

2024-08-26-HRscat_BiopResec_335genes_Diff-Colors.R

t

2024-08-26

```
# HEADER ####
#
# Version: 2024-08-26
#
# 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))

## Warning: package 'tidyverse' was built under R version 4.3.1
## Warning: package 'forcats' was built under R version 4.3.1
## Warning: package 'lubridate' was built under R version 4.3.1

## — Attaching core tidyverse packages ————— tidyverse 2.0.0 —
## ✓ dplyr      1.1.2      ✓ readr      2.1.4
## ✓ forcats   1.0.0      ✓ stringr   1.5.0
## ✓ ggplot2    3.4.2      ✓ tibble    3.2.1
## ✓ lubridate  1.9.2      ✓ tidyr     1.3.0
## ✓ purrr     1.0.1
## — 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
```

```
## Warning: package 'ggrepel' was built under R version 4.3.1
## Warning: package 'svglite' was built under R version 4.3.3
## Warning: package 'crosstable' was built under R version 4.3.3
##
## Attaching package: 'crosstable'
##
## The following object is masked from 'package:purrr':
##
##     compact
##
## Warning: package 'flextable' was built under R version 4.3.3
##
## 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 258 genes with significant results for survival
df1 <- read.table("HR_quadrant_plot_biopsy_resec335genes_wo15.txt",
                  header=TRUE, sep='\t')

# Hazard ratios from biopsy and resect samples
# for ALL 335
df2 <- read.table("HR_quadrant_plot_biopsy_resec_ALL335genes.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 available 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 %>%  
  filter(  
    str_detect(`pathway (hallmark)`,  
      "angiogenesis"  )  
  ) %>%  
  select(Genes) %>% pull()
```

```

    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"))
```

```
# 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_earl y	immune	other	proliferation	repair_stress	stromal_EMT	tissue_handli ng	
Cluster_1080 pairedSample s								
1_Good VsPoor	13 (11.7%)	11 (9.9%)	45 (40.5%)	7 (6.3%)	26 (23.4%)	9 (8.1%)	0 (0%)	111 (33.1%)
2_Prolif eration	1 (1.9%)	0 (0%)	9 (17.0%)	34 (64.2%)	9 (17.0%)	0 (0%)	0 (0%)	53 (15.8%)
3_PostT xVsPreT x	1 (2.1%)	11 (23.4%)	23 (48.9%)	0 (0%)	6 (12.8%)	6 (12.8%)	0 (0%)	47 (14.0%)
4_Norm alBreast	3 (2.4%)	10 (8.1%)	47 (37.9%)	1 (0.8%)	28 (22.6%)	20 (16.1%)	15 (12.1%)	124 (37.0%)

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%)


```

# Analysis / Plots #####

# Add infos from gene clustering

df <- df1 %>% left_join(n335info)

df <- df2 %>% 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]

# *****

```

```

# 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, 4.5), n.breaks=7) +
  scale_y_continuous(limits = c(0.6, 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=-1.75, vjust=-34, size=4, color = "blue") +
  annotate("text", x = -Inf, y = Inf,
          label = "good iDFS in post.Tx &\n poor 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 &\n poor 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:

```
p1<-shift_axis_y(p1, y=1)
```

```
## Warning: Removed 2 rows containing missing values (`geom_point()`).
```

```
## Warning: Removed 2 rows containing missing values (`geom_text_repel()`).
```

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

```
## Warning: Removed 2 rows containing missing values (`geom_point()`).
```

```
## Removed 2 rows containing missing values (`geom_text_repel()`).
```

