aNon-target analysis in R (R138)
Hyun Yong Jin (hyjin@scripps.edu)
20151209

```{r}
#Gene list has been generated

setwd("~/TheShell/SeqResults/R138_Jin_RiPr")
library(data.table)
library(ggplot2)
cov <- fread("coverage_matrix_small")
setkey(cov, symbol) #setkey based on gene symbol


# Import genelist, from separate csv files
genelist <-
read.table("06_visualization/Coverage_Nontarget_forloop_20151209/genelist_nontarget_noago.csv",
header=T, sep=",")

nontarget_down_in_tg <- as.vector(genelist$nontarget_down_in_tg)
nontarget_down_in_tg   <-nontarget_down_in_tg[nontarget_down_in_tg !="" ]
noago_down_in_tg <- as.vector(genelist$noago_down_in_tg)
noago_down_in_tg   <-noago_down_in_tg[noago_down_in_tg !="" ]
nontarget_up_in_tko <- as.vector(genelist$nontarget_up_in_tko)
nontarget_up_in_tko   <-nontarget_up_in_tko[nontarget_up_in_tko !="" ]
noago_up_in_tko <- as.vector(genelist$noago_up_in_tko)
noago_up_in_tko   <-noago_up_in_tko[noago_up_in_tko !="" ]

```


#WT-TG comparison first
```{r}

#WT -->notarget_down_in_tg
wt.start <-matrix(0, nrow=150, ncol=0)
for(i in nontarget_down_in_tg) {
      g <- i
      cg <- cov[g]
      strand <- cg$strd[1]
      if(is.na(strand) ==TRUE) {
         next
      }
      if(strand == "+") {
          cg <- cg[order(cg$cdsutr, cg$exonid, cg$nt),]
       } else {
          cg <- cg[order(cg$cdsutr, -cg$exonid, -cg$nt),]
      }
      beforestart <- tail(which(cg$cdsutr ==0), n=100)
      afterstart <-head(which(cg$cdsutr ==1), n=50)
      cdsregion <- c(beforestart, afterstart)
      wt1r <- (cov[c(g), sum(wt1)])
      wt2r <- (cov[c(g), sum(wt2)])
```

```
      wt3r <- (cov[c(g), sum(wt3)])
      wt.nrm.strt <-cg$wt1[cdsregion[1]:tail(cdsregion, n=1)]/(wt1r[wt1r[ ,symbol] == g, V1]+1)
+cg$wt2[cdsregion[1]:tail(cdsregion, n=1)]/(wt2r[wt2r[ ,symbol] == g, V1]+1)
+cg$wt3[cdsregion[1]:tail(cdsregion, n=1)]/(wt3r[wt3r[ ,symbol] == g, V1]+1)
      if(length(cdsregion) == 150) {
      } else {
      cdsregion <- append(rep(c(0), each=150-length(cdsregion)), cdsregion)
      wt.nrm.strt <- append(rep(c(0), each=150-length(wt.nrm.strt)), wt.nrm.strt)
      }
   wt.start <- cbind(wt.start, wt.nrm.strt)
}

dim(wt.start) # supposed to 1044 but matched ones are 976
wt.start.trim<- apply(wt.start, 1, mean, trim=0.15)



#TG -->notarget_down_in_tg
tg.start <-matrix(0, nrow=150, ncol=0) #empty numeric vector (inclease speed)
for(i in nontarget_down_in_tg) {
      g <- i
      cg <- cov[g]
      strand <- cg$strd[1]
      if(is.na(strand) ==TRUE) {     # this is for avoid non-gene name matched ones
         next
      }
      if(strand == "+") {
          cg <- cg[order(cg$cdsutr, cg$exonid, cg$nt),]
       } else {
          cg <- cg[order(cg$cdsutr, -cg$exonid, -cg$nt),]
      }
      beforestart <- tail(which(cg$cdsutr ==0), n=100)
      afterstart <-head(which(cg$cdsutr ==1), n=50)
      cdsregion <- c(beforestart, afterstart)
      tg1r <- (cov[c(g), sum(tg1)])
      tg2r <- (cov[c(g), sum(tg2)])
      tg3r <- (cov[c(g), sum(tg3)])
