HRscat_BiopResec_335genes_Diff-Colors.R

t

2025-01-07

```
# HEADER ####
#
# Version: 2024-09-19
#
# Scatter plots of 1/HR for biopsy vs resect samples for all 335 genes
#
# Coloring according different classifications of genes
#
#
#
#
# SETUP ####

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

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

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

# 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
##
## Attaching package: 'crosstable'
##
##
## The following object is masked from 'package:purrr':
##
##     compact
```

```
##
##
##
## Attaching package: 'flextable'
##
##
## The following object is masked from 'package:purrr':
##
##     compose
```

```
# 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())
}
```

```
# IMPORT ####


# Hazard ratios from biopsy and resect samples for 335 genes
invHRs355 <- read.table("Inverse_HR_biopsy_resec_335genes.txt",
                header=TRUE, sep='\t')


# Gene infos for all 335 genes from clustering
n335info <- read.table("Penelope_n335genes_info.txt",
                       header=TRUE, sep='\t')

# Pathway information for all 2549 HTG genes
pathways <- read_excel("HTG-Pathways.xlsx", na="")
```

```
# Geneset definitions ####
```

Definitions of gene sets for all analyses ——————————

```
# Definition of several sublists of genes (Genesets g0, g1, g2....) based on
# pathway information for all 2549 HTG genes
# These Genesets will be used below to assign individual genes to classes.

data <- pathways

# Geneset containing all genes:

g0.all <- data %>%
  select(Genes) %>% pull()


# Genesets according to pathways:
#
# Pathway information for genes are avaialble from two sources:
#  HTG-Molecular pathway information on HTG-panel in data$`pathway (HTG)`
#  Hallmark pathways in data$`pathway (hallmark)`


g1.immune <- data %>%
  filter(
    str_detect(`pathway (HTG)`, "immuno-oncology")
    | str_detect(
      `pathway (hallmark)`,
      "interferon alpha response|interferon gamma response"
    )
  ) %>%
  select(Genes) %>% pull()


g2.proliferation <- data %>%
  filter(
    str_detect(`pathway (HTG)`, "cell cycle")
    | str_detect(
      `pathway (hallmark)`,
      "E2F targets|G2M checkpoint|mitotic spindle"
    )
  ) %>%
  select(Genes) %>% pull()


g3a.stromalEMT <- data %>%
  filter(str_detect(`pathway (hallmark)`,
      "coagulation|epithelial mesenchymal transition|fatty acid metabolism|myogenesis"
    )
  ) %>%
  select(Genes) %>% pull()


g3b.angiogen <- data %>%
```

```r
  filter(
    str_detect(`pathway (HTG)`, "angiogenesis")
    | str_detect(
      `pathway (hallmark)`,
      "angiogenesis"
    )
  ) %>%
  select(Genes) %>% pull()



# g4.stromal.NonImmune <- g3.stromalEMT[!(g3.stromalEMT %in% g1.immune)]


g5.DNArepair <- data %>%
  filter(
    str_detect(`pathway (HTG)`, "DNA repair")
    | str_detect(
      `pathway (hallmark)`, "DNA repair"
    )
  ) %>%
  select(Genes) %>% pull()


g6.stemcell <- data %>%
  filter(
    str_detect(`pathway (HTG)`, "stem cells")
  ) %>%
  select(Genes) %>% pull()


g7.estrogen.early <- data %>%
  filter(
    str_detect(`pathway (hallmark)`, "estrogen response early")
  ) %>%
  select(Genes) %>% pull()


g8.estrogen.late <- data %>%
  filter(
    str_detect(`pathway (hallmark)`, "estrogen response late")
  ) %>%
  select(Genes) %>% pull()

g9.stress.apopt.hypox <- data %>%
  filter(
    str_detect(`pathway (HTG)`, "apoptosis|hypoxia|stress toxicity")
    | str_detect(
      `pathway (hallmark)`, "apoptosis|hypoxia"
    )
  ) %>%
  select(Genes) %>% pull()

g15.tissueHandl <- c("RGS2", "RASD1", "PER1", "SPRY1", "JUN",
```

```
                        "NR4A1", "EGR1", "DUSP1", "FOS", "SERPINE1",
                        "CYR61", "BTG2", "JUNB", "SLC2A3", "GADD45B")
```

```
# Gene.class assignments ####
```

Gene assignment to unique class ————————————————

```
# Individual genes will be assigned to a unique gene.class based on
# their membership in Genesets (as defined above based on pathway information).
#
# Ranking of unique assignments based on membership in the above genesets:
# a) Tissue handling
# b) DNA repair, stress, hypoxia, apoptosis
# c) estrogen response early
# d) proliferation
# e) immune
# f) stromal-EMT, stem cell , angiogenesis
# g) other (not in any of the above genesets)
#
# These unique assignments will be used for color coding in scatter plots

gene.class <- data %>% select(Genes) %>%
  mutate(gene.class = "other") %>%
  mutate(gene.class = if_else(Genes %in% g6.stemcell, "stromal_EMT", gene.class)) %>%
  mutate(gene.class = if_else(Genes %in% g3a.stromalEMT, "stromal_EMT", gene.class)) %>%
  mutate(gene.class = if_else(Genes %in% g3b.angiogen, "stromal_EMT", gene.class)) %>%
  mutate(gene.class = if_else(Genes %in% g1.immune, "immune", gene.class)) %>%
  mutate(gene.class = if_else(Genes %in% g2.proliferation, "proliferation", gene.class))
%>%
  mutate(gene.class = if_else(Genes %in% g7.estrogen.early, "estrogen_early",
gene.class)) %>%
  mutate(gene.class = if_else(Genes %in% g9.stress.apopt.hypox, "repair_stress",
gene.class)) %>%
  mutate(gene.class = if_else(Genes %in% g5.DNArepair, "repair_stress", gene.class)) %>%
  mutate(gene.class = if_else(Genes %in% g15.tissueHandl, "tissue_handling", gene.class))



# Add gene.class to n335info

n335info <- n335info %>% left_join(gene.class, by = c("Gene" = "Genes"))
```

```r
# Compare gene.class assignment to gene clusters ####


# We compare the frequencies of gene.class assignments within the
# four main gene clusters in the PenelopeB dataset:


crosstable(n335info, Cluster_1080pairedSamples, by=gene.class,
           total = "row",
           percent_digits=1,
           percent_pattern="{n} ({p_row})")  %>%
  as_flextable(compact = TRUE)  %>%
  align(align = "right", part = "body") %>%
  fontsize(size = 7, part = "all") %>%
  width(width=0.8)
```

	gene.class							Total
	estrogen_early	immune	other	proliferation	repair_stress	stromal_EMT	tissue_handling	
Cluster_1080pairedSamples								
1_GoodVsPoor	13 (11.7%)	11 (9.9%)	45 (40.5%)	7 (6.3%)	26 (23.4%)	9 (8.1%)	0 (0%)	111 (33.1%)
2_Proliferation	1 (1.9%)	0 (0%)	9 (17.0%)	34 (64.2%)	9 (17.0%)	0 (0%)	0 (0%)	53 (15.8%)
3_PostTxVsPreTx	1 (2.1%)	11 (23.4%)	23 (48.9%)	0 (0%)	6 (12.8%)	6 (12.8%)	0 (0%)	47 (14.0%)
4_NormalBreast	3 (2.4%)	10 (8.1%)	47 (37.9%)	1 (0.8%)	28 (22.6%)	20 (16.1%)	15 (12.1%)	124 (37.0%)

```r
# We further stratify the cluster '1_GoodVsPoor' into sub-clusters of genes
# and compare the frequencies of gene.class assignments:


crosstable(n335info, SubCluster_1080pairedSamples, by=gene.class,
           total = "row",
           percent_digits=1,
           percent_pattern="{n} ({p_row})")  %>%
  as_flextable(compact = TRUE)  %>%
  align(align = "right", part = "body") %>%
  fontsize(size = 7, part = "all") %>%
  width(width=0.8)
```

	gene.class							Total
	estrogen_early	immune	other	proliferation	repair_stress	stromal_EMT	tissue_handling	
SubCluster_1080pairedSamples								
1_A_IFN	0 (0%)	7 (70.0%)	1 (10.0%)	0 (0%)	2 (20.0%)	0 (0%)	0 (0%)	10 (3.0%)
1_B_RepairStress	0 (0%)	2 (18.2%)	1 (9.1%)	0 (0%)	3 (27.3%)	5 (45.5%)	0 (0%)	11 (3.3%)
1_C_EstrogenResp	12 (42.9%)	2 (7.1%)	11 (39.3%)	1 (3.6%)	2 (7.1%)	0 (0%)	0 (0%)	28 (8.4%)
1_D_RepairStress	1 (1.6%)	0 (0%)	32 (51.6%)	6 (9.7%)	19 (30.6%)	4 (6.5%)	0 (0%)	62 (18.5%)
2_Proliferation	1 (1.9%)	0 (0%)	9 (17.0%)	34 (64.2%)	9 (17.0%)	0 (0%)	0 (0%)	53 (15.8%)
3_PostTxVsPreTx	1 (2.1%)	11 (23.4%)	23 (48.9%)	0 (0%)	6 (12.8%)	6 (12.8%)	0 (0%)	47 (14.0%)
4_NormalBreast	3 (2.4%)	10 (8.1%)	47 (37.9%)	1 (0.8%)	28 (22.6%)	20 (16.1%)	15 (12.1%)	124 (37.0%)

```r
# Analysis / Plots ####


# Add infos from gene clustering


df <- invHRs355 %>% left_join(n335info)

## Joining with `by = join_by(Gene)`

df$quadrant <-dplyr::case_when(df$post.Tx > 1 & df$pre.Tx > 1  ~ "Q1", #Q1... both pre
and post are >1
                               df$post.Tx > 1 & df$pre.Tx <1  ~ "Q2",  #Q2...  pre is <1
and post is >1
                               df$post.Tx < 1 & df$pre.Tx <1  ~ "Q3",  #Q3...  both pre
and post are <1
                               df$pre.Tx > 1 & df$post.Tx <1  ~ "Q4")  #Q4...  pre is >1
and post is <1




# We will generate several plots with different coloring schemes:

# color =
  # 1. Cluster_1080pairedSamples    [only main clusters]
  # 2. gene.class                   [functional annotation]
  # 3. SubCluster_1080pairedSamples [subclusters of cluster1]

# ****************************************
```

```r
# 1.  Coloring: Cluster_1080pairedSamples labels  ####
#  (only the main clusters)

# For colorcoding we will use the following
# 1_GoodVsPoor      #orange
# 2_Proliferation  #deepskyblue
# 3_PostTxVsPreTx  # darkgreen
# 4_NormalBreast   #lightgoldenrod4

# Code:
# scale_color_manual(values=c("orange", "deepskyblue", "darkgreen",
#                             "lightgoldenrod4")) +



scatplot1 <- ggplot(df, aes(x = pre.Tx, y =post.Tx,
                            color = Cluster_1080pairedSamples),
                    guides(fill = FALSE, color = FALSE)) +
  scale_color_manual(values=c("orange", "deepskyblue","darkgreen",
                              "lightgoldenrod4")) +
  geom_point(size=3.5) +
  geom_text_repel(aes(label = Gene),size=3, box.padding = 0.2,
                  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(0.25, 4.7),  n.breaks=7) +
  scale_y_continuous(limits = c(0.5, 1.5),n.breaks=7) +
  theme(plot.title = element_text(hjust = 0.5,vjust = 10,))

# Add annotations

p1 <- scatplot1 +
  annotate("segment", x = -Inf, xend = Inf, y = 1, linewidth=1, yend = 1) +
  annotate("segment", x = 1, xend = 1, y = -Inf, linewidth=1, yend = Inf) +
  annotate("text", x = -Inf, y = 1,
           label = "inverse HR in post.Tx \u2192 improved survival",
           angle = 90,size=4, hjust=0.5, vjust=-1.0,  color = "red") +
  annotate("text", x = 1, y = -Inf,
           label = "inverse HR in pre.Tx \u2192 improved survival",
           hjust=-2.4,vjust=-37, 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) +
```

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

# Place x- and y-axis at "Hazard Ratio = 1",   and include color legend:

p1<-shift_axis_y(p1, y=1)
p1<-shift_axis_x(p1, x=1) +
  labs(color = "Gene clusters")  + theme(legend.position = c(0.9, 0.8))
```

```
## Warning: A numeric `legend.position` argument in `theme()` was deprecated in ggplot2
## 3.5.0.
## ℹ Please use the `legend.position.inside` argument of `theme()` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```r
# Save the plot as svg file

ggsave ("./1_HRscat_BiopResec_335genes_col-mainclust_labs.svg",
        plot=p1, width=14, height=10)
dev.off()
```

```
## null device
##           1
```

```r
# Generate an additional plot without gene names:  ####

# plot WITHOUT gene labels but increased dots
#              and increased text size in legend
#              margins increased to fit to increased legend size

scatplot2 <- ggplot(df, aes(x = pre.Tx, y = post.Tx,
                            color = Cluster_1080pairedSamples),
                      guides(fill = FALSE, color = FALSE)) +
  scale_color_manual(values=c("orange", "deepskyblue","darkgreen",
                              "lightgoldenrod4")) +
  geom_point(size=11) +
  # geom_text_repel(aes(label = Gene),size=3, box.padding = 0.2, max.overlaps=Inf) +
  theme(panel.background = element_rect(fill="white"),
        plot.margin = margin(3, 4, 1, 1, "cm"),axis.title=element_blank(),
        axis.ticks.length=unit(.2, "cm")) +
  scale_x_continuous(limits = c(0.25, 4.7),  n.breaks=7) +
  scale_y_continuous(limits = c(0.5, 1.5),n.breaks=7)


# Add ONLY SELECTED annotations WITH INCREASED TEXT SIZES

p2 <- scatplot2 +
  annotate("segment", x = -Inf, xend = Inf, y = 1, linewidth=1, yend = 1) +
  annotate("segment", x = 1, xend = 1, y = -Inf, linewidth=1, yend = Inf) +
  # annotate("text", x = -Inf, y = 1,
  #          label = "inverse HR in post.Tx \u2192 improved survival",
  #          angle = 90,size=7, hjust=0.5, vjust=-1.0,  color = "red") +
  # annotate("text", x = 1, y = -Inf,
  #          label = "inverse HR in pre.Tx \u2192 improved survival",
  #          hjust=-1.00,vjust=-20, size=7, color = "blue") +
```

```
  # annotate("text", x = -Inf, y = Inf,
  #          label = "good iDFS in post.Tx &\npoor iDFS in pre.Tx",
  #          vjust = 1, hjust=0, size=5) +
  # annotate("text", x = -Inf, y = -Inf,
  #          label = "poor iDFS in post.Tx & pre.Tx",
  #          vjust = 0, hjust=0, size=5) +
  # annotate("text", x = Inf, y = Inf,
  #          label = "good iDFS in post.Tx & pre.Tx",
  #          vjust = 1, hjust=1, size=5) +
  # annotate("text", x = Inf, y = -Inf,
  #          label = "good iDFS in pre.Tx &\npoor iDFS in post.Tx",
  #          vjust = 0, hjust=1, size=5) +
  coord_cartesian(clip = "off")  # Allow annotations to be outside the plot area


# Place x- and y-axis at "Hazard Ratio = 1",
#   and include color legend with increased text size:

p2 <- shift_axis_y(p2, y=1)
p2 <- shift_axis_x(p2, x=1) +
  labs(color = "Gene clusters")  +
  theme(legend.position = c(0.95, 0.9),
        legend.title=element_text(size=28), # increase legend title size
        legend.text=element_text(size=28))  # increase legend text size


ggsave ("./2_HRscat_BiopResec_335genes_col-mainclust_Nolabs.svg",
        plot = p2, width=14, height=10)
dev.off()
```

```
## null device
##           1
```

```
# ************************************************************
```

```r
# 2. Coloring:  gene.class labels ####

  # "estrogen_early",   #limegreen
  # "immune",           #orchid
  # "other",            #grey
  # "proliferation",    #deepskyblue
  # "repair_stress",    #orange
  # "stromal_EMT",      #lightgoldenrod4
  # "tissue_handling"   #yellow

# Code:
  # scale_color_manual(values=c("limegreen", "orchid", "deepskyblue","orange",
  #                             "lightgoldenrod4",  "yellow")) +