Save the plot as svg file

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

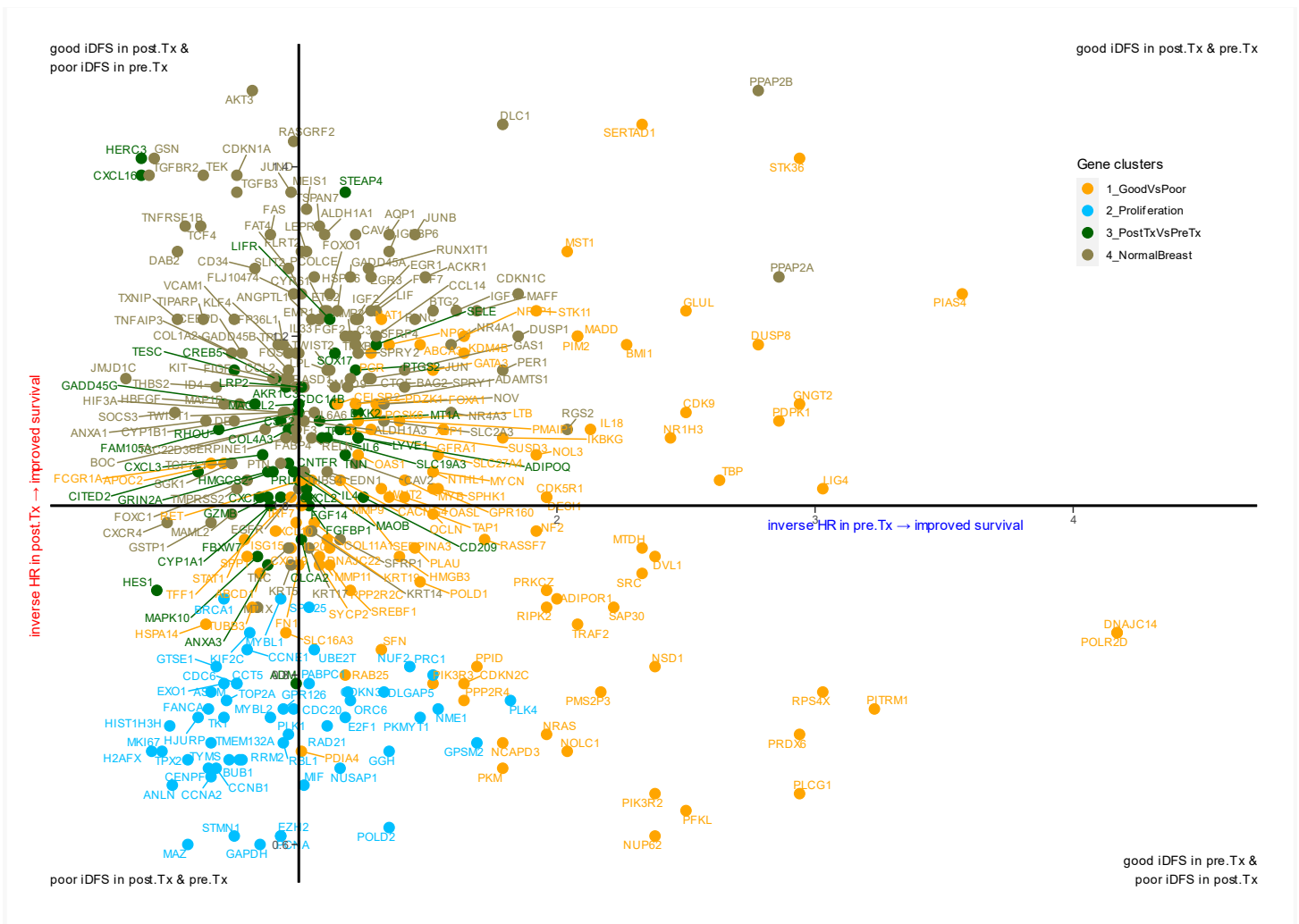
```
## Warning: Removed 2 rows containing missing values (`geom_point()`).
```

```
## Removed 2 rows containing missing values (`geom_text_repel()`).
```

```
dev.off()
```

```
## null device
```

```
##          1
```



Generate an additional plot without gene names:

plot without gene Labels but increased dots

```
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=6) +
  # 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, 4.5), n.breaks=7) +
  scale_y_continuous(limits = c(0.6, 1.5), n.breaks=7) +
  theme(plot.title = element_text(hjust = 0.5, vjust = 10,))
```

