Ambika HDAC inhibitor screen analysis

F. Heigwer

8/17/2018

Overview

This vignette is used to analyze a CRISPR screen performed by treating SW480 cells using 8 nM panobinostat in 0.1 % DMSO and 0.1 % DMSO as a control sample. It was screened in duplicates per treatment. Raw counts per sgRNA are the basis of the analysis.

All plots were used as given for the main figure panels. The compiled vignette pdf serves as an overview.

Data input

```
treat1<-read_delim(file = "../raw_data_counts/counts_panobinostat_screen/P1_S7_L001_R1_001.txt",delim =
treat2<-read_delim(file = "../raw_data_counts/counts_panobinostat_screen/P2_S8_L001_R1_001.txt",delim =

treatment<-"P"

system(paste("mkdir",treatment))

dmso1<-read_delim(file = "../raw_data_counts/counts_panobinostat_screen/DMSO1_S1_L001_R1_001.txt",delim
dmso2<-read_delim(file = "../raw_data_counts/counts_panobinostat_screen/DMSO2_S2_L001_R1_001.txt",delim

mydata<-treat1 %>%
  left_join(treat2,by="sgRNA") %>%
  left_join(dmso1,by="sgRNA") %>%
  left_join(dmso2,by="sgRNA") %>%
  extract(sgRNA,c("sgRNA","gene_id"),"((\\w+)_.*)")

names(mydata)<-c("sgRNA","gene_id","treat_1","treat_2","dmso_1","dmso_2")

mydata %<>%
  gather(treatment,count,-sgRNA,-gene_id) %>%
  extract(treatment,c("treatment","replicate"),"(\\w+)_(\\d)")

anno<-read_delim("../meta_data/gene_names.txt",delim="\t")
```

Lets normalize the data

```
mydata %<>%
  mutate(count=if_else(count>10,count,NA_real_)) %>%    #filter low read counts
  group_by(treatment,replicate) %>% #group by treatment and replicate
  mutate(normval=asinh(count/median(count,na.rm = T)))
```

Lets do some data QC

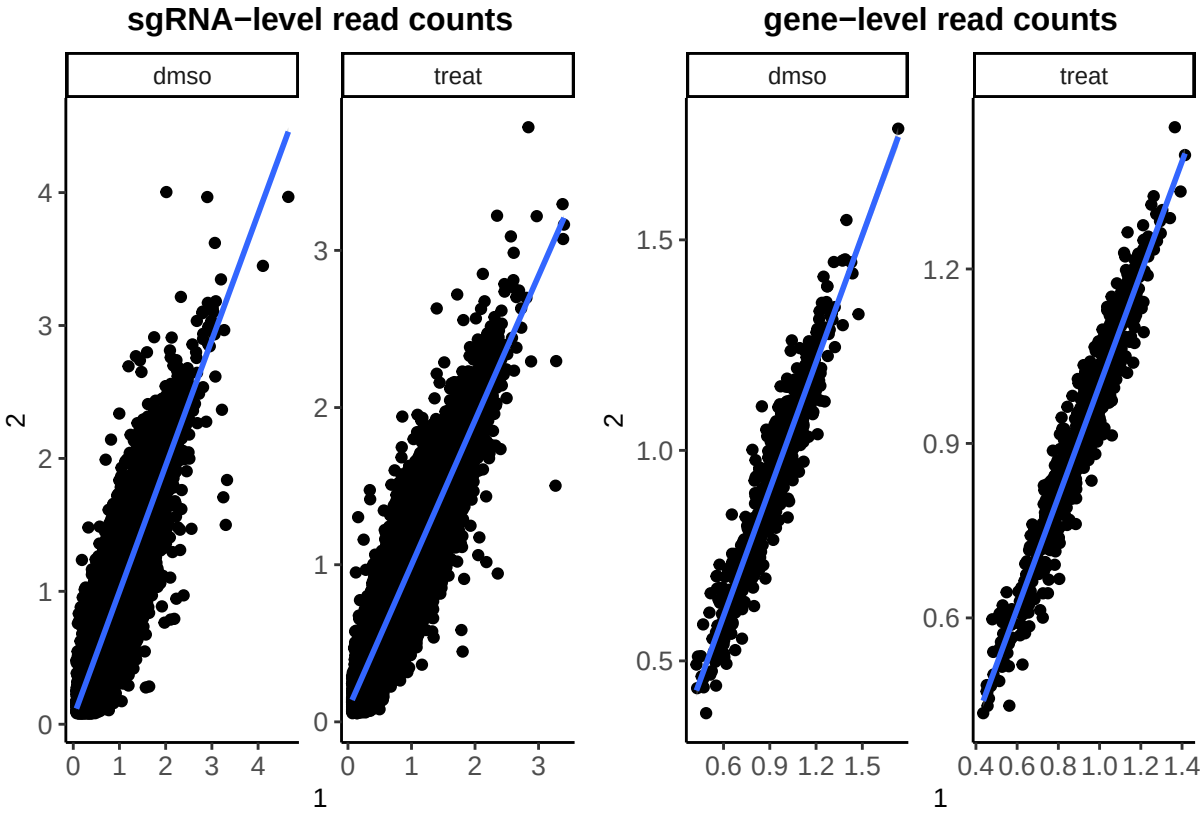
sgRNA-level replicate correlation

```
## $dmso
##   term   X1 X2
## 1    1
## 2    2  .88
##
## $treat
##   term   X1 X2
## 1    1
## 2    2  .91
```