      tg.nrm.strt <-cg$tg1[cdsregion[1]:tail(cdsregion, n=1)]/(tg1r[tg1r[ ,symbol] == g, V1]+1)
+cg$tg2[cdsregion[1]:tail(cdsregion, n=1)]/(tg2r[tg2r[ ,symbol] == g, V1]+1)
+cg$tg3[cdsregion[1]:tail(cdsregion, n=1)]/(tg3r[tg3r[ ,symbol] == g, V1]+1)
      if(length(cdsregion) == 150) {
      } else {
      cdsregion <- append(rep(c(0), each=150-length(cdsregion)), cdsregion)
      tg.nrm.strt <- append(rep(c(0), each=150-length(tg.nrm.strt)), tg.nrm.strt)
      }
   tg.start <- cbind(tg.start, tg.nrm.strt)
}

dim(tg.start)
tg.start.trim<- apply(tg.start, 1, mean, trim=0.15)


#generate matrix for graph
plot <- as.data.frame(cbind(wt.start.trim, tg.start.trim))

#Plotting: WT vs TG --> tg_res_targets
```

```
ggplot(plot, aes(x=as.numeric(row.names(plot))))+
   geom_line(aes(y=wt.start.trim), colour="black", size=2)+
   geom_line(aes(y=tg.start.trim), colour="green", size=2)+
   geom_vline(xintercept=100, linetype="dashed", size=2, color="gray")+
   ggtitle("Nontargets_down_in_tg")+
   labs(x="Start Codon", y="Relative Ribosome Occupancy")+
   geom_segment(mapping=aes(x=100, xend=150, y=-0.0002, yend=-0.0002), size=12, color="dark
grey") +
   geom_segment(mapping=aes(x=0, xend=150, y=-0.0002, yend=-0.0002), size=4, color="dark grey")+
   theme (panel.background = element_rect(fill='white'),    axis.text.x= element_text(color="black"),
axis.text.y = element_blank(),    plot.title = element_text(face="italic", size=14), axis.title =
element_text(size=15)) +
   scale_y_continuous(limit=c(-0.0003, 0.0045))

Save as 6x8
Save as 4x6




#WT -->noago_down_in_tg
wt.start <-matrix(0, nrow=150, ncol=0)
for(i in noago_down_in_tg) {
      g <- i
      cg <- cov[g]
      strand <- cg$strd[1]
      if(is.na(strand) ==TRUE) {
         next
      }
      if(strand == "+") {
          cg <- cg[order(cg$cdsutr, cg$exonid, cg$nt),]
       } else {
          cg <- cg[order(cg$cdsutr, -cg$exonid, -cg$nt),]
      }
      beforestart <- tail(which(cg$cdsutr ==0), n=100)
      afterstart <-head(which(cg$cdsutr ==1), n=50)
      cdsregion <- c(beforestart, afterstart)
      wt1r <- (cov[c(g), sum(wt1)])
      wt2r <- (cov[c(g), sum(wt2)])
      wt3r <- (cov[c(g), sum(wt3)])
      wt.nrm.strt <-cg$wt1[cdsregion[1]:tail(cdsregion, n=1)]/(wt1r[wt1r[ ,symbol] == g, V1]+1)
+cg$wt2[cdsregion[1]:tail(cdsregion, n=1)]/(wt2r[wt2r[ ,symbol] == g, V1]+1)
+cg$wt3[cdsregion[1]:tail(cdsregion, n=1)]/(wt3r[wt3r[ ,symbol] == g, V1]+1)
      if(length(cdsregion) == 150) {
      } else {
      cdsregion <- append(rep(c(0), each=150-length(cdsregion)), cdsregion)
      wt.nrm.strt <- append(rep(c(0), each=150-length(wt.nrm.strt)), wt.nrm.strt)
      }
   wt.start <- cbind(wt.start, wt.nrm.strt)
}

dim(wt.start) # supposed to 755 but matched ones are 687
wt.start.trim<- apply(wt.start, 1, mean, trim=0.15)
```

```
#TG -->nogo_down_in_tg
tg.start <-matrix(0, nrow=150, ncol=0) #empty numeric vector (inclease speed)
for(i in noago_down_in_tg) {
      g <- i
      cg <- cov[g]
      strand <- cg$strd[1]
      if(is.na(strand) ==TRUE) {     # this is for avoid non-gene name matched ones