Add annotations

```
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=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=-1.75,vjust=-34, 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:

p2 <- shift_axis_y(p2, y=1)

## Warning: Removed 2 rows containing missing values (`geom_point()`).

p2 <- shift_axis_x(p2, x=1) +
  labs(color = "Gene clusters") + theme(legend.position = c(0.9, 0.8))

## Warning: Removed 2 rows containing missing values (`geom_point()`).

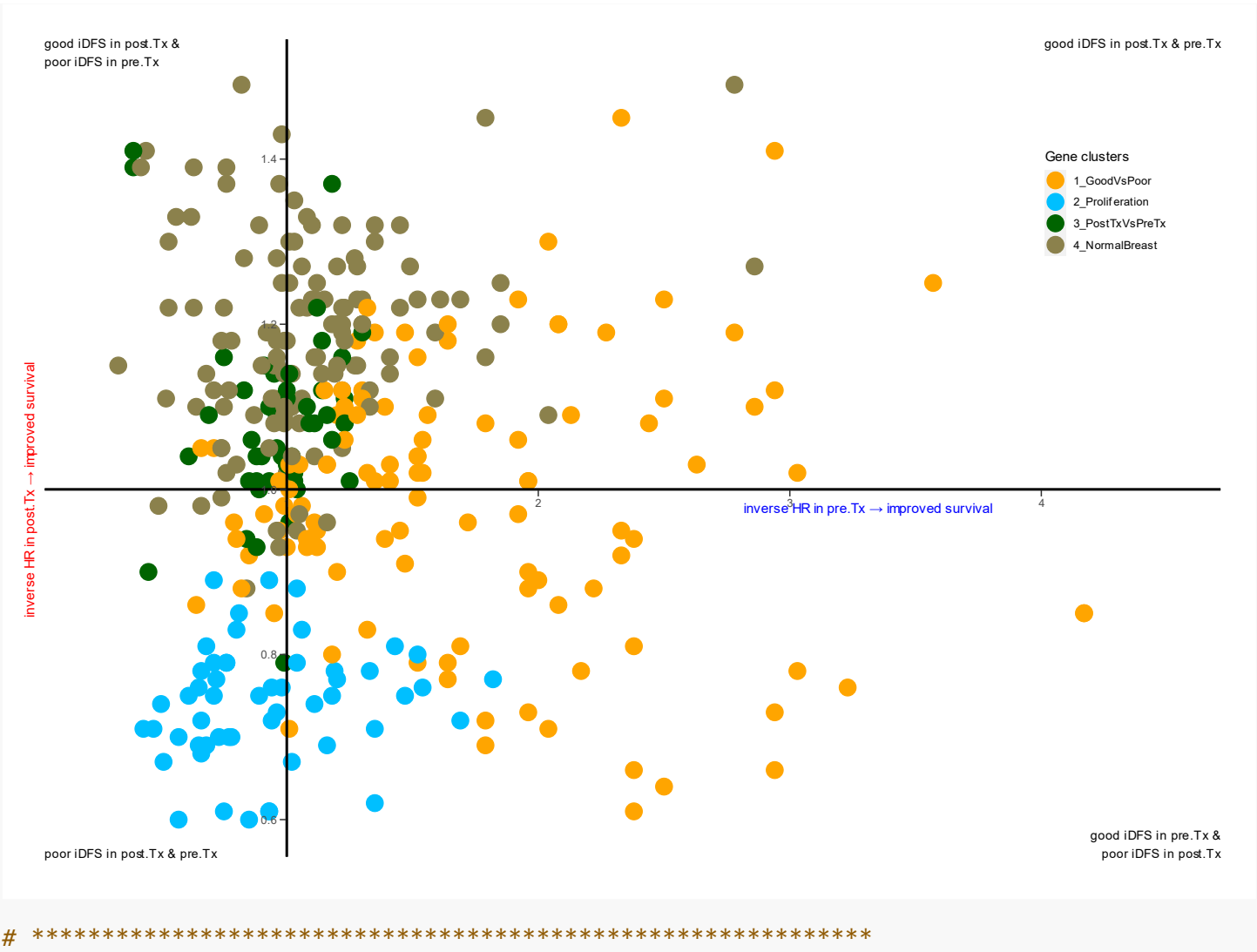
ggsave ("./2_HRscat_BiopResec_335genes_col-mainclust_NoIabs.svg",
        plot = p2, width=14, height=10)

## Warning: Removed 2 rows containing missing values (`geom_point()`).

dev.off()

## null device
##          1

