sankey_BRM_AIMS_Fig2H.R

t

2024-11-18

```
# HEADER ####
#
# Version: 2024-11-18
#
# Figure 2H: Sankey plots of AIMS subtypes in pre-Tx / post-Tx / metastasis
#
# !! IMPORTANT !!:
#
# The devtools() package must already be installed
#   for the installation of ggsankey()
#
#
# SETUP ####

Sys.setenv(lang = "en_US")
```

Install required packages if missing ————————————————–

```
# Package ggsankey from GitHub using devtools
if ("ggsankey" %in% rownames(installed.packages) == FALSE) {
  devtools::install_github("davidsjoberg/ggsankey")
}
```

```
## WARNING: Rtools is required to build R packages, but is not currently installed.
##
## Please download and install Rtools 4.4 from https://cran.r-
project.org/bin/windows/Rtools/.
```

```
## Skipping install of 'ggsankey' from a github remote, the SHA1 (b675d0d5) has not
changed since last install.
##   Use `force = TRUE` to force installation
```

```
# Package names for install from CRAN
packs <- c("ggplot2", "dplyr", "ggalluvial", "networkD3")

# Install packages not yet installed
installed_packages <- packs %in% rownames(installed.packages())
if (any(installed_packages == FALSE)) {
  install.packages(packs[!installed_packages])
}
```

Load required packages ————————————————–

```
invisible(library(ggsankey))

invisible(lapply(packs, library, character.only = TRUE))
```

