

Figure 5 E-F and Sup Fig 5 D

RAC

25/08/2020

```
#load data
BP_anno<-read.csv("../data/BP-annotated-metadata.hg38.100ntupdown.binN201.bed", header = F, sep = "\t")

colnames(BP_anno)<-c("chr", "start", "stop", "geneID", "DistToLandmark", "strand")

BP_anno_wrangled<-
  BP_anno %>%
  separate(geneID, c("BP_Comp_Score", "CLIPtags", "coordBPsite", "BP_distTo3ss", "polyYTrack", "percentageY"))

#count the number of introns in the annoation
BP_annotation_number<-
BP_anno_wrangled %>%
  filter(DistToLandmark == "0") %>%
  group_by(BP_distTo3ss) %>%
  summarise(BP_number = n())

#load count files

xiCLIP_BP_5primepos_counts<-read.csv("../data/xiCLIP.5primepos.BP-annotated-metadata_hg38_100ntupdown.binN201.bed", header = F, sep = "\t")
xiCLIP_BP_3endOfRead2_counts<-read.csv("../data/xiCLIP.3endOfRead2.BP-annotated-metadata_hg38_100ntupdown.binN201.bed", header = F, sep = "\t")

xiCLIP_BP_counts<-rbind(xiCLIP_BP_5primepos_counts, xiCLIP_BP_3endOfRead2_counts)

colnames(xiCLIP_BP_counts)<-c("Sample", "chr", "start", "stop", "geneID", "DistToLandmark", "strand", "count")

#wrapgle count file
xiCLIP_BP_counts_wrangled<-
  xiCLIP_BP_counts %>%
  separate(Sample, into=c("Sample", "Rep", "Timepoint", "ReadType")) %>%
  filter(Timepoint != "negative") %>%
  separate(geneID, c("BP_Comp_Score", "CLIPtags", "coordBPsite", "BP_distTo3ss", "polyYTrack", "percentageY")) %>%
  filter(!(Sample == "CBP20" & Rep == "3")) %>%
  filter(Sample == "RBM7")

#wrapgle data
xiCLIP_BP_counts_wrangled_calculated<-
xiCLIP_BP_counts_wrangled %>%
  #group together bin(DistToLandmark and the landmark itself to calculates sum of bins.)
  group_by(Sample, Rep, Timepoint, DistToLandmark, ReadType, BP_distTo3ss) %>%
```

```

summarise(binSum=sum(count)) %>%
ungroup() %>%
spread(DistToLandmark,binSum, fill = 0) %>%
gather("DistToLandmark","binSum", c(`-100`:`100`)) %>%
#normalise to the number of annotations
left_join(BP_annotation_number) %>%
mutate(normSum = case_when(
  binSum > 0 ~ binSum/BP_number,
  TRUE ~ 0))

## `summarise()` has grouped output by 'Sample', 'Rep', 'Timepoint',
## 'DistToLandmark', 'ReadType'. You can override using the `.groups` argument.
## Joining, by = "BP_distTo3ss"

for_plot <-

xiCLIP_BP_counts_wrangled_calculated %>%
mutate(DistToLandmark_n = as.numeric(as.character(DistToLandmark)),
  Timepoint = factor(Timepoint, levels=c("negative", "PBSDRB", "t00","t05", "t10", "t15", "t20",
  ReadType = factor(ReadType, c("5primepos","3endOfRead2"))) %>%
mutate(Timepoint_f = case_when(
  Timepoint == "PBSDRB" ~ "t00",
  TRUE ~ as.character(Timepoint)
)) %>%
mutate(Timepoint_f = factor(Timepoint_f, levels=c("negative","t00","t05", "t10", "t15", "t20", "t40",
filter(BP_distTo3ss %in% c(18:38) & ReadType %in% c("5primepos","3endOfRead2"))) %>%
mutate(ReadType = gsub("5primepos","cross-link", gsub("3endOfRead2","3'CLIP", ReadType))) %>%
mutate(ReadType = factor(ReadType, levels = c("cross-link", "3'CLIP"))) %>%
filter(Sample == "RBM7" ) %>%
group_by(Sample,Rep,Timepoint_f,ReadType) %>%
mutate(norm_to_max = normSum/max(normSum)) %>%
ungroup() %>%
mutate(new_x = as.numeric(DistToLandmark_n) - as.numeric(BP_distTo3ss)) %>%
filter(Timepoint_f %in% c("t00","t40","t60","DMSO"))

