

Supplementary Fig.1 f

RAC

22/09/2022

```
#script used
```

```
#for f in 2_QC_mapped/CBP*DMSO*SoUmiDedupRemSec.bam ; do OUT=$(basename $f | sed 's/SoUmiDedupRemSec.
```

above script was used to extract insert size from mapped bamfile. Insert sizes above 1000nt were removed, as these probably were due to spliced reads. Data below shows box plots of insert size from CBP20 1-2 and CBP80 1-3

```
library(dplyr)
```

```
##
```

```
## 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
```

```
library(tidyr)
```

```
library(ggplot2)
```

```
df<-read.csv("../data/readLengthCBPlongextra.txt", sep = " ", header = F)
```

```
colnames(df)<-c("sample","length")
```

```
fig<-
```

```
df %>%
```

```
  separate(sample, c("protein","rep", "timepoint")) %>%
```

```
  filter(!(protein == "CBP20" & rep == "3")) %>%
```

```
  ggplot(aes(x=rep, y=length, fill = protein)) +
```

```
  geom_boxplot(position = position_dodge2(preserve = "single"), outlier.shape = NA, width = 0.8) +
```

```
  geom_point(stat="summary", position = position_dodge2(width = 0.8), size = 1) +
```

```
  coord_cartesian(ylim = c(0,120)) +
```

```
  labs(subtitle = "CBP - DMSO library insert size analysis",
```

```
        caption = "under 1000nt to remove reads mapped over splice junctions") +
```

```
  xlab("Replicate") +
```

```
  ylab("Read length")
```

```
## Warning: Expected 3 pieces. Additional pieces discarded in 2384648 rows [1, 2,
```

```
## 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].
```

```
fig +
```

```
  theme_bw() +
```

```

theme(text = element_text(size=10)) +
xlab("") +
ylab("") +
labs(subtitle = "",
      caption = "") +
facet_grid(. ~ timepoint)

```

No summary function supplied, defaulting to `mean_se()`