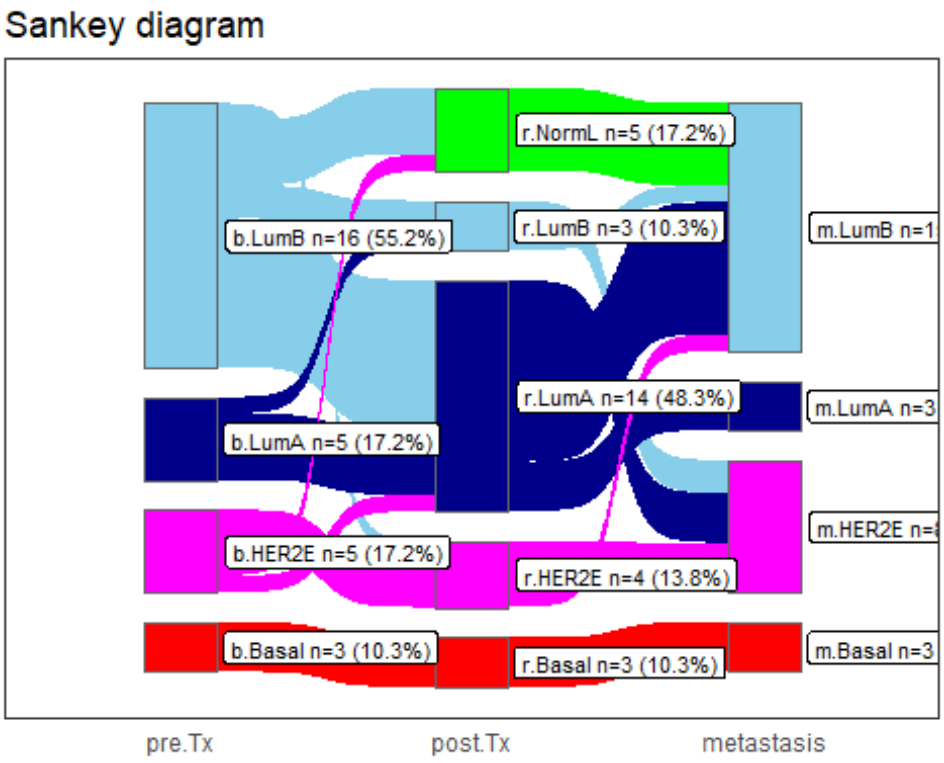
```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

## Warning: package 'ggalluvial' was built under R version 4.4.2

## Warning: package 'networkD3' was built under R version 4.4.2
```

```r
set.seed(321)
```

```r
# IMPORT ####

# Import AIMS subtypes of pre-Tx biopsy, post-Tx resect, and metast. sample

sank_df <- read.delim("BRM_AIMS.txt", header=T, stringsAsFactors = T, skipNul=T)

# Analysis / Plots ####

df <- sank_df %>%
  make_long(pre.Tx,post.Tx, metastasis) |> mutate(next_node =
forcats::fct_inorder(next_node))

## Warning: attributes are not identical across measure variables; they will be
## dropped

# counts and percentages
  TotalCount = nrow(sank_df)
  dagg <- df%>%
    dplyr::group_by(node)%>%
    tally()

  dagg <- dagg%>%
    dplyr::group_by(node)%>%
    dplyr::mutate(pct = n/TotalCount)


# visualising plot
  df2 <- merge(df, dagg,   by.x = 'node',by.y = 'node')
  pl <- ggplot(df2, aes(x = x
                        , node = node
                        , next_x = next_x
                        , next_node = next_node
                        , fill = factor(node),
                        , label = paste0(node," n=", n, ' (',  round(pct* 100,1), '%)' ))
)

  pl <- pl +geom_sankey(width = 1/4, flow.alpha = 1.5,  node.color = "gray40",
show.legend = TRUE)+
        geom_sankey_label(size = 3, color = "black", fill= "white", hjust = -0.2)+
        theme_bw()+
        theme(legend.position = "none")+
        theme(axis.title = element_blank(),
                          axis.text.y = element_blank(),
                          axis.ticks = element_blank(),
                          panel.grid = element_blank())+

        labs(fill = 'Nodes')+
        labs(title = "Sankey diagram")+
        scale_fill_manual(values
=c('b.Basal'="red",'b.HER2E'="magenta",'b.LumA'='darkblue','b.LumB'="skyblue",'b.NormL'="
green",

'r.Basal'="red",'r.HER2E'="magenta",'r.LumA'='darkblue','r.LumB'="skyblue",'r.NormL'="gre
en",
```

```
'm.Basal'="red",'m.HER2E'="magenta",'m.LumA'='darkblue','m.LumB'="skyblue",'mr.NormL'="gr
een"))
  pl
```




```
ggsave( "./Fig2H_Sankey_plotBRM_AIMS_rev.pdf", pointsize = 12, bg = "white")

## Saving 5 x 4 in image

# SESSION INFO ####
sessionInfo()

## R version 4.4.1 (2024-06-14 ucrt)
## Platform: x86_64-w64-mingw32/x64
## Running under: Windows 11 x64 (build 22631)
##
## Matrix products: default
##
##
## locale:
## [1] LC_COLLATE=German_Germany.utf8  LC_CTYPE=German_Germany.utf8
## [3] LC_MONETARY=German_Germany.utf8 LC_NUMERIC=C
## [5] LC_TIME=German_Germany.utf8
##
## time zone: Europe/Berlin
## tzcode source: internal
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] networkD3_0.4      ggalluvial_0.12.5  dplyr_1.1.4        ggplot2_3.5.1
## [5] ggsankey_0.0.99999
##
```

```
## loaded via a namespace (and not attached):
##  [1] tidyr_1.3.1       generics_0.1.3    utf8_1.2.4        stringi_1.8.4
##  [5] digest_0.6.37     magrittr_2.0.3    evaluate_1.0.0    grid_4.4.1
##  [9] pkgload_1.4.0     fastmap_1.2.0     processx_3.8.4    pkgbuild_1.4.5
## [13] sessioninfo_1.2.2 urlchecker_1.0.1  ps_1.7.7          promises_1.3.0
## [17] purrr_1.0.2       fansi_1.0.6       scales_1.3.0      textshaping_0.4.0
## [21] cli_3.6.3         shiny_1.9.1       rlang_1.1.4       ellipsis_0.3.2
## [25] munsell_0.5.1     withr_3.0.1       remotes_2.5.0     cachem_1.1.0
## [29] yaml_2.3.10       devtools_2.4.5    tools_4.4.1       memoise_2.0.1
## [33] colorspace_2.1-1  httpuv_1.6.15     forcats_1.0.0     curl_5.2.2
## [37] vctrs_0.6.5       R6_2.5.1          mime_0.12         lifecycle_1.0.4
## [41] stringr_1.5.1     fs_1.6.4          htmlwidgets_1.6.4 usethis_3.0.0
## [45] miniUI_0.1.1.1    ragg_1.3.2        pkgconfig_2.0.3   callr_3.7.6
## [49] pillar_1.9.0      later_1.3.2       gtable_0.3.5      glue_1.7.0
## [53] profvis_0.4.0     Rcpp_1.0.13       systemfonts_1.1.0 tidyselect_1.2.1
## [57] xfun_0.47         tibble_3.2.1      rstudioapi_0.16.0 knitr_1.48
## [61] farver_2.1.2      xtable_1.8-4      igraph_2.1.1      htmltools_0.5.8.1
## [65] labeling_0.4.3    rmarkdown_2.28    compiler_4.4.1
```