        next
      }
      if(strand == "+") {
          cg <- cg[order(cg$cdsutr, cg$exonid, cg$nt),]
       } else {
          cg <- cg[order(cg$cdsutr, -cg$exonid, -cg$nt),]
      }
      beforestart <- tail(which(cg$cdsutr ==0), n=100)
      afterstart <-head(which(cg$cdsutr ==1), n=50)
      cdsregion <- c(beforestart, afterstart)
      tg1r <- (cov[c(g), sum(tg1)])
      tg2r <- (cov[c(g), sum(tg2)])
      tg3r <- (cov[c(g), sum(tg3)])
      tg.nrm.strt <-cg$tg1[cdsregion[1]:tail(cdsregion, n=1)]/(tg1r[tg1r[ ,symbol] == g, V1]+1)
+cg$tg2[cdsregion[1]:tail(cdsregion, n=1)]/(tg2r[tg2r[ ,symbol] == g, V1]+1)
+cg$tg3[cdsregion[1]:tail(cdsregion, n=1)]/(tg3r[tg3r[ ,symbol] == g, V1]+1)
      if(length(cdsregion) == 150) {
      } else {
      cdsregion <- append(rep(c(0), each=150-length(cdsregion)), cdsregion)
      tg.nrm.strt <- append(rep(c(0), each=150-length(tg.nrm.strt)), tg.nrm.strt)
      }
   tg.start <- cbind(tg.start, tg.nrm.strt)
}

dim(tg.start)
tg.start.trim<- apply(tg.start, 1, mean, trim=0.15)


#generate matrix for graph
plot <- as.data.frame(cbind(wt.start.trim, tg.start.trim))

#Plotting: WT vs TG
ggplot(plot, aes(x=as.numeric(row.names(plot))))+
   geom_line(aes(y=wt.start.trim), colour="black", size=2)+
   geom_line(aes(y=tg.start.trim), colour="green", size=2)+
   geom_vline(xintercept=100, linetype="dashed", size=2, color="gray")+
   ggtitle("No_ago_down_in_tg")+
   labs(x="Start Codon", y="Relative Ribosome Occupancy")+
   geom_segment(mapping=aes(x=100, xend=150, y=-0.0002, yend=-0.0002), size=12, color="dark
grey") +
   geom_segment(mapping=aes(x=0, xend=150, y=-0.0002, yend=-0.0002), size=4, color="dark grey")+
   theme (panel.background = element_rect(fill='white'),    axis.text.x= element_text(color="black"),
axis.text.y = element_blank(),    plot.title = element_text(face="italic", size=14), axis.title =
element_text(size=15)) +
   scale_y_continuous(limit=c(-0.0003, 0.0048))

Save as 6x8
Save as 4x6
```

```

#WT-TKO comparison next
```{r}

#WT -->notarget_up_in_tko
wt.start <-matrix(0, nrow=150, ncol=0)
for(i in nontarget_up_in_tko) {
      g <- i
      cg <- cov[g]
      strand <- cg$strd[1]
      if(is.na(strand) ==TRUE) {
        next
      }
      if(strand == "+") {
          cg <- cg[order(cg$cdsutr, cg$exonid, cg$nt),]
       } else {
          cg <- cg[order(cg$cdsutr, -cg$exonid, -cg$nt),]
      }
      beforestart <- tail(which(cg$cdsutr ==0), n=100)
      afterstart <-head(which(cg$cdsutr ==1), n=50)
      cdsregion <- c(beforestart, afterstart)
      wt1r <- (cov[c(g), sum(wt1)])
      wt2r <- (cov[c(g), sum(wt2)])
      wt3r <- (cov[c(g), sum(wt3)])
      wt.nrm.strt <-cg$wt1[cdsregion[1]:tail(cdsregion, n=1)]/(wt1r[wt1r[ ,symbol] == g, V1]+1)
+cg$wt2[cdsregion[1]:tail(cdsregion, n=1)]/(wt2r[wt2r[ ,symbol] == g, V1]+1)
+cg$wt3[cdsregion[1]:tail(cdsregion, n=1)]/(wt3r[wt3r[ ,symbol] == g, V1]+1)
      if(length(cdsregion) == 150) {
      } else {
      cdsregion <- append(rep(c(0), each=150-length(cdsregion)), cdsregion)
