

# Plot haplotyped methylation from long reads

*Xuanken Tay*

```
library(tidyverse)
library(data.table)
library(gridExtra)
```

## Load and prepare data

```
loadRData <- function(infile) {
  # loads an RData file and returns it
  # assume that the variable saved in same name as the file
  load(infile)
  get(ls()[ls() == gsub(basename(infile), pattern=".RData$",
                        replacement="")])
}

mouse_per_read_stats_path <-
  "/wehisn/bioinf/lab_speed/txk/methylation_data/mouse/mouse_per_read_stats_scaled_llr.RData"
cpg_path <-
  "/wehisn/bioinf/lab_speed/txk/cpg_data/GRCm38/CG_GRCm38.tsv"

mouse_per_read_stats <- loadRData(mouse_per_read_stats_path) %>%
  dplyr::rename(start=pos) %>%
  mutate(end=start+1) %>%
  as.data.table()

cpg <- read_tsv(cpg_path, col_types="cii", col_names=FALSE)
names(cpg) <- c("chr", "start", "end")
cpg <- cpg %>%
  mutate(chr=sub("^chr", "", chr))
```

## Extract b6xcast and castxb6

```
mouse_per_read_stats <- mouse_per_read_stats %>%
  mutate(is_b6xcast=grepl('b6_maternal|cast_paternal', mouse_per_read_stats$strain),
         strain=sub("b6_|cast_", "", strain))
b6xcast <- mouse_per_read_stats %>%
  filter(is_b6xcast)
castxb6 <- mouse_per_read_stats %>%
  filter(!is_b6xcast)

rm(mouse_per_read_stats)

head(b6xcast, 3)
```

```
##   chr   start log_lik_ratio          read_id   strain
## 1    1 3003226      -1.484 367e2c44-c70e-4827-ae22-5dfd1508afac maternal
## 2    1 3003339       0.913 367e2c44-c70e-4827-ae22-5dfd1508afac maternal
```

```
## 3 1 3003379 -3.535 367e2c44-c70e-4827-ae22-5dfd1508afac maternal
##      beta      end is_b6xcast
## 1 0.815176 3003227      TRUE
## 2 0.286386 3003340      TRUE
## 3 0.971667 3003380      TRUE
```

```
head(castxb6, 3)
```

```
## chr start log_lik_ratio read_id strain
## 1 1 3003226 -1.600 b1d8e98c-cad5-4f49-93ac-3089bb727c2c maternal
## 2 1 3003226 -0.099 999a6f0c-d211-4e66-9eec-6edd9250f496 maternal
## 3 1 3003339 4.039 b1d8e98c-cad5-4f49-93ac-3089bb727c2c maternal
##      beta      end is_b6xcast
## 1 0.832018 3003227      FALSE
## 2 0.524730 3003227      FALSE
## 3 0.017310 3003340      FALSE
```

## Functions to plot methylation level per-read

```
plotMethylation <- function(df, cpG, chrM, start, end, overhang=20000,
                             title="Title") {
  ## Spaghetti plot by Scott Gigante
  region_start <- start - overhang
  region_end <- end + overhang
  df <- df %>%
    filter(chr==chrM, start>region_start, end<region_end)

  p <- ggplot(df) +
    geom_smooth(method="loess",
               aes(x=start, y=beta, group=read_id, color=strain),
               span=2, se=FALSE) +
    ggplot2::ylim(0, 1) +
    geom_point(data=cpG %>% filter(chr==chrM,
                                   start > region_start,
                                   end < region_end),
              aes(x=start), y=0, pch='|') +
    geom_ribbon(data=data_frame(x=c(start, end)),
              aes(x), fill='red', color='transparent',
              alpha=0.2, ymin=0, ymax=1) +

    scale_color_manual(values=c('maternal'='#F8766D',
                                'paternal'='#00BFC4')) +

    theme_bw() +
    ggtitle(title) +
    xlab(paste0("chr ", chrM, " pos"))
  p
}

plotReadMethylation <- function(df, chrM, start, end, overhang=2000,
                                title="Title") {
  region_start <- start - overhang
  region_end <- end + overhang
  df <- df %>%
```

```

    filter(chr==chr, start>region_start, end<region_end)
p <- ggplot(df, aes(x=start, group=read_id, y=1, color=beta)) +
  geom_line(position=ggstance::position_dodgev(height=0.2), size=3) +
  scale_color_gradient2(low="blue", mid="white", high="red", midpoint=0.5, space="Lab") +
  geom_vline(xintercept=c(start, end), linetype="dashed") +
  facet_grid(rows=vars(strain)) +
  theme_bw() +
  theme(axis.text.y=element_blank(),
        axis.ticks.y=element_blank()) +
  labs(x=paste0("chr ", chr, " pos"), y="read", title=title)
p
}