```



```
# 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, 4.5), n.breaks=7) +
  scale_y_continuous(limits = c(0.6, 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=-1.75, vjust=-34, size=4, color = "blue") +
  annotate("text", x = -Inf, y = Inf,
          label = "good iDFS in post.Tx & \n poor 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 & \n poor 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)

## Warning: Removed 2 rows containing missing values (`geom_point()`).
## Warning: Removed 2 rows containing missing values (`geom_text_repel()`).

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

## Warning: Removed 2 rows containing missing values (`geom_point()`).
## Removed 2 rows containing missing values (`geom_text_repel()`).

# Save the plot as svg file

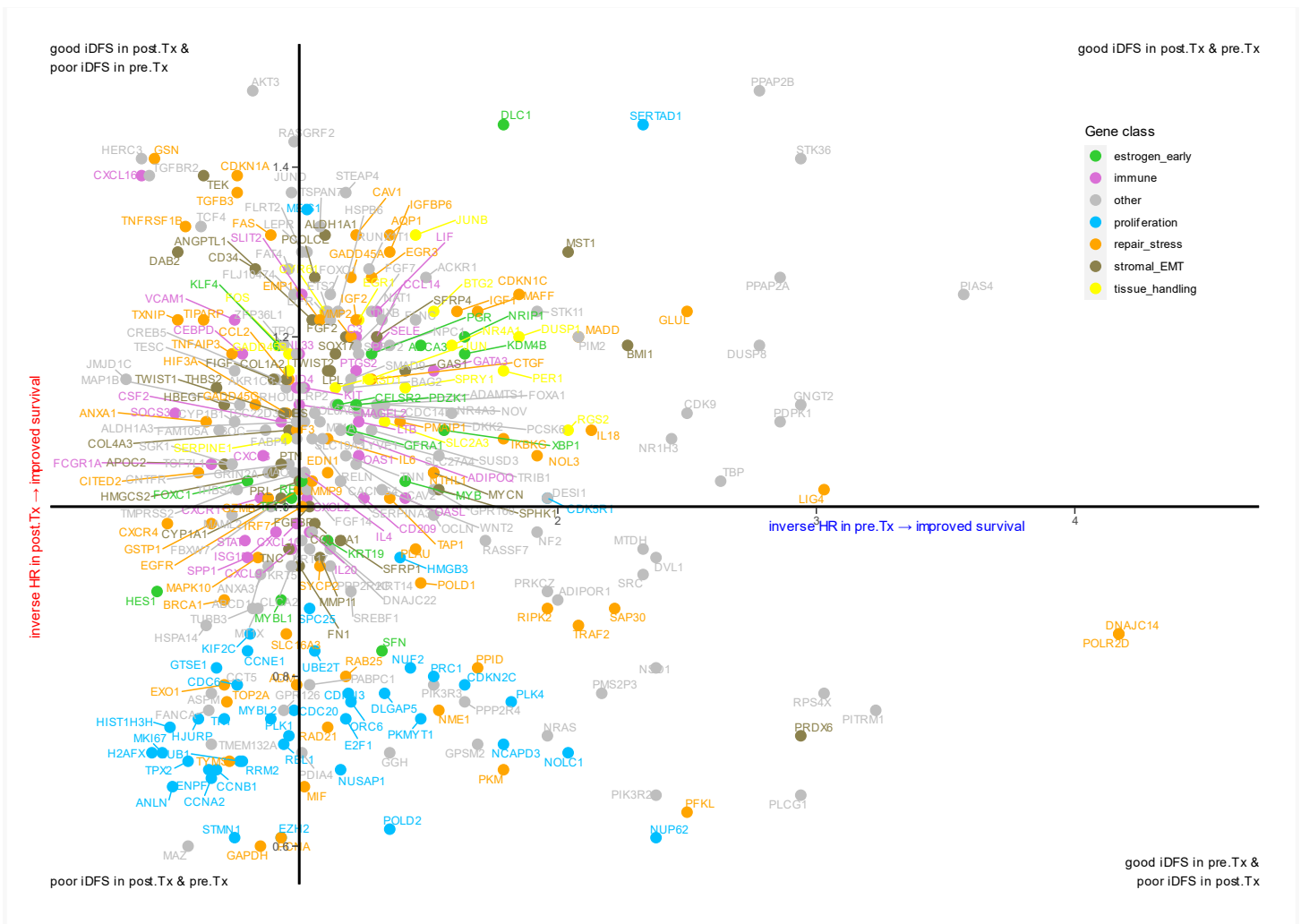
ggsave ("./3_HRscat_BiopResec_335genes_col-GeneClass_labs.svg",
        plot=p3, width=14, height=10)

## Warning: Removed 2 rows containing missing values (`geom_point()`).
## Removed 2 rows containing missing values (`geom_text_repel()`).

dev.off()

## null device
##          1