```

#Figure 5 E & F #graph shows the aggregated cross-link and 3'CLIP positions around branchpoints. Y axis shows gene groups based on the BP distance to the downstream 3'SS.

```

plot<-
for_plot %>%
ggplot(aes(x=DistToLandmark_n, y=reorder(BP_distTo3ss, desc(BP_distTo3ss)))) +
geom_raster(aes(fill = norm_to_max)) +
scale_fill_distiller(palette = "Spectral",
  direction = -1) +
  #limits=c(0,0.2)) +
facet_grid(Timepoint_f~ReadType, scales="free_y") +
coord_cartesian(xlim = c(-50,50) ) +
scale_x_continuous(breaks = c(-50,-25,0,+18,+38,+50),
  labels = c("-50","-25","BP","+18","+38","+50")) +
scale_y_discrete(breaks = c(18,28,38),
  labels = c("+18","+28","+38")) +
xlab("Distance to landmark (nt)") +
ylab("Distance from BP to downstream 3'SS") +
theme_bw()+
theme(text = element_text(size = 8),

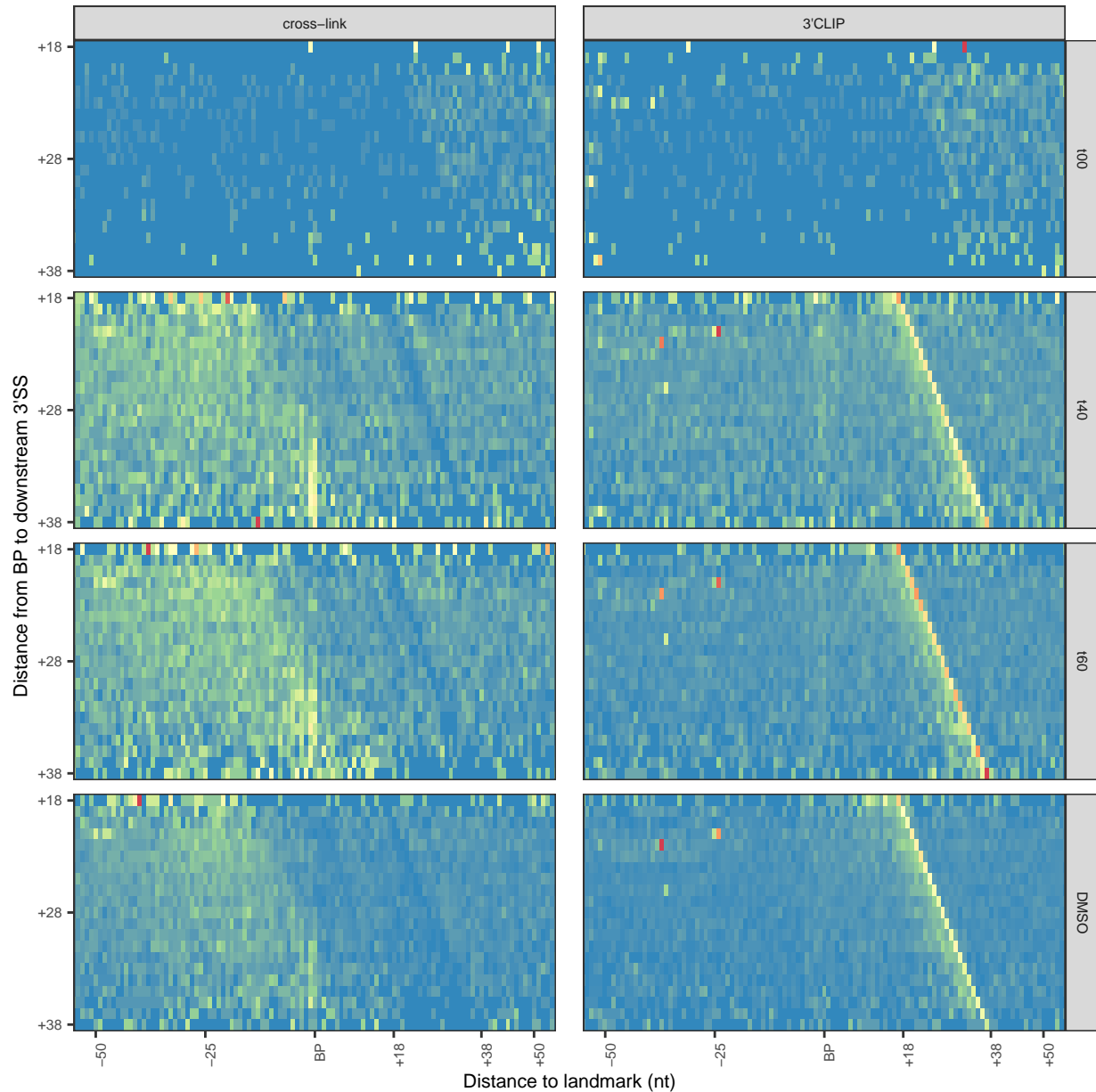
```

```

legend.position = "none",
axis.text.x = element_text(angle = 90, hjust = 1),
panel.spacing.y = unit(0.4, "lines"),
panel.spacing.x = unit(0.8, "lines"))

print(plot)

```



#supplementary Figure 5 D #graph shows the aggregated cross-link sites around branch points

```
#xiCLIP_BP_counts_wrangled_calculated_graph %>%
```

```

xiCLIP_BP_counts_wrangled_calculated_graph<-
  xiCLIP_BP_counts_wrangled_calculated %>%
  mutate(DistToLandmark_n = as.numeric(as.character(DistToLandmark)),
         Timepoint = factor(Timepoint, levels=c("negative", "PBSDRB", "t00","t05", "t10", "t15", "t20",