```

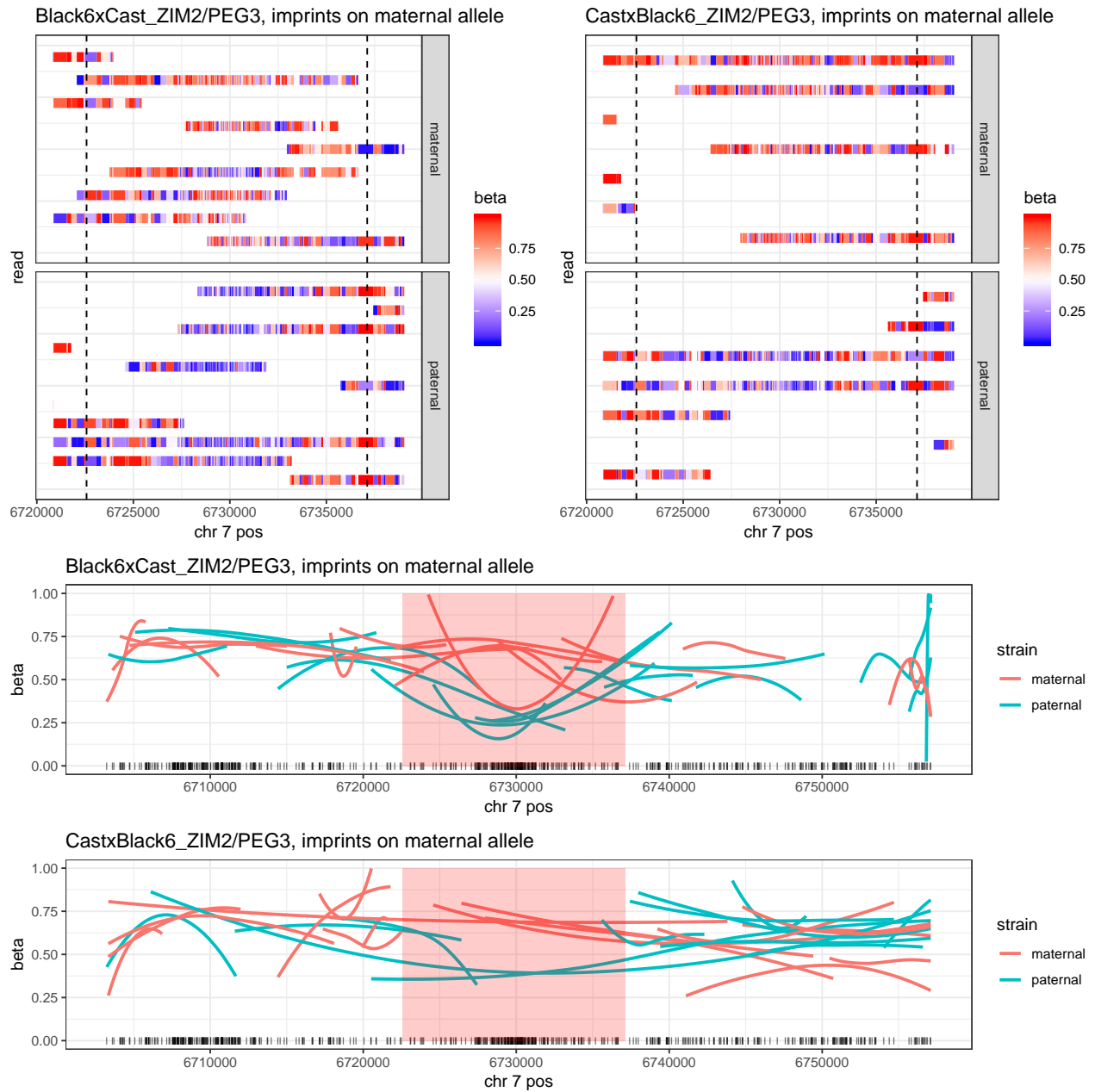
## Actual plotting on some known ICRs

```

layout <- rbind(c(1, 2), c(1, 2), c(3, 3), c(4, 4))

peg3 <- c(7, 6722576, 6737116)
p1 <- plotMethylation(b6xcast, cpg, peg3[1], peg3[2], peg3[3],
  title=paste0("Black6xCast_ZIM2/PEG3, imprints on maternal allele"))
p2 <- plotMethylation(castxb6, cpg, peg3[1], peg3[2], peg3[3],
  title=paste0("CastxBlack6_ZIM2/PEG3, imprints on maternal allele"))
pr1 <- plotReadMethylation(b6xcast, peg3[1], peg3[2], peg3[3],
