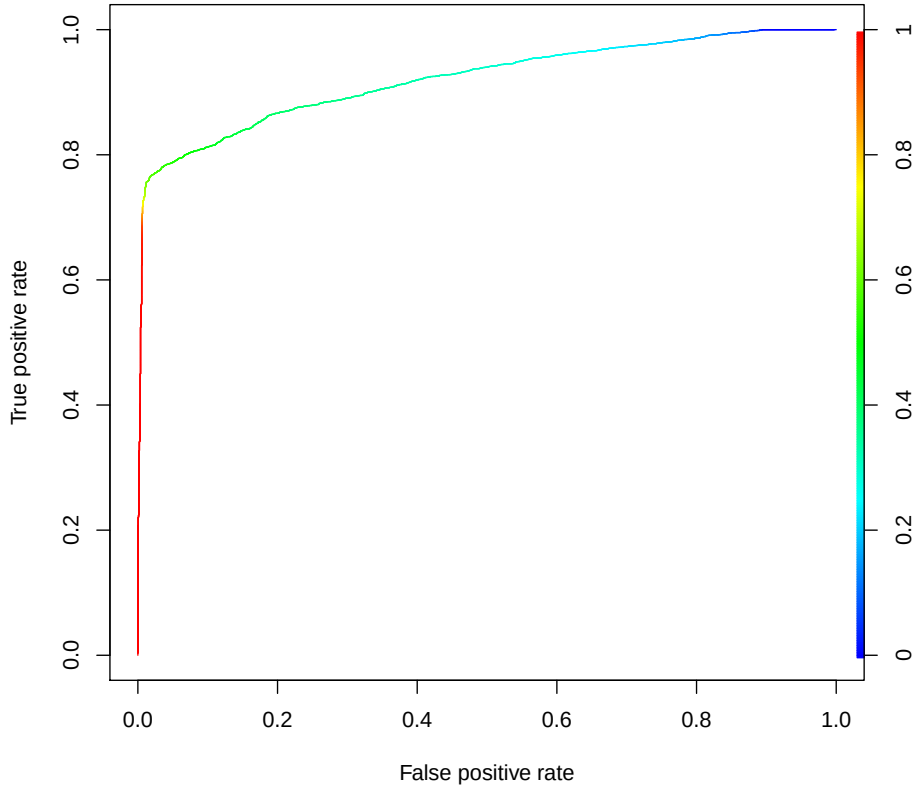
```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Call:
## glm(formula = Variant ~ Ac + Met + Type + Start + Stop + silenced +
##     noexprs, family = binomial(link = "logit"), data = train_df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -5.6138  -0.7408   0.0020   0.1488   2.7447
##
## Coefficients: (1 not defined because of singularities)
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)   6.643e-01  1.012e-01   6.566 5.17e-11 ***
## Ac           -1.340e-04  9.630e-06 -13.916  < 2e-16 ***
## Met           3.826e-04  3.141e-05  12.180  < 2e-16 ***
## Type5prima   -2.894e-01  1.016e-01  -2.847  0.00441 **
## TypeORF      -1.050e+00  8.838e-02 -11.880  < 2e-16 ***
## Typeother    -2.885e+01  2.288e+02  -0.126  0.89963
## Start        -2.339e-07  4.308e-08  -5.430 5.62e-08 ***
## Stop                 NA         NA      NA       NA
## silencedTRUE  4.138e+00  3.254e-01  12.715  < 2e-16 ***
## noexprsTRUE   1.642e+00  2.608e-01   6.296 3.05e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 11193.5  on 8485  degrees of freedom
## Residual deviance:  5498.4  on 8477  degrees of freedom
## AIC: 5516.4
##
## Number of Fisher Scoring iterations: 17
##
##         FALSE TRUE
##   FALSE  2964  252
##   TRUE   1060 4387
## [1] "Accuracy 0.848551310169687"

## Error in `[<-.data.frame`(`*tmp*`, "all_0.5", value = c(0, 1, 1,
0, 0, :  replacement has 8845 rows, data has 8663

## [1] "Accuracy of null model NaN"

## Loading required package:  gplots
##
## Attaching package:  'gplots'
```

```
## The following object is masked from 'package:stats':
##
##     lowess
```




```
##
## 3prima 5prima    ORF  other
##    140     89     21      2
##
## 3prima 5prima    ORF
##    171    213    676
##           llh       llhNull            G2      McFadden          r2ML
## -2749.2229441 -5596.7623637  5695.0788391     0.5087833     0.4888615
##          r2CU
##     0.6672848
```