      wt.nrm.strt <- append(rep(c(0), each=150-length(wt.nrm.strt)), wt.nrm.strt)
      }
   wt.start <- cbind(wt.start, wt.nrm.strt)
}

dim(wt.start) # supposed to 781 but matched ones are 730
wt.start.trim<- apply(wt.start, 1, mean, trim=0.15)



#TKO -->nontarget_up_in_tko
tko.start <-matrix(0, nrow=150, ncol=0) #empty numeric vector (inclease speed)
for(i in nontarget_up_in_tko) {
      g <- i
      cg <- cov[g]
      strand <- cg$strd[1]
```
```

```
      if(is.na(strand) ==TRUE) {     # this is for avoid non-gene name matched ones
        next
      }
      if(strand == "+") {
          cg <- cg[order(cg$cdsutr, cg$exonid, cg$nt),]
       } else {
          cg <- cg[order(cg$cdsutr, -cg$exonid, -cg$nt),]
      }
      beforestart <- tail(which(cg$cdsutr ==0), n=100)
      afterstart <-head(which(cg$cdsutr ==1), n=50)
      cdsregion <- c(beforestart, afterstart)
      tko1r <- (cov[c(g), sum(tko1)])
      tko2r <- (cov[c(g), sum(tko2)])
      tko3r <- (cov[c(g), sum(tko3)])
      tko.nrm.strt <-cg$tko1[cdsregion[1]:tail(cdsregion, n=1)]/(tko1r[tko1r[ ,symbol] == g, V1]+1)
+cg$tko2[cdsregion[1]:tail(cdsregion, n=1)]/(tko2r[tko2r[ ,symbol] == g, V1]+1)
+cg$tko3[cdsregion[1]:tail(cdsregion, n=1)]/(tko3r[tko3r[ ,symbol] == g, V1]+1)
      if(length(cdsregion) == 150) {
      } else {
      cdsregion <- append(rep(c(0), each=150-length(cdsregion)), cdsregion)
      tko.nrm.strt <- append(rep(c(0), each=150-length(tko.nrm.strt)), tko.nrm.strt)
      }
   tko.start <- cbind(tko.start, tko.nrm.strt)
}

dim(tko.start)
tko.start.trim<- apply(tko.start, 1, mean, trim=0.15)


#generate matrix for graph
plot <- as.data.frame(cbind(wt.start.trim, tko.start.trim))

#Plotting: WT vs TKO
ggplot(plot, aes(x=as.numeric(row.names(plot))))+
   geom_line(aes(y=wt.start.trim), colour="black", size=2)+
   geom_line(aes(y=tko.start.trim), colour="red", size=2)+
   geom_vline(xintercept=100, linetype="dashed", size=2, color="gray")+
   ggtitle("Nontargets_up_in_ko")+
   labs(x="Start Codon", y="Relative Ribosome Occupancy")+
   geom_segment(mapping=aes(x=100, xend=150, y=-0.0002, yend=-0.0002), size=12, color="dark
grey") +
   geom_segment(mapping=aes(x=0, xend=150, y=-0.0002, yend=-0.0002), size=4, color="dark grey")+
   theme (panel.background = element_rect(fill='white'),    axis.text.x= element_text(color="black"),
axis.text.y = element_blank(),    plot.title = element_text(face="italic", size=14), axis.title =
element_text(size=15)) +
   scale_y_continuous(limit=c(-0.0003, 0.0045))

Save as 6x8
Save as 4x6



#WT -->noago_up_in_tko
wt.start <-matrix(0, nrow=150, ncol=0)
for(i in noago_up_in_tko) {
      g <- i
```

```
      cg <- cov[g]
      strand <- cg$strd[1]
      if(is.na(strand) ==TRUE) {
         next
      }
      if(strand == "+") {
          cg <- cg[order(cg$cdsutr, cg$exonid, cg$nt),]
       } else {
          cg <- cg[order(cg$cdsutr, -cg$exonid, -cg$nt),]
      }
      beforestart <- tail(which(cg$cdsutr ==0), n=100)
      afterstart <-head(which(cg$cdsutr ==1), n=50)