  title=paste0("Black6xCast_ZIM2/PEG3, imprints on maternal allele"))
pr2 <- plotReadMethylation(castxb6, peg3[1], peg3[2], peg3[3],
  title=paste0("CastxBlack6_ZIM2/PEG3, imprints on maternal allele"))
grid.arrange(pr1, pr2, p1, p2, layout_matrix=layout)

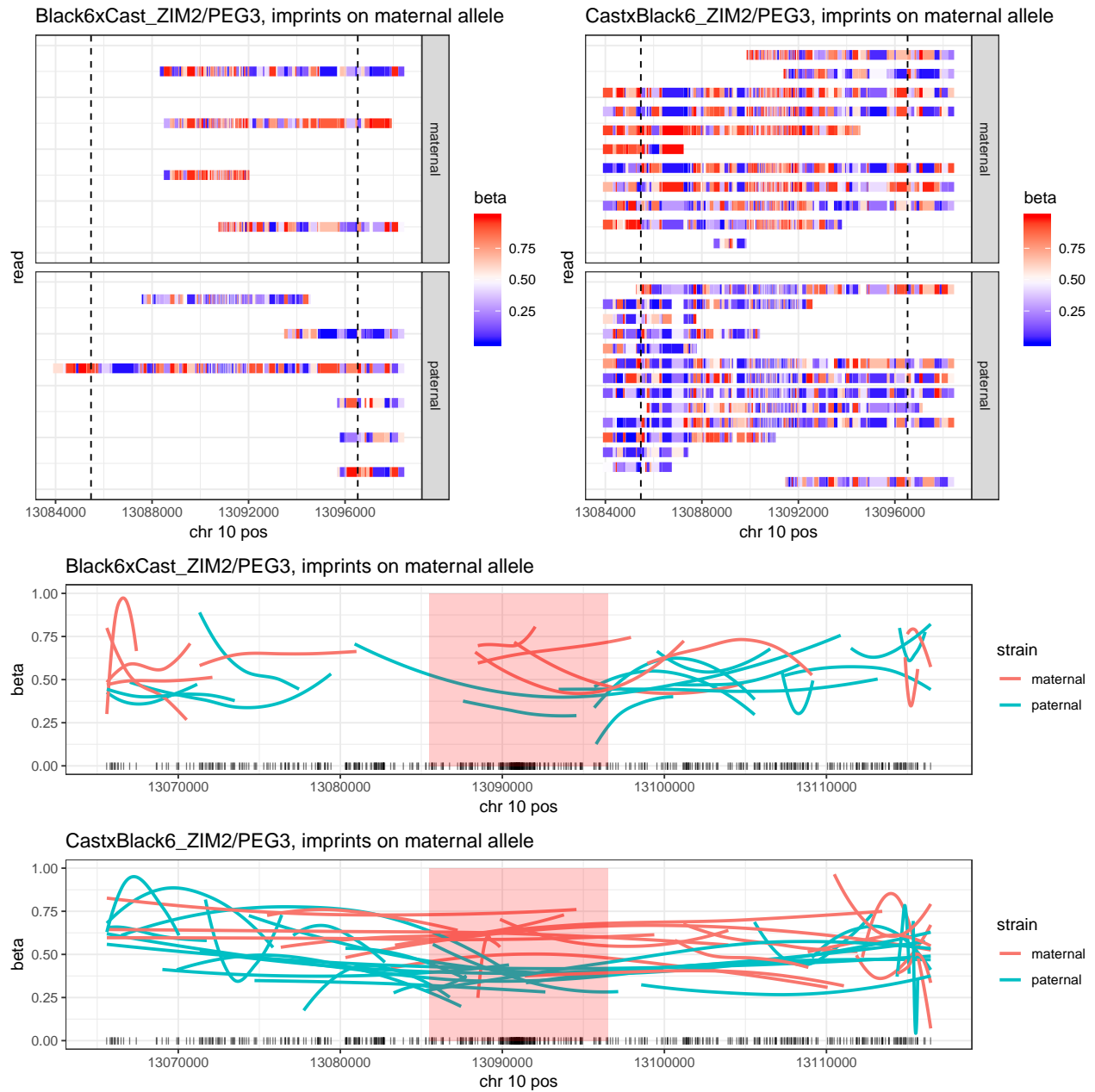
```



```

zac1 <- c(10, 13085470, 13096527)
p1 <- plotMethylation(b6xcast, cpg, zac1[1], zac1[2], zac1[3],
  title=paste0("Black6xCast_ZIM2/PEG3, imprints on maternal allele"))
p2 <- plotMethylation(castxb6, cpg, zac1[1], zac1[2], zac1[3],
  title=paste0("CastxBlack6_ZIM2/PEG3, imprints on maternal allele"))
pr1 <- plotReadMethylation(b6xcast, zac1[1], zac1[2], zac1[3],
  title=paste0("Black6xCast_ZIM2/PEG3, imprints on maternal allele"))
pr2 <- plotReadMethylation(castxb6, zac1[1], zac1[2], zac1[3],
  title=paste0("CastxBlack6_ZIM2/PEG3, imprints on maternal allele"))
grid.arrange(pr1, pr2, p1, p2, layout_matrix=layout)