# Generate scatter plot

scatplot3 <- ggplot(df, aes(x = pre.Tx, y =post.Tx,
                            color = gene.class),
                    guides(fill = FALSE, color = FALSE)) +
  scale_color_manual(values=c("limegreen", "orchid", "grey", "deepskyblue",
                              "orange", "lightgoldenrod4",  "yellow")) +
  geom_point(size=3.5) +
  geom_text_repel(aes(label = Gene),size=3, box.padding = 0.2,
                  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(0.25, 4.7),  n.breaks=7) +
  scale_y_continuous(limits = c(0.5, 1.5),n.breaks=7) +
  theme(plot.title = element_text(hjust = 0.5,vjust = 10,))

# Add annotations

p3 <- scatplot3 +
  annotate("segment", x = -Inf, xend = Inf, y = 1, linewidth=1, yend = 1) +
  annotate("segment", x = 1, xend = 1, y = -Inf, linewidth=1, yend = Inf) +
  annotate("text", x = -Inf, y = 1,
           label = "inverse HR in post.Tx \u2192 improved survival",
           angle = 90,size=4, hjust=0.5, vjust=-1.0,  color = "red") +
  annotate("text", x = 1, y = -Inf,
           label = "inverse HR in pre.Tx \u2192 improved survival",
           hjust=-2.4,vjust=-37, 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

# Place x- and y-axis at "Hazard Ratio = 1",   and include color legend:

p3<-shift_axis_y(p3, y=1)
p3<-shift_axis_x(p3, x=1) +
  labs(color = "Gene class")  + theme(legend.position = c(0.9, 0.8))

# Save the plot as svg file

ggsave ("./3_HRscat_BiopResec_335genes_col-GeneClass_labs.svg",
        plot=p3, width=14, height=10)
dev.off()

## null device
##           1

# Generate an additional plot without Gene labels:  ####

# plot WITHOUT gene labels but increased dots
#              and increased text size in legend
#              margins increased to fit to increased legend size

scatplot4 <- ggplot(df, aes(x = pre.Tx, y = post.Tx,
                            color = gene.class),
                    guides(fill = FALSE, color = FALSE)) +
  scale_color_manual(values=c("limegreen", "orchid", "grey", "deepskyblue",
                              "orange", "lightgoldenrod4",  "yellow")) +
  geom_point(size=11) +
  # geom_text_repel(aes(label = Gene),size=3, box.padding = 0.2, max.overlaps=Inf) +
  theme(panel.background = element_rect(fill="white"),
        plot.margin = margin(3, 4, 1, 1, "cm"),axis.title=element_blank(),
        axis.ticks.length=unit(.2, "cm")) +
  scale_x_continuous(limits = c(0.25, 4.7),  n.breaks=7) +
  scale_y_continuous(limits = c(0.5, 1.5),n.breaks=7)

# Add ONLY SELECTED annotations WITH INCREASED TEXT SIZES

p4 <- scatplot4 +
  annotate("segment", x = -Inf, xend = Inf, y = 1, linewidth=1, yend = 1) +
  annotate("segment", x = 1, xend = 1, y = -Inf, linewidth=1, yend = Inf) +
  # annotate("text", x = -Inf, y = 1,
  #          label = "inverse HR in post.Tx \u2192 improved survival",
  #          angle = 90,size=7, hjust=0.5, vjust=-1.0,  color = "red") +
  # annotate("text", x = 1, y = -Inf,
  #          label = "inverse HR in pre.Tx \u2192 improved survival",
  #          hjust=-1.00,vjust=-20, size=7, color = "blue") +
  # annotate("text", x = -Inf, y = Inf,
  #          label = "good iDFS in post.Tx &\npoor iDFS in pre.Tx",
  #          vjust = 1, hjust=0, size=5) +
  # annotate("text", x = -Inf, y = -Inf,
  #          label = "poor iDFS in post.Tx & pre.Tx",
  #          vjust = 0, hjust=0, size=5) +
```

```r
  # annotate("text", x = Inf, y = Inf,
  #          label = "good iDFS in post.Tx & pre.Tx",
  #          vjust = 1, hjust=1, size=5) +
  # annotate("text", x = Inf, y = -Inf,
  #          label = "good iDFS in pre.Tx &\npoor iDFS in post.Tx",
  #          vjust = 0, hjust=1, size=5) +
  coord_cartesian(clip = "off")  # Allow annotations to be outside the plot area

