z scores

John Stansfield

December 4, 2017

set up

```r
library(readr)
library(data.table)
library(HiCcompare)
library(dplyr)
library(pROC)
library(ggplot2)
```

```r
# Z scores for M, Diff and distance weighting
.calc_zscores <- function(hic.table) {
  # calculate z scores
  Zm <- (hic.table$adj.M - mean(hic.table$adj.M)) / sd(hic.table$adj.M)
  hic.table[, raw_diff := adj.IF2 - adj.IF1]
  Zd <- (hic.table$raw_diff - mean(hic.table$raw_diff)) / sd(hic.table$raw_diff)
  Zmean <- (Zm + Zd) / 2
  hic.table[, ':=' (Zm = Zm, Zd = Zd, Zmean = Zmean)]
  # calculate distance weighting
  dist_weight <- 1 - ((hic.table$D + 1)/max(hic.table$D + 1))
  hic.table[, D_wt := dist_weight]
  hic.table[, Zwt := Zmean * D_wt]
  hic.table[, p.val := 2*pnorm(abs(Zwt), lower.tail = FALSE)]
  hic.table[, p.adj := p.adjust(p.val, method = 'fdr')]
  MD.plot2(hic.table$adj.M, hic.table$D, hic.table$p.adj)
  MD.plot2(hic.table$adj.M, hic.table$D, hic.table$p.val)
}


# Z scores for M, Diff with NO distance weighting
.calc_zscores2 <- function(hic.table) {
  # calculate z scores
  Zm <- (hic.table$adj.M - mean(hic.table$adj.M)) / sd(hic.table$adj.M)
  hic.table[, raw_diff := adj.IF2 - adj.IF1]
  Zd <- (hic.table$raw_diff - mean(hic.table$raw_diff)) / sd(hic.table$raw_diff)
  Zmean <- (Zm + Zd) / 2
  hic.table[, ':=' (Zm = Zm, Zd = Zd, Zmean = Zmean)]
  # calculate distance weighting
  # dist_weight <- 1 - ((hic.table$D + 1)/max(hic.table$D + 1))
  # hic.table[, D_wt := dist_weight]
  # hic.table[, Zwt := Zmean * D_wt]
  hic.table[, p.val := 2*pnorm(abs(Zwt), lower.tail = FALSE)]
  hic.table[, p.adj := p.adjust(p.val, method = 'fdr')]
  MD.plot2(hic.table$adj.M, hic.table$D, hic.table$p.adj)
  MD.plot2(hic.table$adj.M, hic.table$D, hic.table$p.val)
}
```

```r
# Z scores for M, Diff and distance weighting calculated by Distance
.calc_zscores3 <- function(hic.table) {
  hic.table[, raw_diff := adj.IF2 - adj.IF1]
  # split table up for each distance
  temp_list <- S4Vectors::split(hic.table, hic.table$D)
  # combined top 15% of distances into single data.table
  all_dist <- sort(unique(hic.table$D))
  dist_85 <- ceiling(0.85 * length(all_dist))
  temp_list2 <- temp_list[1:dist_85]
  temp_list2[[dist_85+1]] <- data.table::rbindlist(temp_list[(dist_85+1):length(temp_list)])
  temp_list <- temp_list2
  rm("temp_list2")
  # z score by distance
  temp_list <- lapply(temp_list, function(x) {
    Zm <- (x$adj.M - mean(x$adj.M)) / sd(x$adj.M)
    Zd <- (x$raw_diff - mean(x$raw_diff)) / sd(x$raw_diff)
    Zmean <- (Zm + Zd) / 2
    x[, ':=' (Zm = Zm, Zd = Zd, Zmean = Zmean)]
    return(x)
  })
  # recombine into one table
  hic.table <- rbindlist(temp_list)
  # calculate distance weighting
  dist_weight <- 1 - ((hic.table$D + 1)/max(hic.table$D + 1))
  hic.table[, D_wt := dist_weight]
  hic.table[, Zwt := Zmean * D_wt]
  hic.table[, p.val := 2*pnorm(abs(Zwt), lower.tail = FALSE)]
  hic.table[, p.adj := p.adjust(p.val, method = 'fdr')]
  MD.plot2(hic.table$adj.M, hic.table$D, hic.table$p.adj)
  MD.plot2(hic.table$adj.M, hic.table$D, hic.table$p.val)
}
```

Z scores for M, Diff and distance weighting

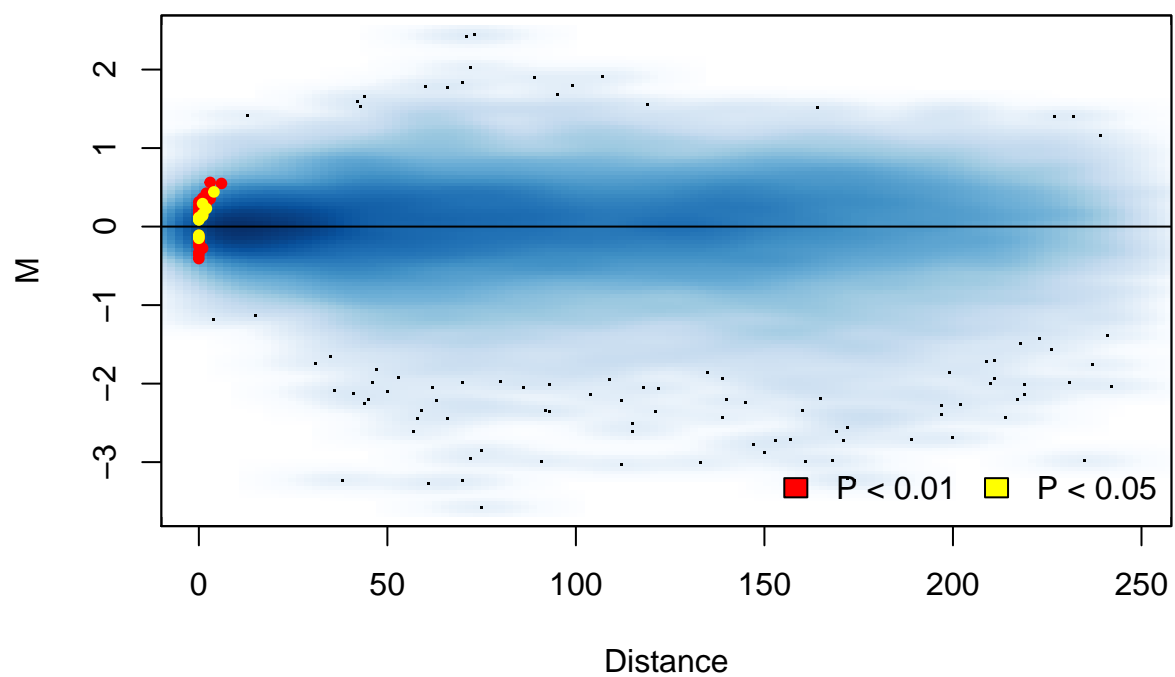First plot is FDR adjusted p-values. Second plot is raw p-values.
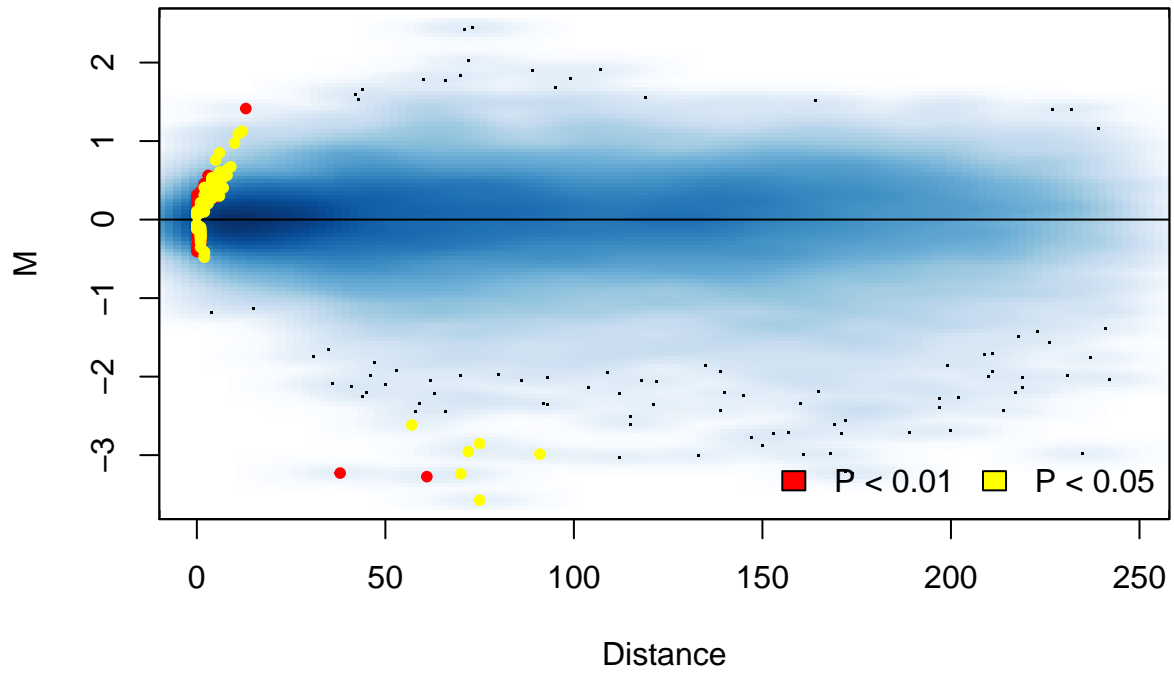