      cdsregion <- c(beforestart, afterstart)
      wt1r <- (cov[c(g), sum(wt1)])
      wt2r <- (cov[c(g), sum(wt2)])
      wt3r <- (cov[c(g), sum(wt3)])
      wt.nrm.strt <-cg$wt1[cdsregion[1]:tail(cdsregion, n=1)]/(wt1r[wt1r[ ,symbol] == g, V1]+1)
+cg$wt2[cdsregion[1]:tail(cdsregion, n=1)]/(wt2r[wt2r[ ,symbol] == g, V1]+1)
+cg$wt3[cdsregion[1]:tail(cdsregion, n=1)]/(wt3r[wt3r[ ,symbol] == g, V1]+1)
      if(length(cdsregion) == 150) {
      } else {
      cdsregion <- append(rep(c(0), each=150-length(cdsregion)), cdsregion)
      wt.nrm.strt <- append(rep(c(0), each=150-length(wt.nrm.strt)), wt.nrm.strt)
      }
   wt.start <- cbind(wt.start, wt.nrm.strt)
}

dim(wt.start) # supposed to 627 matched 578
wt.start.trim<- apply(wt.start, 1, mean, trim=0.15)



#TKO -->noago_up_in_tko
tko.start <-matrix(0, nrow=150, ncol=0) #empty numeric vector (inclease speed)
for(i in noago_up_in_tko) {
      g <- i
      cg <- cov[g]
      strand <- cg$strd[1]
      if(is.na(strand) ==TRUE) {     # this is for avoid non-gene name matched ones
         next
      }
      if(strand == "+") {
          cg <- cg[order(cg$cdsutr, cg$exonid, cg$nt),]
       } else {
          cg <- cg[order(cg$cdsutr, -cg$exonid, -cg$nt),]
      }
      beforestart <- tail(which(cg$cdsutr ==0), n=100)
      afterstart <-head(which(cg$cdsutr ==1), n=50)
      cdsregion <- c(beforestart, afterstart)
      tko1r <- (cov[c(g), sum(tko1)])
      tko2r <- (cov[c(g), sum(tko2)])
      tko3r <- (cov[c(g), sum(tko3)])
      tko.nrm.strt <-cg$tko1[cdsregion[1]:tail(cdsregion, n=1)]/(tko1r[tko1r[ ,symbol] == g, V1]+1)
+cg$tko2[cdsregion[1]:tail(cdsregion, n=1)]/(tko2r[tko2r[ ,symbol] == g, V1]+1)
+cg$tko3[cdsregion[1]:tail(cdsregion, n=1)]/(tko3r[tko3r[ ,symbol] == g, V1]+1)
      if(length(cdsregion) == 150) {
```

```
    } else {
    cdsregion <- append(rep(c(0), each=150-length(cdsregion)), cdsregion)
    tko.nrm.strt <- append(rep(c(0), each=150-length(tko.nrm.strt)), tko.nrm.strt)
    }
  tko.start <- cbind(tko.start, tko.nrm.strt)
}

dim(tko.start)
tko.start.trim<- apply(tko.start, 1, mean, trim=0.15)


#generate matrix for graph
plot <- as.data.frame(cbind(wt.start.trim, tko.start.trim))

#Plotting: WT vs TKO
ggplot(plot, aes(x=as.numeric(row.names(plot))))+
  geom_line(aes(y=wt.start.trim), colour="black", size=2)+
  geom_line(aes(y=tko.start.trim), colour="red", size=2)+
  geom_vline(xintercept=100, linetype="dashed", size=2, color="gray")+
  ggtitle("Noago_up_in_ko")+
  labs(x="Start Codon", y="Relative Ribosome Occupancy")+
  geom_segment(mapping=aes(x=100, xend=150, y=-0.0002, yend=-0.0002), size=12, color="dark
grey") +
  geom_segment(mapping=aes(x=0, xend=150, y=-0.0002, yend=-0.0002), size=4, color="dark grey")+
  theme (panel.background = element_rect(fill='white'),   axis.text.x= element_text(color="black"),
axis.text.y = element_blank(),   plot.title = element_text(face="italic", size=14), axis.title =
element_text(size=15)) +
  scale_y_continuous(limit=c(-0.0003, 0.0045))

Save as 6x8
Save as 4x6




```