```

Generate an additional plot without Gene Labels:

plot without gene Labels but increased dots

```
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=6) +
  # 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, 4.5), n.breaks=7) +
  scale_y_continuous(limits = c(0.6, 1.5),n.breaks=7) +
  theme(plot.title = element_text(hjust = 0.5,vjust = 10,))
```

Add annotations

```
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=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=-1.75,vjust=-34, size=4, color = "blue") +
annotate("text", x = -Inf, y = Inf,
      label = "good iDFS in post.Tx &\n poor 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 &\n poor 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:

p4 <- shift_axis_y(p4, y=1)

## Warning: Removed 2 rows containing missing values (`geom_point()`).

p4 <- shift_axis_x(p4, x=1) +
  labs(color = "Gene class") + theme(legend.position = c(0.9, 0.8))

## Warning: Removed 2 rows containing missing values (`geom_point()`).

# Save the plot as svg file

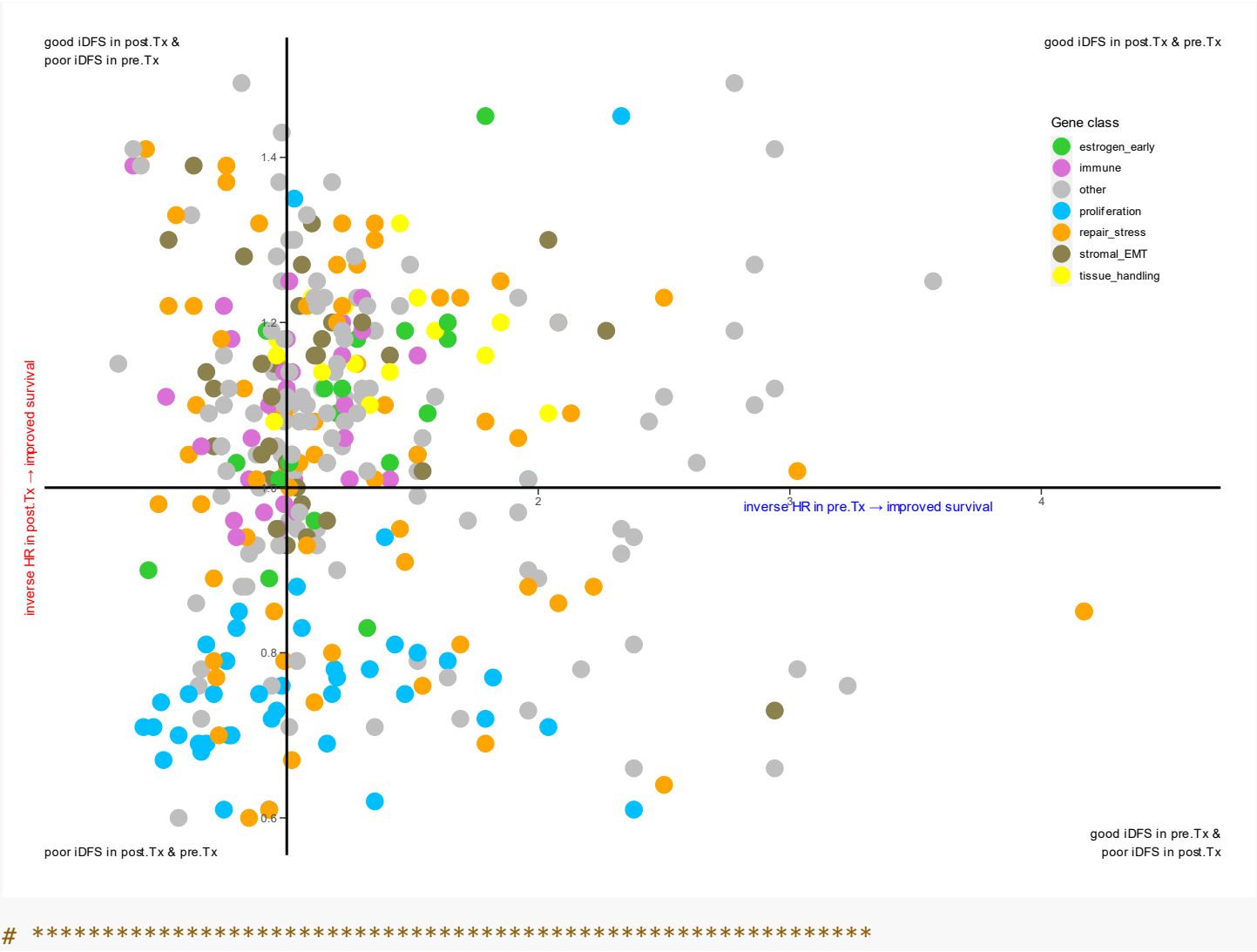
ggsave ("./4_HRscat_BiopResec_335genes_col-GeneClass_NoIabs.svg",
      plot=p4, width=14, height=10)

## Warning: Removed 2 rows containing missing values (`geom_point()`).

dev.off()

## null device
##          1