```r
hic.table <- amyg_dplfc1[[1]]
```

```r
.calc_zscores(hic.table)
```
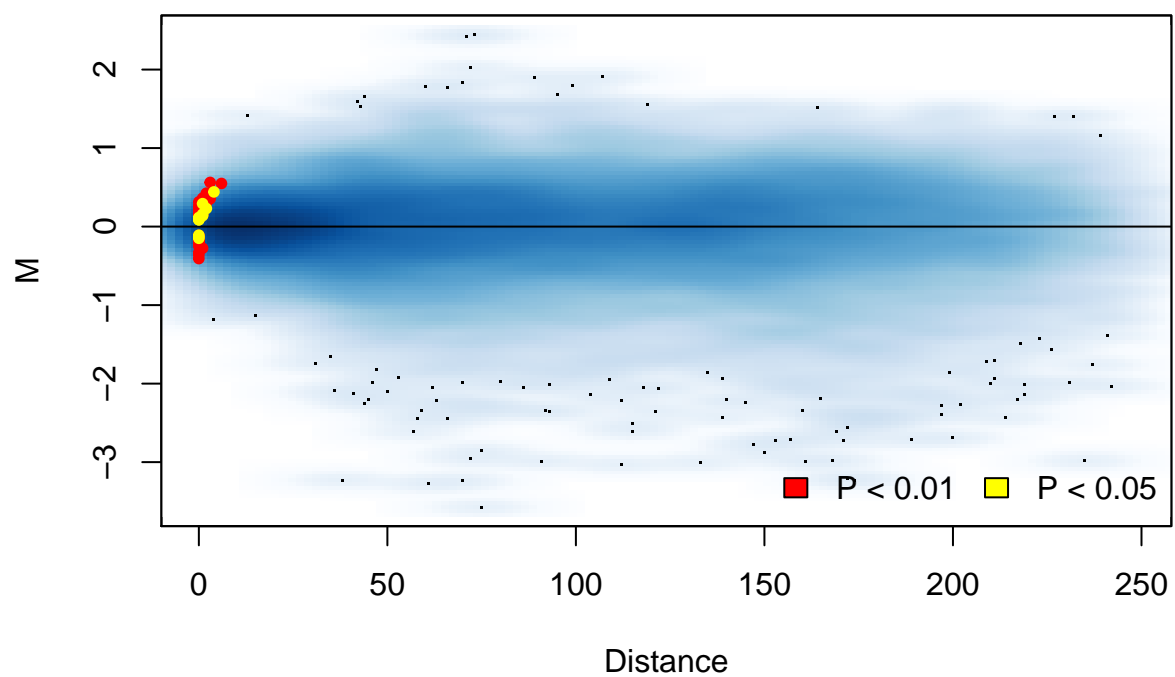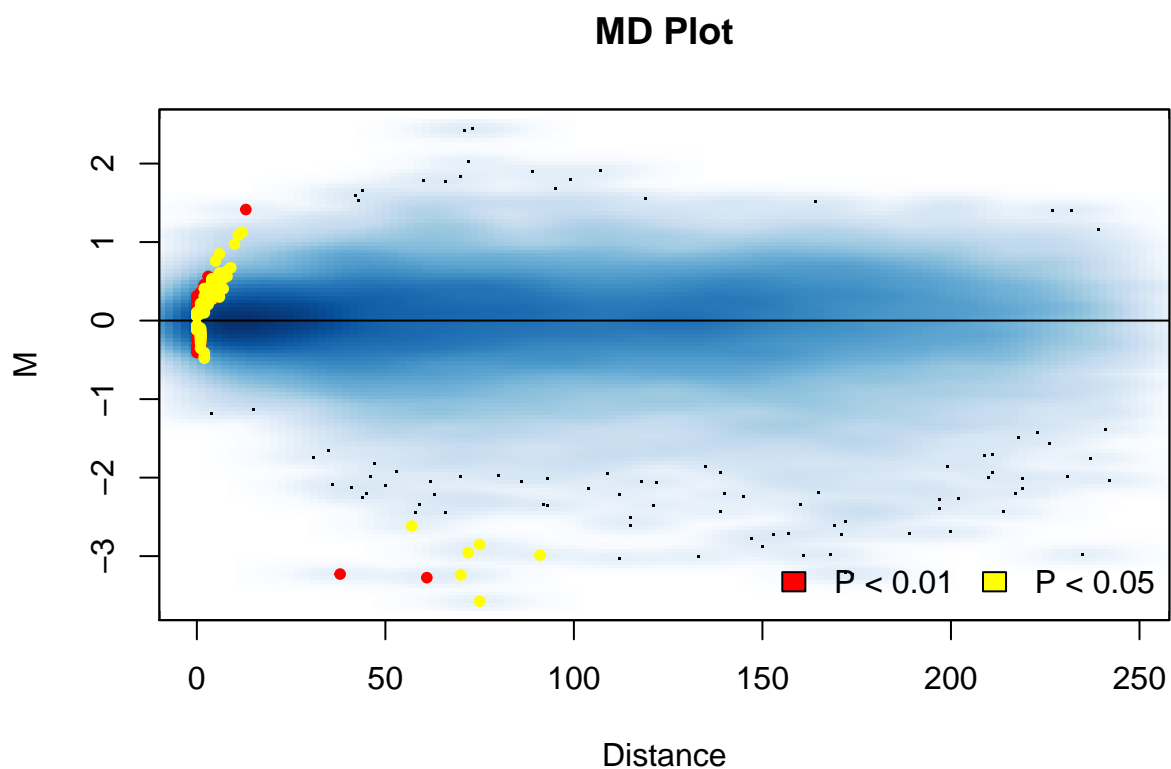
MD Plot

MD Plot

Z scores for M, Diff with NO distance weighting

First plot is FDR adjusted p-values. Second plot is raw p-values.

```
.calc_zscores2(hic.table)
```
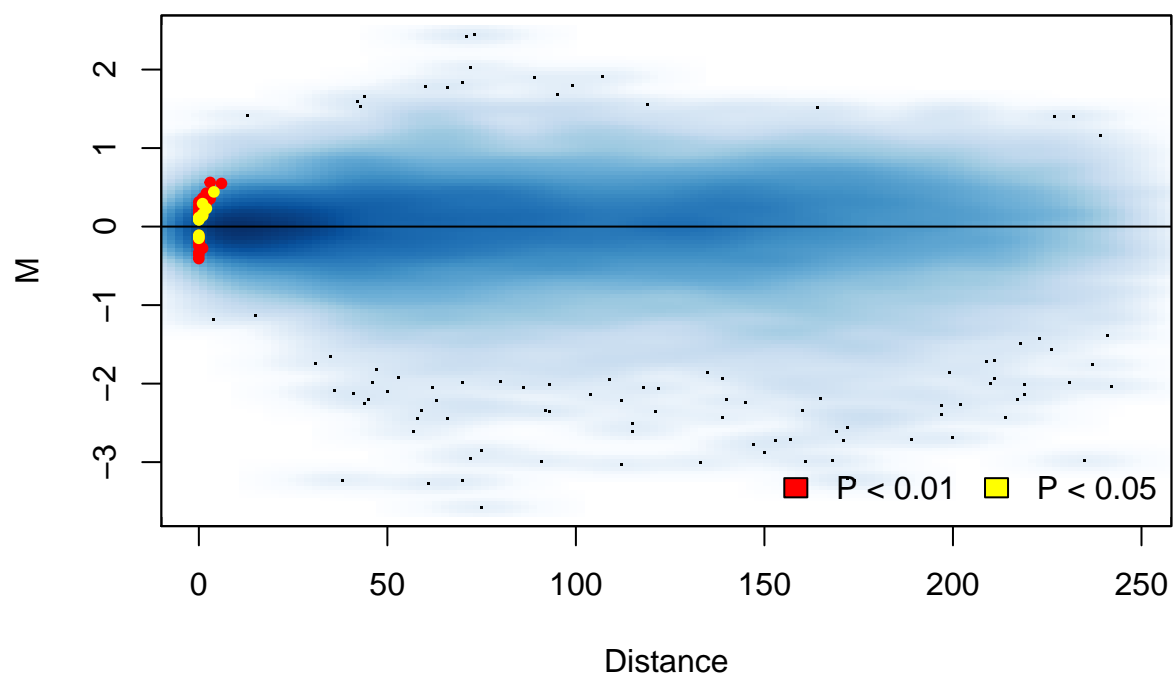
MD Plot

MD Plot

Z scores for M, Diff and distance weighting calculated by Distance

First plot is FDR adjusted p-values. Second plot is raw p-values.

```
hic.table <- amyg_dplfc1[[1]]
.calc_zscores3(hic.table)
```

MD Plot

MD Plot