

```
#####Figure 7C Average Norm Counts at MRB7260 EPS#####
library(dplyr)
#install.packages("tidyr")
library(tidyr)
#install.packages("ggplot2")
library(ggplot2)
#install.packages("reshape2")
library(reshape2)

#open the clean table organized by KD and induction (this will include 2 MRB7260 induced & 10 uninduced samples)
RPS12allC1 <- read.csv(file.choose(), header = TRUE, sep = ",")

#Need to average all ten uninduced samples and remove 2913 samples that are in the file
AvgUninducedRPS12 <- RPS12allC1 %>% filter(KD != "29-13" & TET == "FALSE") %>% group_by(edit_stop, gene) %>%
  summarize(Avg = sum(norm_count)/10) %>% mutate(KD = "AvgUn")

#If want to separate out only PhyH KD first and rename to MRB7260
AvgInducedRPS12 <- RPS12allC1 %>% filter(KD == "PhyH" & TET == "TRUE") %>% group_by(edit_stop, gene) %>%
  summarize(Avg = sum(norm_count)/2) %>% mutate(KD = "MRB7260")

#Combine data - if want to look at 29-13 you can also add WT factor
TotaldataRPS12 <- bind_rows(AvgUninducedRPS12, AvgInducedRPS12)

#define which sites you want to look at
EPS7260 <- c(25, 29, 30, 39, 45, 46, 48, 53, 58, 59, 60, 61, 71, 72, 78)

TotaldataRPS12b <- TotaldataRPS12 %>% filter(edit_stop %in% EPS7260)

allbar <- ggplot(TotaldataRPS12b, aes(x=KD, y=Avg)) + geom_bar(stat="identity", position = "stack") +
  facet_wrap(~edit_stop, scales = "free") + scale_fill_manual(values=c("black"), name="Junction Length",
labels=c("0", "1-10", "11-50", "50+")) +
  theme(panel.margin = unit(1, "lines"))

#Final graph
allbar

#####Figure 8 Top Junction Sequences gRNA 2#####
library(dplyr)
#install.packages("tidyr")
library(tidyr)
#install.packages("ggplot2")
library(ggplot2)
#install.packages("reshape2")
library(reshape2)

# open the clean table organized by KD and induction (this will include 2 MRB7260 induced & 10 uninduced samples)
RPS12allC1 <- read.csv(file.choose(), header = TRUE, sep = ",")

#to look at the second gRNA - edit_stop<40, edit_stop>22)
RPS12allg2 <- RPS12allC1 %>% filter(edit_stop<40, edit_stop>22, KD!="29-13", !(edit_stop==9 & junc_len==0))

#Average all the 10 uninduced samples
RPS12allglun <- RPS12allg2 %>% filter(TET==FALSE) %>% group_by(edit_stop, junc_len, junc_seq) %>%
  mutate(AvgNm = sum(norm_count)/10) %>% select(edit_stop, junc_seq, junc_len, TET, AvgNm) %>%
  mutate(KD="AvgUn") %>% distinct(., .keep_all = TRUE)
glimpse(RPS12allglun)
head(RPS12allglun)

# make an equivalent table with the top junction sequences
# found in the KDs, taking the average of the norm count across replicates

RPS12allglin <- RPS12allg2 %>% filter(TET==TRUE & KD=="PhyH") %>% group_by(KD, edit_stop, junc_len, junc_seq) %>%
  mutate(AvgNm = sum(norm_count)/2) %>% select(edit_stop, junc_seq, junc_len, TET, AvgNm, KD) %>%
  distinct(., .keep_all = TRUE)
head(RPS12allglin)

# combine the tables together to make graphing easier

#Contains all average junction sequences for Unind and KD
RPS12allglav <- bind_rows(RPS12allglun, RPS12allglin) %>% group_by(KD) %>%
  mutate(total = sum(AvgNm)) %>% rowwise() %>% mutate(perc = 100*(AvgNm/total))
View(RPS12allglav)

# put the top sequences by KD in a chart compare side by side (>100 avg copies)
RPS12g1top100 <- RPS12allglav %>% filter(AvgNm > 100) %>% arrange(KD, desc(AvgNm))
glimpse(RPS12g1top100)
View(RPS12g1top100)
summary(RPS12g1top100)

write.table(RPS12g1top100, "RPS12gRNA2topSeq.csv", sep = ",")

#####Figure S3A Total Junction Lengths of RPS12#####
library(dplyr)
```

```

#install.packages("tidyr")
library(tidyr)
#install.packages("ggplot2")
library(ggplot2)
#install.packages("reshape2")
library(reshape2)

# open the clean table organized by KD and induction (this will include 2 MRB7260 induced & 10 uninduced samples)
RPS12allC1 <- read.csv(file.choose(), header = TRUE, sep = ",")

# introduce junction length bins
RPS12allC1$bin1 <- cut(RPS12allC1$junc_len, breaks = c(-Inf,1,11,51,Inf), right = FALSE)
levels <- levels(RPS12allC1$bin1)

#Need to average all ten uninduced samples and remove 2913 samples that are in the file
AvgUninduced <- RPS12allC1 %>% filter(KD != "29-13" & TET == "FALSE") %>% group_by(edit_stop, bin1) %>%
  summarize(Avg = sum(norm_count)/10) %>% mutate(KD = "AvgUn")

#Need to include only the MRB7260 induced samples
#If want to separate out PhyH first and rename to MRB7260
AvgInduced <- RPS12allC1 %>% filter(KD == "PhyH"& TET == "TRUE") %>% group_by(edit_stop, bin1) %>%
  summarize(Avg = sum(norm_count)/2) %>% mutate(KD = "MRB7260")

