

Pan Cancer MK2 transcript clinical annotation

Karthik Suresh

3/11/2020

We have already downloaded transcript data for MK2 across a variety of cancers. Here, we need to collate all the transcript data, go find all the relevant clinical data using GDC_query, annotate, and save a giant file containing transcript and clinical data across many cancers.

```
library(knitr)
opts_chunk$set(eval=FALSE, echo=TRUE)

knitr::opts_chunk$set(echo = TRUE, eval=FALSE)
library("SummarizedExperiment")
library("dplyr")
library("DT")
library("TCGAbiolinks")
library("plyr")
library("dplyr")
rm(list=ls())

filelist <- as.data.frame(list.files("../rawdata/panca.mk2.oncolnc/"))
colnames(filelist)<- c("filename")
filelist$cancer.type <- lapply(as.character(filelist$filename), function(x) return(unlist(strsplit(x,"_"))))

GetTranscriptData <- function(file, ca.type) {

  if(!exists("MasterTranscriptList")) {
    MasterTranscriptList <- read.csv(paste0("../rawdata/panca.mk2.oncolnc/", file))
    MasterTranscriptList$cancer.type <- ca.type
  }
  else {
    currTranscriptList <- read.csv(paste0("../rawdata/panca.mk2.oncolnc/", file))
    currTranscriptList$cancer.type <- ca.type
    MasterTranscriptList <- rbind(MasterTranscriptList, currTranscriptList)
  }
  return(1)
}

filelist$collectTranscript <- mapply(GetTranscriptData, filelist$filename, filelist$cancer.type)

GetClinicalData <- function(ca.type) {
  if(!exists("MasterClinicalList")) {

    MasterClinicalList <- GDCquery_clinic(project = paste0("TCGA-",ca.type), type = "clinical")
    MasterClinicalList$cancer.type <- ca.type
```

```

}
else {
  currClinicalList <- GDCquery_clinic(project = paste0("TCGA-",ca.type), type = "clinical")
  currClinicalList$cancer.type <- ca.type
  MasterClinicalList <- rbind.fill(MasterClinicalList, currClinicalList)

}

return(1)

}

filelist$collectClinical <- lapply(filelist$cancer.type, GetClinicalData)
MasterList <- left_join(MasterClinicalList, MasterTranscriptList, by=c("submitter_id"="Patient"))

write.table(MasterList, file = "../rawdata/MasterMK2data.csv", quote=FALSE, sep="\t", col.names=TRUE, r

```

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this: