

survival.R

u1413890

2022-11-10

```
#####plotting survival curve with full cohort data.
```

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.2 --
```

```
## v ggplot2 3.4.0      v purrr  0.3.5
```

```
## v tibble  3.1.8      v dplyr  1.0.10
```

```
## v tidyr   1.2.1      v stringr 1.4.1
```

```
## v readr   2.1.3      v forcats 0.5.2
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()     masks stats::lag()
```

```
library(tinytex)
```

```
#1st thing, creating a clean dataset that can be easily used for plotting.
```

```
#pull data from compiled_cohorts.csv, which has
```

```
read_csv("compiled_cohorts.csv", show_col_types = FALSE)%>%
```

```
#selecting variables that are needed for survival curve plots,
```

```
#plus identifiers if needed to link back to main data.
```

```
select_at(vars(mouse_num,
```

```
  dob,
```

```
  virus,
```

```
  injection_date,
```

```
  death_date,
```

```
  tumor_noticed,
```

```
  behavior_noticed,
```

```
  strain.x,
```

```
  strain.y,
```

```
  src,
```

```
  exclude))%>%
```

```
#selecting just strains temporarily to make the strains data easier to check.
```

```
#select_at(vars(strain.x, strain.y))%>%
```

```
separate(strain.y, sep = ";",
```

```
  into = c("tva", "pten", "h11", "ink", "atrx"))%>%
```

```
mutate(tva = "N-TVA:")%>%
```

```
mutate(strain.x = gsub(".*:", "", strain.x)%>%trimws())%>%
```

```
mutate(pten = if_else(grepl("Pten f/f", strain.x), "Pten f/f", pten))%>%
```

```
mutate(h11 = if_else(grepl("H11LSL-Cas9", strain.x), "H11LSL-Cas9 f/f", trimws(h11)))%>%
```

```
mutate(ink = if_else(grepl("Ink4a/Arf f/f", strain.x), "Ink4a/Arf f/f", ink))%>%
```

```
mutate(atrx = if_else(grepl("Atrxf/f", strain.x), "Atrxf/f", atrx))%>%
```

```
select(-strain.x)%>%
```

```

mutate(virus = if_else(virus == "Rcas Cre", "RCAS-Cre",
                      if_else(virus == "RCAS Cre-U6sgRNA-ex2-NF1", "RCAS-Cre-pENTR-U6-sgRNA-NF1-ex2", virus)),
relocate(h11, .after = tva))>%
mutate(pten = if_else(is.na(pten), "Pten +/+", trimws(pten)))>%
mutate(ink = if_else(is.na(ink), "Ink4a/Arf +/+", trimws(ink)))>%
mutate(atrx = if_else(is.na(atrx), "ATRX +/+", atrx))>%
mutate(atrx = gsub("atrx", "ATRX", atrx, ignore.case = TRUE))>%trimws())>%
mutate(atrx = if_else(atrx == "ATRXf/f", "ATRX f/f", atrx))>%
group_by(mouse_num)>%
arrange(desc(death_date))>%
slice(1)>%
ungroup()>%
mutate(strain = paste(sep = ";", paste0(tva, h11), pten, ink, atrx))>%
mutate(nf1_ko = if_else(grepl("NF1", virus), "nf1 KO", "nf1 wt"))>%
mutate(pten_ko = if_else(grepl("f/f", pten), "pten KO", "pten wt"))>%
mutate(ink_ko = if_else(grepl("f/f", ink), "ink KO", "ink wt"))>%
mutate(atrx_ko = if_else(grepl("f/f", atrx), "atrx KO", "atrx wt"))>%
mutate(genes_ko = paste(nf1_ko, pten_ko, ink_ko, atrx_ko, sep = ";"))>%
filter(is.na(exclude))>%
select(-exclude)>%
relocate(virus, .after = strain)>%
mutate(age_death_capped = as.numeric(death_date - injection_date))>%
relocate(age_death_capped, .after = death_date)>%
mutate(age_death_capped = if_else(age_death_capped > 148, 300, age_death_capped))>%

write_csv("survival_compiled.csv")
#count(tva)>%
#view()

leg_order<-
read_csv("survival_compiled.csv")>%
  count(genes_ko)>%
  select(1)>%
  pull(genes_ko)

## Rows: 299 Columns: 20
## -- Column specification -----
## Delimiter: ","
## chr  (13): tva, h11, pten, ink, atrx, src, strain, virus, nf1_ko, pten_ko, i...
## dbl  (2): mouse_num, age_death_capped
## date (5): dob, injection_date, death_date, tumor_noticed, behavior_noticed
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.

library(survminer)

## Loading required package: ggpubr

```

```

library(survival)

##
## Attaching package: 'survival'
##
## The following object is masked from 'package:survminer':
##
##      myeloma

survfit(Surv(time = age_death_capped)~genes_ko, data = read_csv("survival_compiled.csv"))%>%

ggsurvplot(xlim = c(0, 150),
            ylim = c(0, 1.02),
            size = 2,
            alpha = .75,
            break.x.by = 25,
            break.y.by = .25,
            axes.offset = FALSE,
            legend = "right",
            ggtheme = theme_classic(),
            palette = c("locuszoom"),
            xlab = "Time Post Injection (Days)",
            legend.title = "Genes KO",
            legend.lab = leg_order
)

## Rows: 299 Columns: 20
## -- Column specification -----
## Delimiter: ","
## chr  (13): tva, h11, pten, ink, atrx, src, strain, virus, nf1_ko, pten_ko, i...
## dbl  (2): mouse_num, age_death_capped
## date (5): dob, injection_date, death_date, tumor_noticed, behavior_noticed
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
## Rows: 299 Columns: 20
## -- Column specification -----
## Delimiter: ","
## chr  (13): tva, h11, pten, ink, atrx, src, strain, virus, nf1_ko, pten_ko, i...
## dbl  (2): mouse_num, age_death_capped
## date (5): dob, injection_date, death_date, tumor_noticed, behavior_noticed
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.

```