```



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, 4.5), n.breaks=7) +
  scale_y_continuous(limits = c(0.6, 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=-1.75,vjust=-34, size=4, color = "blue") +
  annotate("text", x = -Inf, y = Inf,
            label = "good iDFS in post.Tx &\n poor 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)

## Warning: Removed 2 rows containing missing values (`geom_point()`).
## Warning: Removed 2 rows containing missing values (`geom_text_repel()`).

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

## Warning: Removed 2 rows containing missing values (`geom_point()`).
## Removed 2 rows containing missing values (`geom_text_repel()`).

# Save the plot as svg file

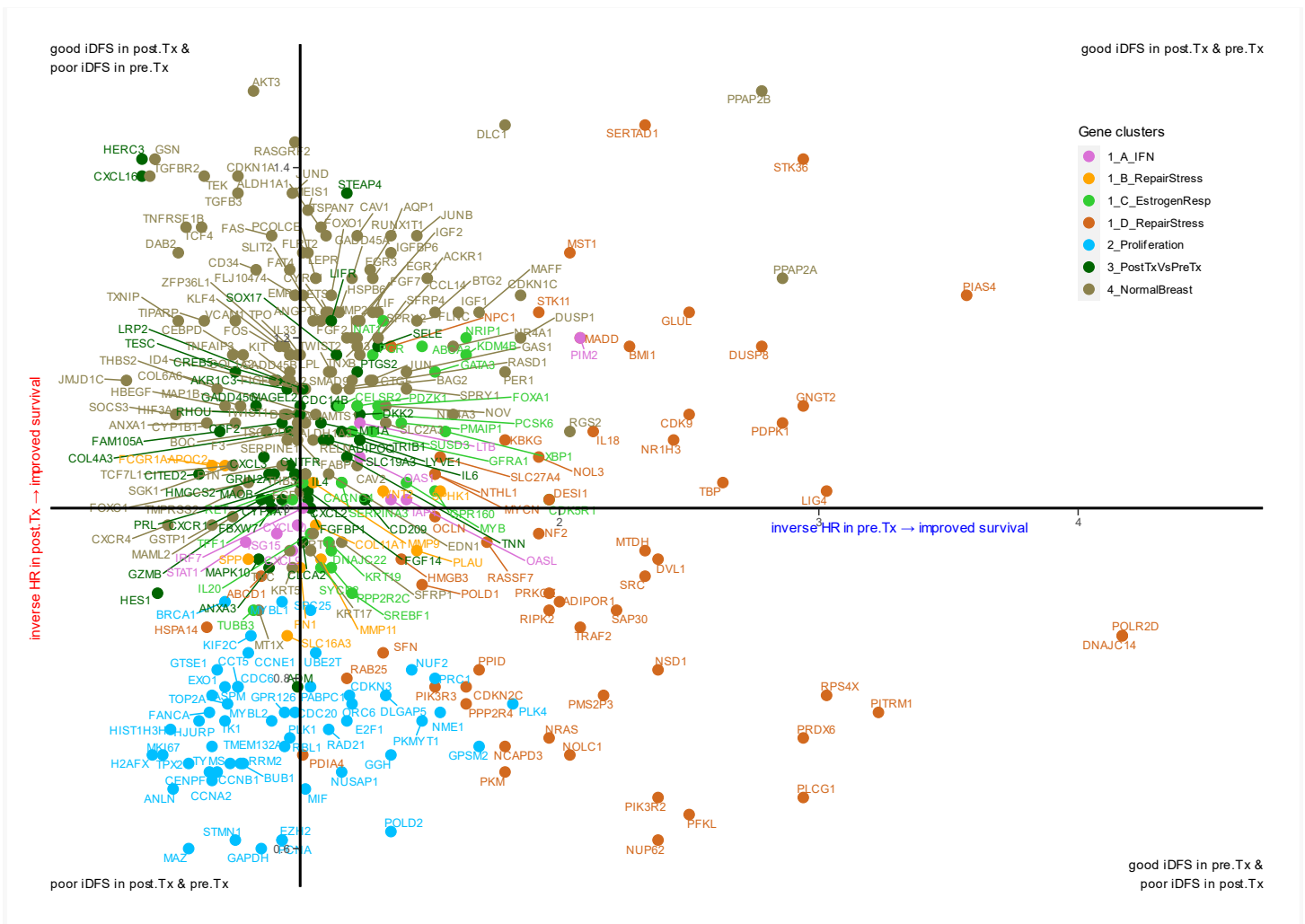
ggsave ("./5_HRscat_BiopResec_335genes_col-subclust_labs.svg",
  plot=p5, width=14, height=10)