gene-level replicate correlation

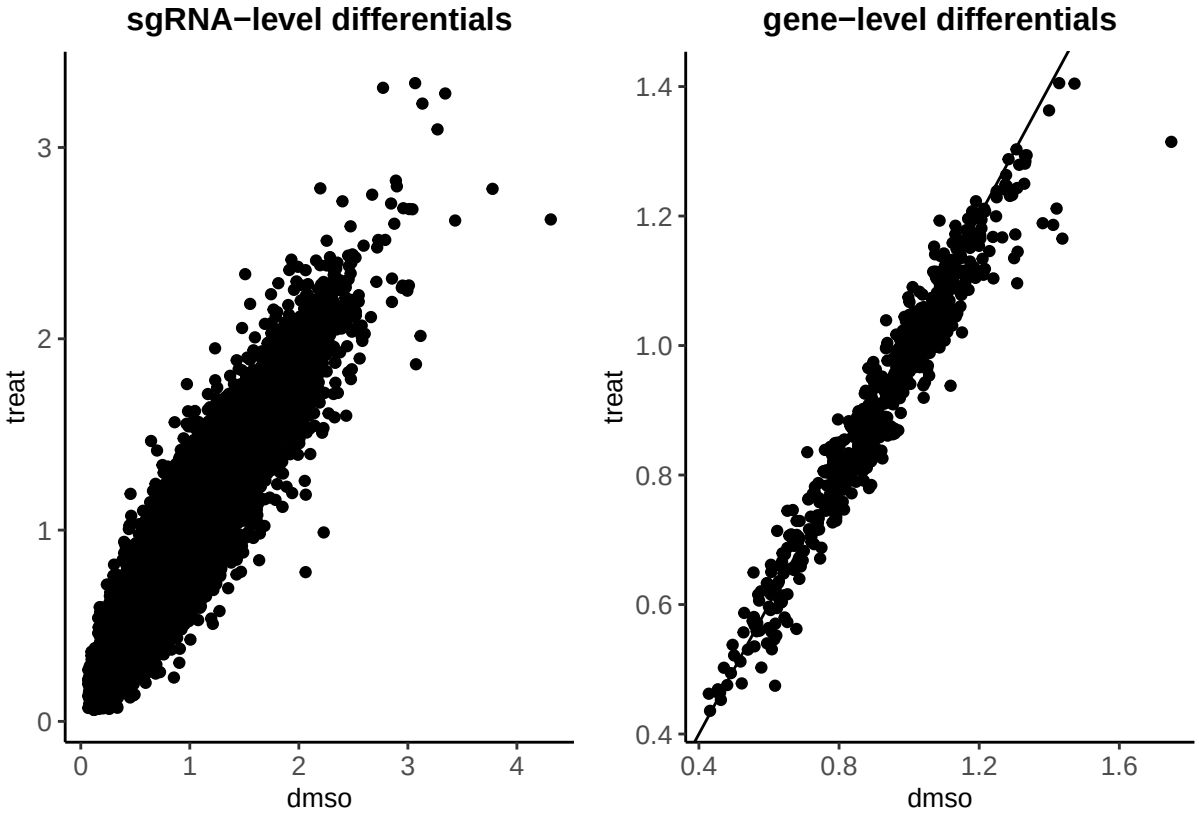
```
## $dmso
##   term   X1 X2
## 1    1
## 2    2  .94
##
## $treat
##   term   X1 X2
## 1    1
## 2    2  .97
```

