metadata_script_for_lusha

Jenea Adams

6/26/2023

packages

```
knitr::opts_chunk$set(echo = T, results = "show", fig.width=8, fig.height=6, tidy.opts=list(width.cutof
options(scipen = 0, digits = 3)  # controls base R output
# check if you have ISLR package, if not, install it
if(!require('pacman')) {install.packages('pacman')}
```

```
## Loading required package: pacman
```

```
pacman::p_load(dplyr, readr, tidyverse, ggplot2, formatR)
```

Read in the data

```
setwd("~/OneDrive - Children's Hospital of Philadelphia/COG-pediatric-AML/metadata_v2")
wgs = read.csv("WGS.csv", header = T)  # point to path of wgs Cavatica CSV
rna = read.csv("RNA.csv", header = T)  # point to path of rna Cavatica CSV
new_pat.met = read.csv("NCT01371981-D4-Dataset.csv")  # point to path of new metadata file
```

Change the name of rna case_id –> usi

```
colnames(rna)[1] = "usi"
# head(rna)

colnames(wgs)[1] = "usi"  # same for wgs
# head(wgs)
```

subset cavatica data

RNA

N individuals with WGS (normal) x RNA-seq tumor x Pat.Met –> 390

```
rna_norm = filter(rna, sample_type == "Normal")

rna_tum = filter(rna, sample_type == "Tumor")



# x2 = filter(rna_tum, rna_tum$usi %in% wgs$usi & rna_tum$usi %in% regno$usi & !rna_tum %in%
# rna_norm)

rna.wgs = merge(rna, wgs, by = "usi")
rna_tum.wgs = merge(rna_tum, wgs, by = "usi")
```

WGS

```
wgs_norm = filter(wgs, sample_type == "Normal")

wgs_tum = filter(wgs, sample_type == "Tumor")
```

Filter out Study Ineligible Patients

```
new_pat.met_eligible = new_pat.met %>%
    filter(!grepl("02", consort_clsf))
```

Number of eligible patients with RNAseq data

```
length(intersect(rna$usi, new_pat.met_eligible$usi))  ## --> 1613
```

```
## [1] 1040
```

creating table

row 1

WGS normal, !WGS tumor, !RNAseq normal, RNA-seq tumor, Patient Metadata

```
row1 = filter(new_pat.met_eligible, new_pat.met_eligible$usi %in% rna_tum$usi & new_pat.met_eligible$usi
    wgs$usi & !new_pat.met_eligible$usi %in% rna_norm$usi)

nrow(row1)
```

```
## [1] 390
```

row 2

!wgs_normal, !wgs_tumor, !rna_normal, rna_tumor, patient metadata

```
row2 = filter(new_pat.met_eligible, new_pat.met_eligible$usi %in% rna_tum$usi & !new_pat.met_eligible$us
    wgs_norm$usi & !new_pat.met_eligible$usi %in% wgs_tum$usi & !new_pat.met_eligible$usi %in% rna_norm$
nrow(row2)
```

```
## [1] 650
```

row 3

wgs_normal, !wgs_tumor, !rna_normal, rna_tumor, !patient metadata

```
row3 = filter(wgs, wgs$usi %in% rna_tum$usi & wgs$usi %in% wgs_norm$usi & !wgs$usi %in% wgs_tum$usi &
    !wgs$usi %in% rna_norm$usi & !wgs$usi %in% new_pat.met_eligible$usi)
nrow(row3)
```

```
## [1] 0
```

row 4

wgs_norm, !wgs_tum, !rna_norm, !rna_tum, pat met

```
row4 = filter(wgs, wgs$usi %in% wgs_norm$usi & !wgs$usi %in% wgs_tum$usi & !wgs$usi %in% rna_norm$usi &
    !wgs$usi %in% rna_tum$usi & wgs$usi %in% new_pat.met_eligible)
nrow(row4)
```

```
## [1] 0
```

row 5

wgs_norm, !wgs_tum, !rna_norm, !rna_tum, !pat met

```
row5 = filter(wgs, wgs$usi %in% wgs_norm$usi & !wgs$usi %in% wgs_tum$usi & !wgs$usi %in% rna_norm$usi &
    !wgs$usi %in% rna_tum$usi & !wgs$usi %in% new_pat.met_eligible)
nrow(row5)
```

```
## [1] 18
```

row 6

!wgs_norm, !wgs_tum, rna_norm, !rna_tum, !pat met

```
row6 = filter(rna, !rna$usi %in% wgs_norm$usi & !rna$usi %in% wgs_tum$usi & rna$usi %in% rna_norm$usi &
    !rna$usi %in% rna_tum$usi & !rna$usi %in% new_pat.met_eligible)
nrow(row6)
```

```
## [1] 62
```

row 7

```
row7 = filter(rna, !rna$usi %in% wgs_norm$usi & !rna$usi %in% wgs_tum$usi & !rna$usi %in% rna_norm$usi &
    rna$usi %in% rna_tum$usi & !rna$usi %in% new_pat.met_eligible)
nrow(row7)
```

```
## [1] 661
```