## Warning: Removed 2 rows containing missing values (`geom_point()`).
## Removed 2 rows containing missing values (`geom_text_repel()`).

dev.off()

## null device
##      1

```



Generate an additional plot without gene names:

plot without gene labels but increased dots

```
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=6) +
  # 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, 4.5), n.breaks=7) +
  scale_y_continuous(limits = c(0.6, 1.5), n.breaks=7) +
  theme(plot.title = element_text(hjust = 0.5, vjust = 10,))
```

Add annotations

```
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=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=-1.75, vjust=-34, 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:

p6 <- shift_axis_y(p6, y=1)

## Warning: Removed 2 rows containing missing values (`geom_point()`).

p6 <- shift_axis_x(p6, x=1) +
  labs(color = "Gene clusters") + theme(legend.position = c(0.9, 0.8))

## Warning: Removed 2 rows containing missing values (`geom_point()`).

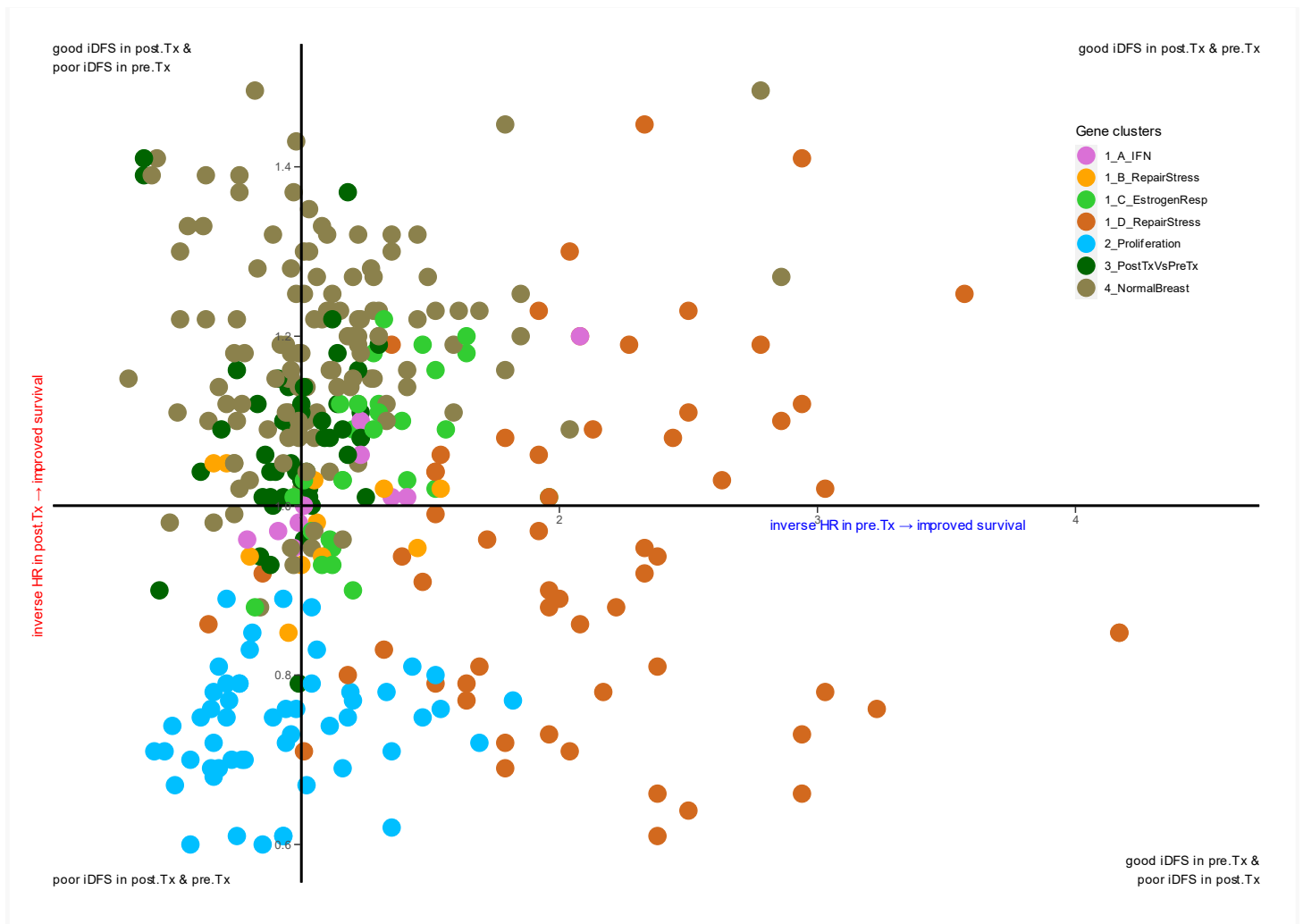
ggsave ("./6_HRscat_BiopResec_335genes_col-subclust_NoIabs.svg",
  plot = p6, width=14, height=10)

