Figure 1 Spec-seq data analysis

Zheng Zuo

04/15/2022

Contents

```
require(dplyr)
require(ggplot2)
require(TFCookbook)
```

Figure 1D

```
load("ZFP3.RData")
inner_join(Sample1.processed, Sample2.processed, by = "Sequence", suffix = c(".Sample1", ".Sample2" )) 
  ggplot(aes(x = `Relative Energy.Sample1`,
             y = `Relative Energy.Sample2`)) +
  geom_point()+
  theme_bw() +
  ggpubr::stat_regline_equation()+
  ggpubr::stat_cor(label.y = 1.5)
```
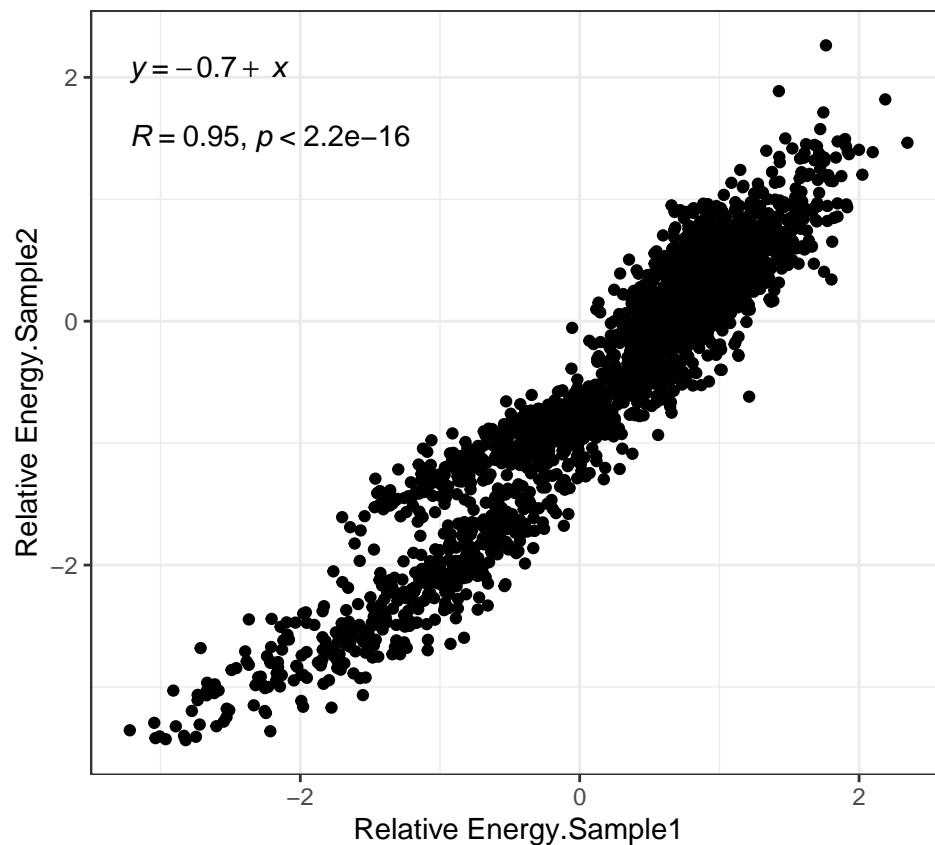
$y = -0.7 + x$

$R = 0.95, p < 2.2e-16$

Relative Energy.Sample2 (y-axis)

Relative Energy.Sample1 (x-axis)

Figure 1F, Build motif models based on all single variants of reference sequence

```
require(dplyr)
Sample1.processed %>%
  dplyr::filter(Mismatch<=1) %>%
  dplyr::rename(Energy=`Relative Energy`) %>%
  TFCookbook::buildEnergyModel() %>%
  TFCookbook::as.PEM() %>%
  TFCookbook::plotEnergyLogo() + ggtitle("Motif of ZFP3\nSample #1") + theme(plot.title = element_text(
```


