

# SFigS3\_GSEA\_scatterplot.R

t

2024-12-18

```
# HEADER #####
#
# Version: 2024-12-18
#
# Figure S3B:
#   Analysis of prognostic gene sets in pre-Tx biopsies
#   and post-Tx residual tumors.
#
#
#
#
# SETUP #####
```

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

Install required packages if missing —————

```
# Package names from CRAN
packs <- c("tidyverse", "ggrepel")

# 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(lapply(packs, library, character.only = TRUE))

## — Attaching core tidyverse packages ————— tidyverse 2.0.0 —
## ✓ dplyr      1.1.4      ✓ readr      2.1.5
## ✓ forcats    1.0.0      ✓ stringr   1.5.1
## ✓ ggplot2     3.5.1      ✓ tibble     3.2.1
## ✓ lubridate  1.9.3      ✓ tidyr      1.3.1
## ✓ purrr       1.0.2
## — Conflicts ————— tidyverse_conflicts() —
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts
to become errors
```

```

# FUNCTION Definitions #####
# *****
#
# The following functions shift_axis_y() and shift_axis_x()
# are applied to place x and y axes in the plots
# to hazard ratios = "1" (instead of "0")

shift_axis_y <- function(p, y=0){
  g <- ggplotGrob(p)
  dummy <- data.frame(y=y)
  ax <- g[["grobs"]][g$layout$name == "axis-b"][[1]]
  p + annotation_custom(grid::grobTree(ax,
                                         vp = grid::viewport(
                                           y=0, height=sum(ax$height))),
                        ymax=y, ymin=y) +
  geom_hline(aes(yintercept=y), data = dummy) +
  theme(axis.text.x = element_blank(),
        axis.ticks.x=element_blank())
}

shift_axis_x <- function(p, x=0){
  g <- ggplotGrob(p)
  dummy <- data.frame(x=x)
  ax <- g[["grobs"]][g$layout$name == "axis-l"][[1]]
  p + annotation_custom(grid::grobTree(ax,
                                         vp = grid::viewport(
                                           x=0, width = sum(ax$height))),
                        xmax=x, xmin=x) +
  geom_vline(aes(xintercept=x), data = dummy) +
  theme(axis.text.y = element_blank(),
        axis.ticks.y=element_blank(),
        legend.position = "none")
}

```

---

```

# IMPORT #####

# Normalized Enrichment Scores (NES) of pathways for iDFS prognosis
# from biopsy and resect samples

df <- read.table("biopsy_resect_HR_basedGSEA_20092024.txt", header=TRUE, sep='\t')

# Define four quadrants Q1-Q4 for coloring
df$quadrant <- dplyr::case_when(df$Resect > 0 & df$Biopsy > 0 ~ "Q1", #Q1... both x and
                                y are positive
                                df$Resect > 0 & df$Biopsy < 0 ~ "Q2",  #Q2... x is neg
                                and y is positive
                                df$Resect < 0 & df$Biopsy < 0 ~ "Q3",  #Q3... both x and
                                y are negative
                                df$Biopsy > 0 & df$Resect < 0 ~ "Q4")  #Q4... x is pos
                                and y is negative

testplot<-ggplot(df, aes(x = Biopsy, y =Resect, color = quadrant), guides(fill = FALSE,
color = FALSE)) +