# Place x- and y-axis at "Hazard Ratio = 1",
#   and include color legend with increased text size:

p4 <- shift_axis_y(p4, y=1)
p4 <- shift_axis_x(p4, x=1) +
  labs(color = "Gene class")  +
  theme(legend.position = c(0.95, 0.9),
        legend.title=element_text(size=28), # increase legend title size
        legend.text=element_text(size=28))  # increase legend text size

# Save the plot as svg file

ggsave ("./4_HRscat_BiopResec_335genes_col-GeneClass_Nolabs.svg",
        plot=p4, width=14, height=10)
dev.off()

## null device
##           1

# ************************************************************
```

```r
# 3.  Coloring: SubCluster_1080pairedSamples labels  ####

# For colorcoding we will use the following
# 1_A_IFN            #orchid
# 1_B_RepairStress  #orange
# 1_C_EstrogenResp  #limegreen
# 1_D_RepairStress  #chocolate
# 2_Proliferation  #deepskyblue
# 3_PostTxVsPreTx  # darkgreen
# 4_NormalBreast   #lightgoldenrod4

# Code:
# scale_color_manual(values=c("orchid", "orange", "limegreen",
#                             "chocolate", "deepskyblue", "darkgreen",
#                             "lightgoldenrod4")) +


# Generate scatter plot

scatplot5 <- ggplot(df, aes(x = pre.Tx, y =post.Tx,
                            color = SubCluster_1080pairedSamples),
                    guides(fill = FALSE, color = FALSE)) +
  scale_color_manual(values=c("orchid", "orange", "limegreen",
                              "chocolate", "deepskyblue", "darkgreen",
                              "lightgoldenrod4")) +
  geom_point(size=3.5) +
  geom_text_repel(aes(label = Gene),size=3, box.padding = 0.2,
                  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(0.25, 4.7),  n.breaks=7) +
  scale_y_continuous(limits = c(0.5, 1.5),n.breaks=7) +
  theme(plot.title = element_text(hjust = 0.5,vjust = 10,))

# Add annotations

p5 <- scatplot5 +
  annotate("segment", x = -Inf, xend = Inf, y = 1, linewidth=1, yend = 1) +
  annotate("segment", x = 1, xend = 1, y = -Inf, linewidth=1, yend = Inf) +
  annotate("text", x = -Inf, y = 1,
           label = "inverse HR in post.Tx \u2192 improved survival",
           angle = 90,size=4, hjust=0.5, vjust=-1.0,  color = "red") +
  annotate("text", x = 1, y = -Inf,
           label = "inverse HR in pre.Tx \u2192 improved survival",
           hjust=-2.4,vjust=-37, 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

# Place x- and y-axis at "Hazard Ratio = 1",   and include color legend:

p5<-shift_axis_y(p5, y=1)
p5<-shift_axis_x(p5, x=1) +
  labs(color = "Gene clusters")  + theme(legend.position = c(0.9, 0.8))

# Save the plot as svg file

ggsave ("./5_HRscat_BiopResec_335genes_col-subclust_labs.svg",
        plot=p5, width=14, height=10)
dev.off()
```

```
## null device
##           1
```

```
# Generate an additional plot without gene names:  ####

# plot WITHOUT gene labels but increased dots
#              and increased text size in legend
#              margins increased to fit to increased legend size

scatplot6 <- ggplot(df, aes(x = pre.Tx, y = post.Tx,
                            color = SubCluster_1080pairedSamples),
                    guides(fill = FALSE, color = FALSE)) +
  scale_color_manual(values=c("orchid", "orange", "limegreen",
                              "chocolate", "deepskyblue", "darkgreen",
                              "lightgoldenrod4")) +
  geom_point(size=11) +
  # geom_text_repel(aes(label = Gene),size=3, box.padding = 0.2, max.overlaps=Inf) +
  theme(panel.background = element_rect(fill="white"),
        plot.margin = margin(3, 4, 1, 1, "cm"),axis.title=element_blank(),
        axis.ticks.length=unit(.2, "cm")) +
  scale_x_continuous(limits = c(0.25, 4.7),  n.breaks=7) +
  scale_y_continuous(limits = c(0.5, 1.5),n.breaks=7)