graphics



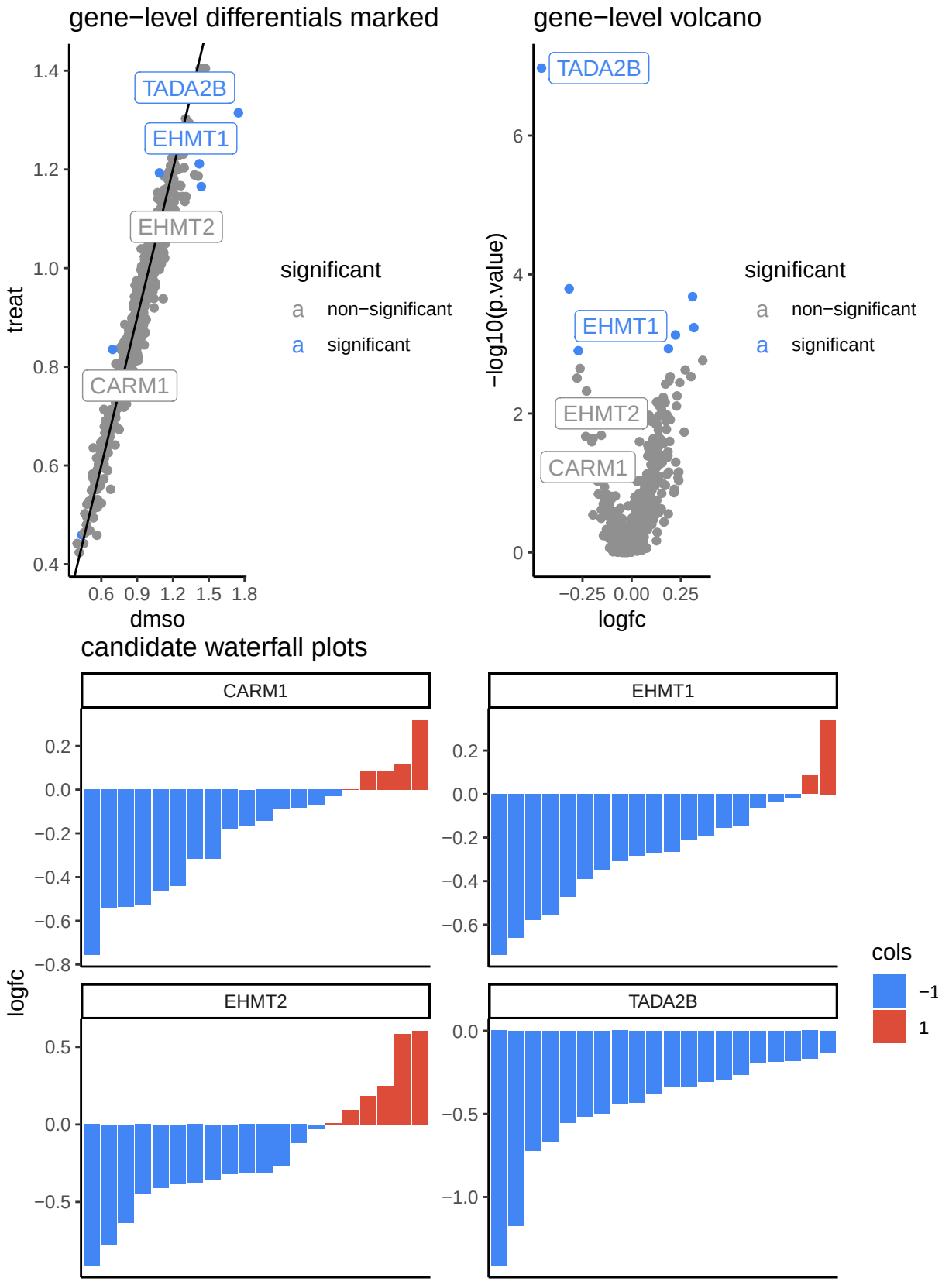
QC looks very nice and good so lets calls some foldchanges and hits thereof