Motif of ZFP3
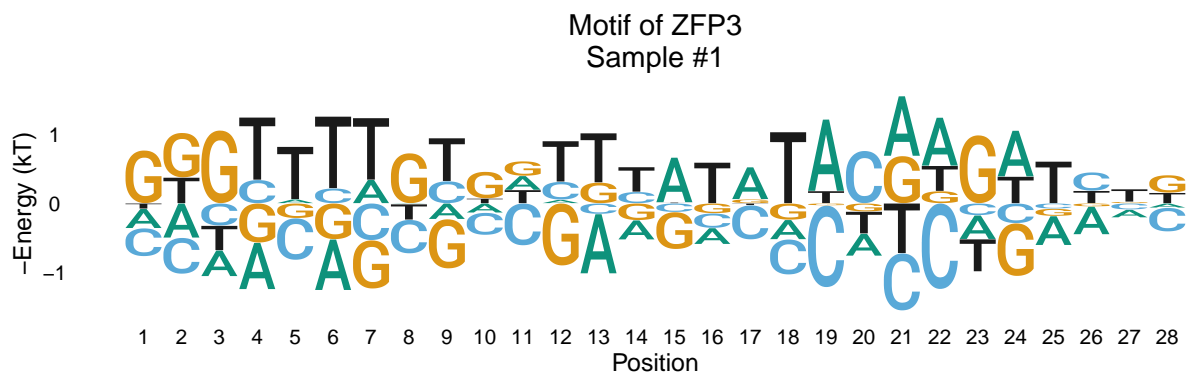Sample #1

−Energy (kT)

Position

Figure 1G

```
Sample1.processed %>%
  dplyr::filter(Mismatch<=1) %>%
  dplyr::rename(Energy=`Relative Energy`) %>%
  TFCookbook::buildEnergyModel() %>%
  as.PEM()
```

```
##           [,1]       [,2]       [,3]       [,4]        [,5]       [,6]
## A  0.28218279  0.4995361  0.3765766  0.6755694 -0.05900597  0.7377799
## C  0.40740973  0.5237523  0.3240545 -0.3402509  0.60580057 -0.2178135
## G -0.76848232 -0.6488556 -1.0583940  0.5711737  0.21559071  0.5222414
## T  0.07888979 -0.3744328  0.3577629 -0.9064922 -0.76238531 -1.0422078
##           [,7]       [,8]       [,9]      [,10]       [,11]       [,12]
## A -0.3412064  0.01195463  0.2294932  0.13616234 -0.2040753 -0.05232286
## C  0.5353914  0.48937425 -0.3217463  0.32286960  0.6111746 -0.26126490
## G  0.7028488 -0.72775861  0.7165971 -0.38868535 -0.2133016  0.90353174
## T -0.8970338  0.22642973 -0.6243440 -0.07034659 -0.1937977 -0.58994399
##          [,13]      [,14]       [,15]      [,16]       [,17]      [,18]
## A  0.8632288  0.2810915 -0.65123038  0.2340916 -0.46634964  0.2829009
## C  0.1548797 -0.1662228  0.12685814  0.2131535  0.49723933  0.5113659
## G -0.3023737  0.2453560  0.53813483  0.1448358 -0.05878225  0.2398132
## T -0.7157347 -0.3602247 -0.01376258 -0.5920809  0.02789256 -1.0340800
##           [,19]      [,20]      [,21]      [,22]      [,23]      [,24]
## A -1.03937560  0.3233298 -0.8689908 -0.7006027  0.3463539 -0.6680696
## C  1.18718473 -0.7618031  0.8316153  1.2447872  0.1737317  0.2886866
## G  0.03072265  0.1200273 -0.6868488 -0.1773274 -0.9885421  0.7670041
## T -0.17853179  0.3184459  0.7242243 -0.3668571  0.4684566 -0.3876211
##          [,25]      [,26]       [,27]       [,28]
## A  0.42301978  0.40465567  0.11525799  0.06402991
## C  0.07764811 -0.26867007  0.09077514  0.34569956
## G  0.10665218  0.04604486 -0.02213776 -0.25668683
## T -0.60732007 -0.18203045 -0.18389538 -0.15304264
```