# Add ONLY SELECTED annotations WITH INCREASED TEXT SIZES

p6 <- scatplot6 +
  annotate("segment", x = -Inf, xend = Inf, y = 1, linewidth=1, yend = 1) +
  annotate("segment", x = 1, xend = 1, y = -Inf, linewidth=1, yend = Inf) +
  # annotate("text", x = -Inf, y = 1,
  #          label = "inverse HR in post.Tx \u2192 improved survival",
  #          angle = 90,size=7, hjust=0.5, vjust=-1.0,  color = "red") +
  # annotate("text", x = 1, y = -Inf,
  #          label = "inverse HR in pre.Tx \u2192 improved survival",
  #          hjust=-1.00,vjust=-20, size=7, color = "blue") +
  # annotate("text", x = -Inf, y = Inf,
```

```
  #           label = "good iDFS in post.Tx &\npoor iDFS in pre.Tx",
  #           vjust = 1, hjust=0, size=5) +
  # annotate("text", x = -Inf, y = -Inf,
  #           label = "poor iDFS in post.Tx & pre.Tx",
  #           vjust = 0, hjust=0, size=5) +
  # annotate("text", x = Inf, y = Inf,
  #           label = "good iDFS in post.Tx & pre.Tx",
  #           vjust = 1, hjust=1, size=5) +
  # annotate("text", x = Inf, y = -Inf,
  #           label = "good iDFS in pre.Tx &\npoor iDFS in post.Tx",
  #           vjust = 0, hjust=1, size=5) +
  coord_cartesian(clip = "off")  # Allow annotations to be outside the plot area


# Place x- and y-axis at "Hazard Ratio = 1",
#   and include color legend with increased text size:

p6 <- shift_axis_y(p6, y=1)
p6 <- shift_axis_x(p6, x=1) +
  labs(color = "Gene clusters")  +
  theme(legend.position = c(0.95, 0.9),
        legend.title=element_text(size=28), # increase legend title size
        legend.text=element_text(size=28))  # increase legend text size


ggsave ("./6_HRscat_BiopResec_335genes_col-subclust_Nolabs.svg",
        plot = p6, width=14, height=10)
dev.off()
```

```
## null device
##           1
```

```r
# SESSION INFO ####
sessionInfo()
```

```
## R version 4.4.2 (2024-10-31 ucrt)
## Platform: x86_64-w64-mingw32/x64
## Running under: Windows 11 x64 (build 26100)
##
## 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] flextable_0.9.6  crosstable_0.7.0 svglite_2.1.3    readxl_1.4.3
##  [5] ggrepel_0.9.6    lubridate_1.9.3  forcats_1.0.0    stringr_1.5.1
##  [9] dplyr_1.1.4      purrr_1.0.2      readr_2.1.5      tidyr_1.3.1
## [13] tibble_3.2.1     ggplot2_3.5.1    tidyverse_2.0.0
##
## loaded via a namespace (and not attached):
##  [1] gtable_0.3.6           xfun_0.47              tzdb_0.4.0
##  [4] vctrs_0.6.5            tools_4.4.2            generics_0.1.3
##  [7] pkgconfig_2.0.3        data.table_1.16.0      checkmate_2.3.2
## [10] uuid_1.2-1             lifecycle_1.0.4        compiler_4.4.2
## [13] farver_2.1.2           textshaping_0.4.0      munsell_0.5.1
## [16] fontquiver_0.2.1       fontLiberation_0.1.0   htmltools_0.5.8.1
## [19] yaml_2.3.10            pillar_1.10.0          crayon_1.5.3
## [22] openssl_2.2.1          fontBitstreamVera_0.1.1 tidyselect_1.2.1
## [25] zip_2.3.1              digest_0.6.37          stringi_1.8.4
## [28] labeling_0.4.3         fastmap_1.2.0          grid_4.4.2
## [31] colorspace_2.1-1       cli_3.6.3              magrittr_2.0.3
## [34] withr_3.0.2            gdtools_0.4.0          scales_1.3.0
## [37] backports_1.5.0        timechange_0.3.0       rmarkdown_2.28
## [40] officer_0.6.6          cellranger_1.1.0       askpass_1.2.0
## [43] ragg_1.3.2             hms_1.1.3              evaluate_1.0.0
## [46] knitr_1.48             rlang_1.1.4            Rcpp_1.0.13
## [49] glue_1.7.0             xml2_1.3.6             rstudioapi_0.16.0
## [52] R6_2.5.1               systemfonts_1.1.0
```