

Isotype distribution analysis

Mikhail Shugay, Anna Obraztsova

11/17/2016

Isotype distribution

```
library(plyr)
library(dplyr)
library(ggplot2)
library(stringr)
library(reshape2)

rna <- data.frame()

old_rna = c("Abdulain", "Ilgen", "Mamaev", "Smirnov", "Vlasov")
young_rna = c("Antipyat", "Epifancev", "Hadjibekov", "Koshkin", "Kovalchuk")

for (sample in old_rna){
  .df <- read.table(paste('data/mixcr_yf_old_RNA/', sample, ".txt.gz", sep = ""), header=T, sep="\t")
  .df$proj <- "old"
  .df$sample <- sample
  rna <- rbind(rna, .df)
}

for (sample in young_rna){
  .df <- read.table(paste('data/mixcr_yf_young_RNA/', sample, ".txt.gz", sep = ""), header=T, sep="\t")
  .df$proj <- "young"
  .df$sample <- sample
  rna <- rbind(rna, .df)
}

new_colnames = c('clone.id', 'clone.count', 'clone.fraction', 'clonal.seq', 'clonal.seq.qual', 'all.v.hits',
  'all.j.hits', 'all.c.hits', 'all.v.alignments', 'all.d.alignments', 'all.j.alignments', 'all.c.alignments',
  'nt.seq.FR1', 'min.qual.FR1', 'nt.seq.CDR1', 'min.qual.CDR1', 'nt.seq.FR2', 'min.qual.FR2', 'nt.seq.CDR2',
  'nt.seq.FR3', 'min.qual.FR3', 'nt.seq.CDR3', 'min.qual.CDR3', 'nt.seq.FR4', 'min.qual.FR4', 'aa.seq.FR1',
  'aa.seq.FR2', 'aa.seq.CDR2', 'aa.seq.FR3', 'aa.seq.CDR3', 'aa.seq.FR4', 'ref.points', 'proj', 'sample')

colnames(rna) <- new_colnames
rna <- mutate(rna, isotype = str_sub(all.c.hits, 1, 4))

rna.2 <- ddply(rna, .(proj, sample), summarize,
  share.IGHA = sum(clone.fraction[which(isotype == "IGHA")]),
  share.IGHD = sum(clone.fraction[which(isotype == "IGHD")]),
  share.IGHE = sum(clone.fraction[which(isotype == "IGHE")]),
  share.IGHG = sum(clone.fraction[which(isotype == "IGHG")]),
  share.IGHM = sum(clone.fraction[which(isotype == "IGHM")]))

rna.3 <- melt(rna.2, id.vars = c("proj", "sample"))
ggplot(rna.3, aes(x=proj, group=proj, y=value)) + geom_boxplot() + facet_grid(.~variable, scales="free")
```