```

```

geom_point(size=2) +
geom_text_repel(aes(label = Pathway),size=3.5, box.padding = 0.5, max.overlaps=Inf) +
theme(panel.background = element_rect(fill="white"),
      plot.margin = margin(1, 1, 1, 1, "cm"),
      axis.title=element_blank(),
      axis.ticks.length=unit(.2, "cm"))+
scale_x_continuous(limits = c(-3, 3), n.breaks=7) +
scale_y_continuous(limits = c(-3, 3),n.breaks=7)+
labs(title = "paired 540 samples",
      caption = "NES: normalized enrichment score \n(pos. NES: genes set enriched in
tumors from patients with good prognosis; neg. NES = gene set enriched in tumors from
patients with poor prognosis)") +
theme(plot.title = element_text(hjust = 0, vjust = 10, margin = margin(t = 10)),
      plot.caption = element_text(hjust = 0))

# Add central axes and axis labels

p1 <- testplot +
  annotate("segment", x = -Inf, xend = Inf, y = 0, linewidth=1.5, yend = 0, color =
"purple") +
  annotate("segment", x = 0, xend = 0, y = -Inf, linewidth=1.5, yend = Inf, color =
"blue") +

  annotate("text", x = -Inf, y = 1, label = "NES for iDFS in post.Tx \u2192 improved
survival", angle = 90,size=4, hjust=1.0, vjust=-2.0, color = "red") +
  annotate("text", x = -1, y = Inf, label = "NES for iDFS in pre.Tx \u2192 improved
survival", hjust=0,vjust=30, size=4, color = "blue") +

  annotate("text", x = -Inf, y = Inf, label = "good iDFS in post.Tx &\npoor iDFS in
pre.Tx", vjust = 1, hjust=0, size=3.5)+
  annotate("text", x = -Inf, y = -Inf, label = "poor iDFS in post.Tx & pre.Tx", vjust =
0, hjust=0, size=3.5)+
  annotate("text", x = Inf, y = Inf, label = "good iDFS in post.Tx & pre.Tx", vjust = 1,
hjust=1, size=3.5)+
  annotate("text", x = Inf, y = -Inf, label = "good iDFS in pre.Tx &\npoor iDFS in
post.Tx", vjust = 0, hjust=1, size=3.5)+

  coord_cartesian(clip = "off") # Allow annotations to be outside the plot area

# Shift axis labels

p1<-shift_axis_y(p1, y=0)
p1<-shift_axis_x(p1, x=0)
p1

```

good iDFS in post Tx & pre Tx

poor iDFS in post Tx & pre Tx

CD117 (HTG)

IL6 JAK STAT3 SIGNALING (HTG)

IFN GAMMA (HTG)

TSSUE SPECIFIC CYCLES (HTG)

OXYDATIVE PHOSPHORYLATION (HTG)

MITOTIC SPINDLE (HTG)

G2M CHECK POINT RESPONSE (HTG)

E2F TARGETS (HTG)

good iDFS in post Tx & pre Tx

poor iDFS in post Tx & pre Tx

JAK STAT PATHWAY (HTG)

HIF1A SIGNALING (HTG)

DNA REPAIR (HTG)

GLYCOLYSIS (HTG)

good iDFS in post Tx & pre Tx

poor iDFS in post Tx & pre Tx

NES: normalized enrichment score

(pos. NES: genes set enriched in tumors from patients with good prognosis; ne:

```
ggsave(file="SFig3B.svg", plot=p1, width=10, height=8)
dev.off()

## null device
##      1
```

# SESSION INFO ####

sessionInfo()

```
## R version 4.4.2 (2024-10-31 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] ggrepel_0.9.6    lubridate_1.9.3 forcats_1.0.0    stringr_1.5.1
## [5] dplyr_1.1.4      purrr_1.0.2     readr_2.1.5     tidyr_1.3.1
## [9] tibble_3.2.1     ggplot2_3.5.1   tidyverse_2.0.0
##
## loaded via a namespace (and not attached):
## [1] gtable_0.3.6      highr_0.11        compiler_4.4.2    Rcpp_1.0.13
## [5] tidyselect_1.2.1  textshaping_0.4.0 systemfonts_1.1.0 scales_1.3.0
## [9] yaml_2.3.10       fastmap_1.2.0     R6_2.5.1          labeling_0.4.3
## [13] generics_0.1.3    knitr_1.48        munsell_0.5.1     svglite_2.1.3
## [17] pillar_1.9.0      tzdb_0.4.0        rlang_1.1.4       utf8_1.2.4
## [21] stringi_1.8.4     xfun_0.47         timechange_0.3.0  cli_3.6.3
## [25] withr_3.0.2       magrittr_2.0.3    digest_0.6.37     grid_4.4.2
## [29] rstudioapi_0.16.0 hms_1.1.3         lifecycle_1.0.4   vctrs_0.6.5
## [33] evaluate_1.0.0    glue_1.7.0        farver_2.1.2      ragg_1.3.2
## [37] fansi_1.0.6       colorspace_2.1-1  rmarkdown_2.28    tools_4.4.2
## [41] pkgconfig_2.0.3   htmltools_0.5.8.1
```