```

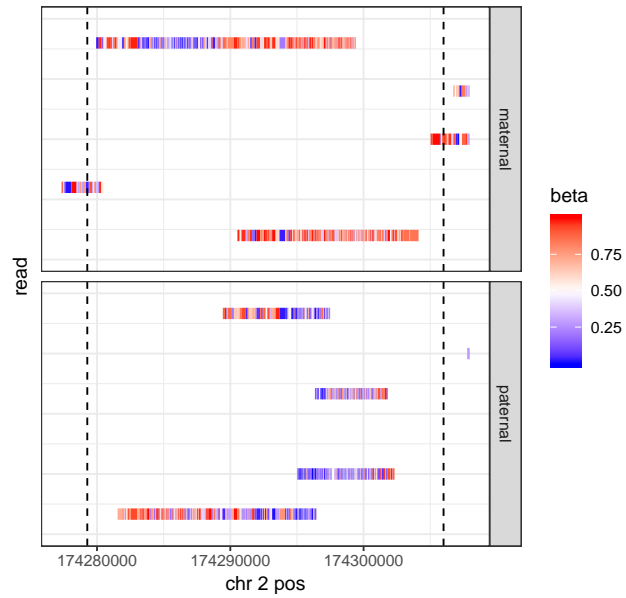


```

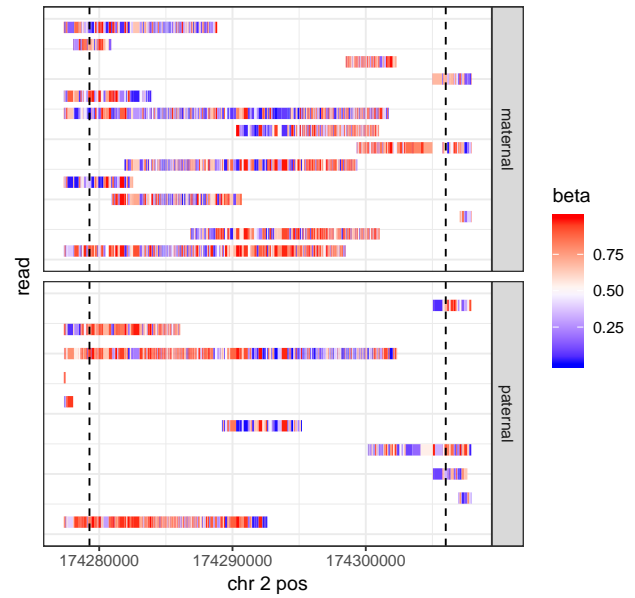
nespas <- c(2, 174279269, 174306000)
p1 <- plotMethylation(b6xcast, cpg, nespas[1], nespas[2], nespas[3],
  title=paste0("Black6xCast_ZIM2/PEG3, imprints side by side"))
p2 <- plotMethylation(castxb6, cpg, nespas[1], nespas[2], nespas[3],
  title=paste0("CastxBlack6_ZIM2/PEG3, imprints side by side"))
pr1 <- plotReadMethylation(b6xcast, nespas[1], nespas[2], nespas[3],
  title=paste0("Black6xCast_ZIM2/PEG3, imprints side by side"))
pr2 <- plotReadMethylation(castxb6, nespas[1], nespas[2], nespas[3],
  title=paste0("CastxBlack6_ZIM2/PEG3, imprints side by side"))
grid.arrange(pr1, pr2, p1, p2, layout_matrix=layout)

```

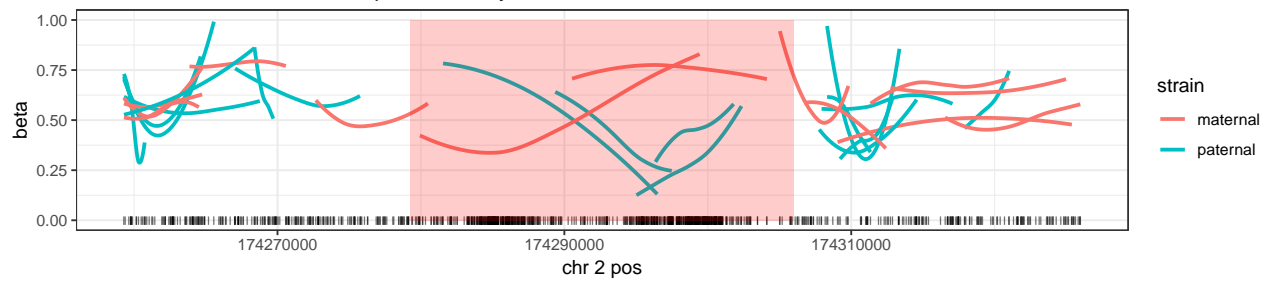
Black6xCast\_ZIM2/PEG3, imprints side by side



CastxBlack6\_ZIM2/PEG3, imprints side by side



Black6xCast\_ZIM2/PEG3, imprints side by side



CastxBlack6\_ZIM2/PEG3, imprints side by side