```

```

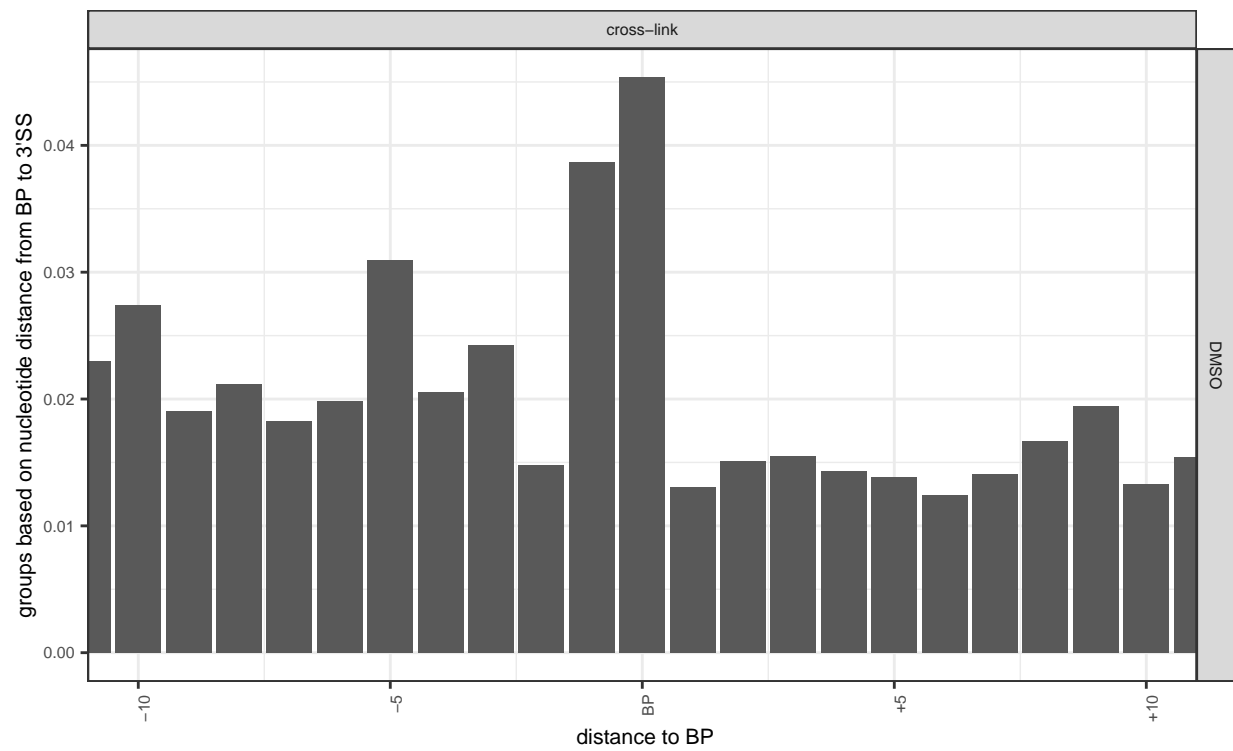
      ReadType = factor(ReadType, c("5primepos", "3endOfRead2")) %>%
mutate(Timepoint_f = case_when(
  Timepoint == "PBSDRB" ~ "t00",
  TRUE ~ as.character(Timepoint)
)) %>%
mutate(Timepoint_f = factor(Timepoint_f, levels=c("negative", "t00", "t05", "t10", "t15", "t20", "t40",

xiCLIP_BP_counts_wrangled_calculated_graph %>%
  filter(BP_distTo3ss %in% c(18:38) & ReadType %in% c("5primepos") & Timepoint == "DMS0") %>%
  mutate(ReadType = gsub("5primepos", "cross-link", gsub("3endOfRead2", "3'CLIP", ReadType))) %>%
  mutate(ReadType = factor(ReadType, levels = c("cross-link", "3'CLIP"))) %>%
  filter(Sample == 'RBM7' ) %>%
  group_by(Sample, Rep, Timepoint_f, ReadType) %>%
  mutate(norm_to_max = normSum/max(normSum)) %>%
  ungroup() %>%
  ggplot(aes(x=DistToLandmark_n, y=normSum) ) +
  geom_bar(stat="summary") +
  facet_grid(Timepoint_f~ReadType, scales="free_y") +
  coord_cartesian(xlim = c(-10,10) ) +
  scale_x_continuous(breaks = c(-10,-5,0,+5,+10),
                     labels = c("-10", "-5", "BP", "+5", "+10") ) +

  xlab(" distance to BP") +
  ylab("groups based on nucleotide distance from BP to 3'SS") +
  theme_bw()+
  theme(text = element_text(size = 8),
        legend.position = "right",
        axis.text.x = element_text(angle = 90, hjust = 1),
        panel.spacing.y = unit(0.4, "lines"),
        panel.spacing.x = unit(0.8, "lines"))

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

```



```
#ggsave("figs/FigureS5.210521.pdf", height = 2, width = 2)
```