#Combine samples into one table
Totaldata <- bind_rows(AvgUninduced,AvgInduced)

## Remove junction length 0 from first ES which is actually pre-edited
## sum all junction bins across whole population
Totaldata2 <- Totaldata %>% filter(!(edit_stop==9 & bin1=="[-Inf,1)"))
Totalbin <- Totaldata2 %>% group_by(KD, bin1) %>% summarize(popsum = sum(Avg))
glimpse(Totalbin)

TotalJL <- ggplot() + ylab("Norm Count") + ggtitle("RPS12 Total Junction Length") +
  geom_bar(data=Totalbin, aes(x = KD, y = popsum, fill=bin1), stat="identity", position="fill") +
  scale_fill_manual(values=c("black", "royalblue1", "yellow", "green 4"),name="Junction Length",
    labels=c("0","1-10","11-50","50+"))

#Final graph
TotalJL

#####Figure S3B Junction Lengths at each MRB7260 EPS for RPS12#####
library(dplyr)
#install.packages("tidyr")
library(tidyr)
#install.packages("ggplot2")
library(ggplot2)
#install.packages("reshape2")
library(reshape2)

#open the clean table organized by KD and induction (this will include 2 MRB7260 induced & 10 uninduced samples)
RPS12allC1 <- read.csv(file.choose(), header = TRUE, sep = ",")

#introduce junction length bins
RPS12allC1$bin1 <- cut(RPS12allC1$junc_len, breaks = c(-Inf,1,11,51,Inf), right = FALSE)
levels <- levels(RPS12allC1$bin1)

#For MRB7260 n=10 here for uninduced
AvgUninduced <- RPS12allC1 %>% filter(KD != "29-13" & TET == "FALSE") %>% group_by(edit_stop, bin1) %>%
  summarize(Avg = sum(norm_count)/10) %>% mutate(KD = "AvgUn")

#If want to separate out PhyH first and rename to MRB7260
AvgInduced <- RPS12allC1 %>% filter(KD == "PhyH"& TET == "TRUE") %>% group_by(edit_stop, bin1) %>%
  summarize(Avg = sum(norm_count)/2) %>% mutate(KD = "MRB7260")

#Combine data
Totaldata <- bind_rows(AvgUninduced,AvgInduced)

#If want to use the same set of data and look at specfic editing sites for that knockdown
#Code to show separate graphs for each site

#define which sites you want to look at
EPS7260 <- c(25,29,30,39,45,46,48,53,58,59,60,61,71,72,78)

Totaldata3 <- Totaldata %>% filter(edit_stop%in%EPS7260)

#bars filled = values set to one
allbar2 <- ggplot(Totaldata3, aes(x=KD, y=Avg, fill = bin1)) + geom_bar(stat="identity", position = "fill") +
  facet_wrap(~edit_stop, nrow=2, scales = "free") + scale_fill_manual(values=c("black", "royalblue2", "yellow",
"green 4"), name="Junction Length", labels=c("0","1-10","11-50","50+")) +
  theme(panel.margin = unit(1, "lines"))

#Final graph
allbar2

```

```
#####Figure S3C Junction Lengths at each ESS for RPS12#####
library(dplyr)
#install.packages("tidyr")
library(tidyr)
#install.packages("ggplot2")
library(ggplot2)
#install.packages("reshape2")
library(reshape2)

#open the clean table organized by KD and induction (this will include 2 MRB7260 induced & 10 uninduced samples)
RPS12allC1 <- read.csv(file.choose(), header = TRUE, sep = ",")

#introduce junction length bins
RPS12allC1$bin1 <- cut(RPS12allC1$junc_len, breaks = c(-Inf,1,11,51,Inf), right = FALSE)
levels <- levels(RPS12allC1$bin1)

#For MRB7260 n=10 here for uninduced
AvgUninduced <- RPS12allC1 %>% filter(KD != "29-13" & TET == "FALSE") %>% group_by(edit_stop, bin1) %>%
  summarize(Avg = sum(norm_count)/10) %>% mutate(KD = "AvgUn")

#If want to separate out PhyH first and rename to MRB7260
AvgInduced <- RPS12allC1 %>% filter(KD == "PhyH" & TET == "TRUE") %>% group_by(edit_stop, bin1) %>%
  summarize(Avg = sum(norm_count)/2) %>% mutate(KD = "MRB7260")

#Combine data
Totaldata <- bind_rows(AvgUninduced, AvgInduced)

#these are for defining the EPS to put little dots over the graph
#Tell which sites I want to look at within gRNA 2-4
MRB7260EPSb <- data.frame(edit_stop=c(25,29,30,39,45,46,48,53,58,59,60,61), Avg=1.05, KD="MRB7260")

#Narrow in on gRNA 2-4 (region that editing stops mostly)
Totaldatab <- Totaldata %>% filter(edit_stop<65, edit_stop>20)

#Figure of all sites with KD and induced
MRB7260JLc <- ggplot() +
  ylab("Percentage") + xlab("Editing Site") + ggtitle("RPS12 Junction Lengths by Editing Site") +
  geom_bar(data=Totaldatab, aes(x = edit_stop, y = Avg, fill=bin1), stat="identity", position="fill") +
  scale_fill_manual(values=c("black", "royalblue1", "yellow", "green 4"), name="Junction Length", labels=c("0", "1-10", "11-50", "50+")) +
  scale_x_reverse(breaks=c(65,60,55,50,45,40,35,30,25,20)) +
  facet_grid(KD~.) +
  geom_point(data=MRB7260EPSb, aes(x = edit_stop, y = Avg), pch=21)

#Final figure
MRB7260JLc
```