Figure S1

```
Sample1.processed %>%
  dplyr::filter(Mismatch<=1) %>%
  dplyr::rename(Energy=`Relative Energy`) %>%
  TFCookbook::buildEnergyModel() -> ZFP3.Sample1.Model

Sample2.processed %>%
  dplyr::filter(Mismatch<=1) %>%
  dplyr::rename(Energy=`Relative Energy`) %>%
  TFCookbook::buildEnergyModel() -> ZFP3.Sample2.Model


ggplot(mapping = aes(x = ZFP3.Sample1.Model$coefficients[2:55],
          y = ZFP3.Sample2.Model$coefficients[2:55]))+
  geom_point() +
 # ggpubr::stat_regline_equation()+
```

```
ggpubr::stat_cor() +
theme_bw() +
xlab("Model parameters derived from Sample 1 data (kT)")+
ylab("Model parameters derived from Sample 2 data (kT)")
```
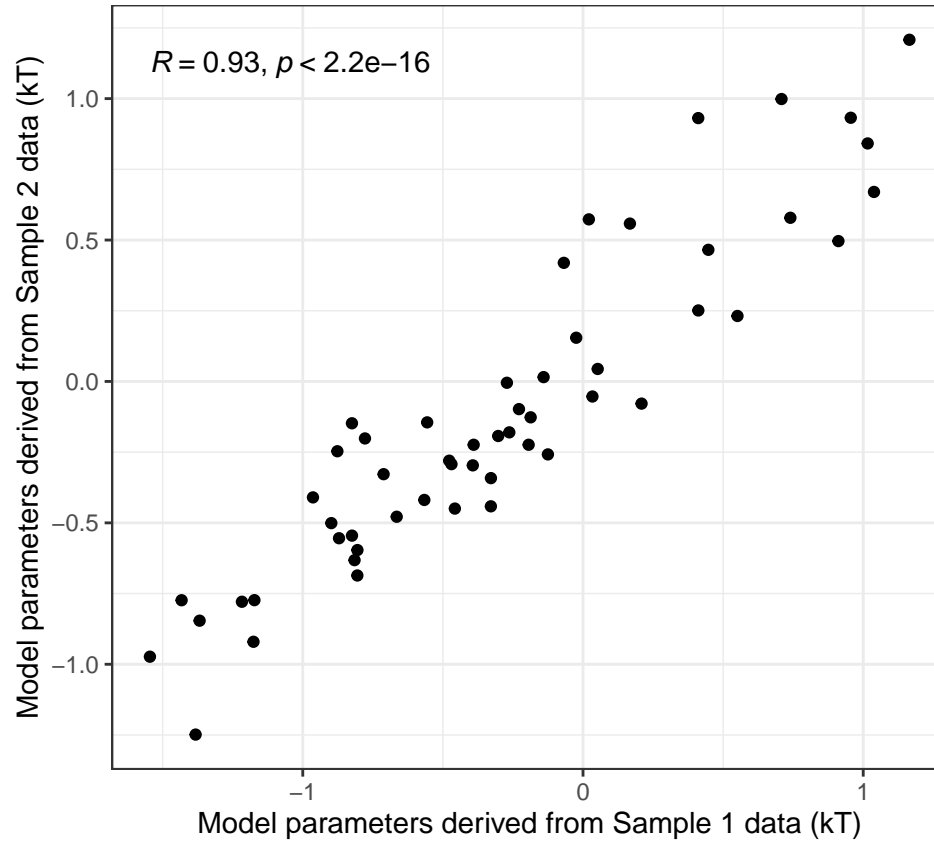


```
#ggsave("Models parameters comparison.svg", height = 4.5, width = 5)
```