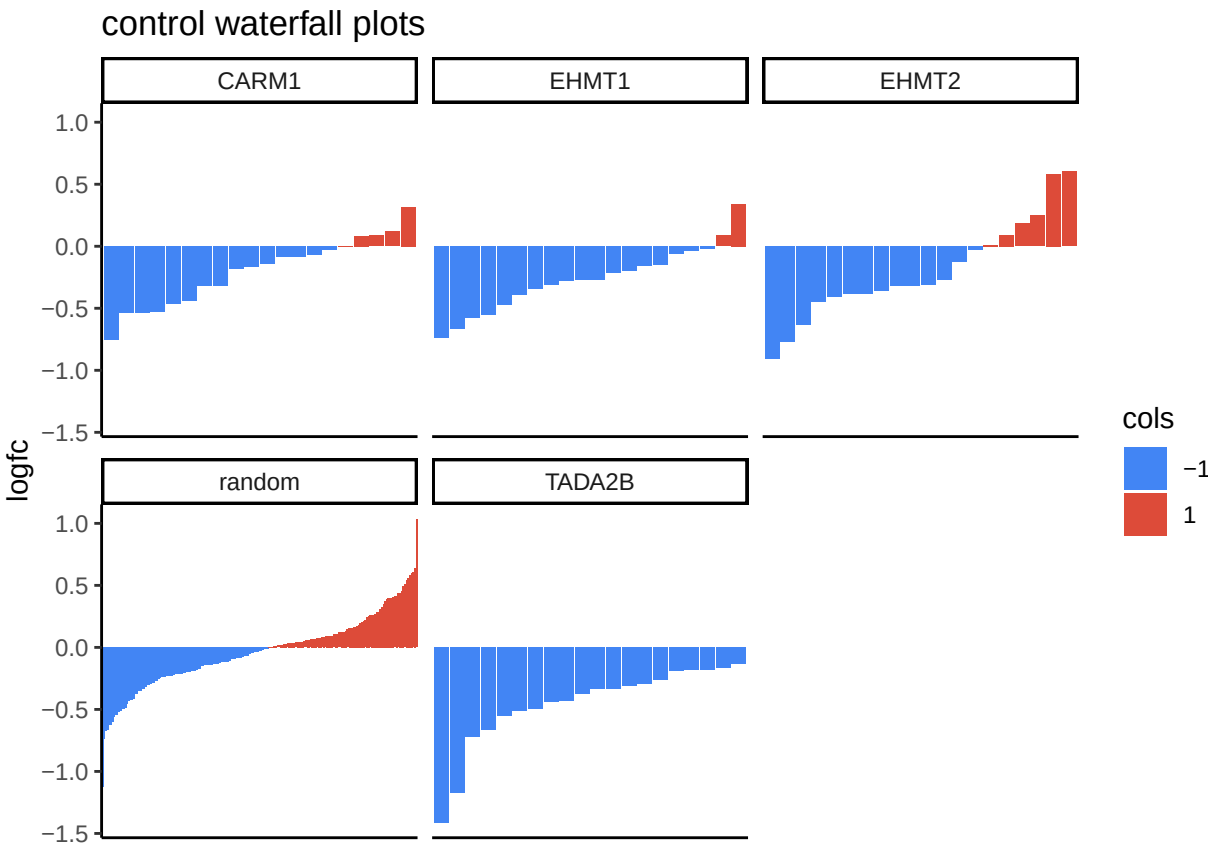
Hit-calling



Statistical analysis

Here we use pairwise wilcoxon rank-sum test between to fold changes of the individual sgRNAs for each gene and the individual negative control sgRNA to determine if the foldchange of the gene represented by independent sgRNAs is significantly different from the non-targeting controls.


gene−level differentials marked
TADA2B
EHMT1
EHMT2
CARM1
treat
dmso
1.4
1.2
1.0
0.8
0.6
0.4
0.6 0.9 1.2 1.5 1.8
significant
a non−significant
a significant
gene−level volcano
TADA2B
EHMT1
EHMT2
CARM1
−log10(p.value)
logfc
6
4
2
0
−0.25 0.00 0.25
significant
a non−significant
a significant
candidate waterfall plots
CARM1
EHMT1
EHMT2
TADA2B
logfc
0.2
0.0
−0.2
−0.4
−0.6
−0.8
0.5
−0.5
−1.0
cols
−1
1



Session info