## Warning: Removed 2 rows containing missing values (`geom_point()`).

dev.off()

## null device
##      1

```



SESSION INFO

sessionInfo()

```
## R version 4.3.0 (2023-04-21 ucrt)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## 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] flextable_0.9.6  crosstable_0.7.0  svglite_2.1.3    readxl_1.4.2
## [5] ggrepel_0.9.3    lubridate_1.9.2   forcats_1.0.0    stringr_1.5.0
## [9] dplyr_1.1.2      purrr_1.0.1       readr_2.1.4      tidyr_1.3.0
## [13] tibble_3.2.1     ggplot2_3.4.2     tidyverse_2.0.0
##
## loaded via a namespace (and not attached):
## [1] gtable_0.3.3      xfun_0.39          tzdb_0.4.0
## [4] vctr_0.6.2        tools_4.3.0        generics_0.1.3
## [7] curl_5.0.2        fansi_1.0.4        pkgconfig_2.0.3
## [10] data.table_1.14.8 checkmate_2.3.0     uuid_1.1-0
## [13] lifecycle_1.0.3   farver_2.1.1        compiler_4.3.0
## [16] textshaping_0.3.6 munsell_0.5.0       httpuv_1.6.11
## [19] fontquiver_0.2.1  fontLiberation_0.1.0 htmltools_0.5.5
## [22] yaml_2.3.7         pillar_1.9.0        later_1.3.1
## [25] crayon_1.5.2       ellipsis_0.3.2      gfonts_0.2.0
## [28] openssl_2.0.6     mime_0.12           fontBitstreamVera_0.1.1
## [31] zip_2.3.0          tidyselect_1.2.0    digest_0.6.31
## [34] stringi_1.7.12     labeling_0.4.2      fastmap_1.1.1
## [37] grid_4.3.0         colorspace_2.1-0    cli_3.6.1
## [40] magrittr_2.0.3     crul_1.5.0          utf8_1.2.3
## [43] withr_2.5.0        promises_1.2.0.1    gdtools_0.3.7
## [46] scales_1.2.1       backports_1.4.1     timechange_0.2.0
## [49] officer_0.6.6      rmarkdown_2.22      cellranger_1.1.0
## [52] ragg_1.2.5         askpass_1.1         hms_1.1.3
## [55] shiny_1.7.4.1      evaluate_0.21       knitr_1.43
## [58] rlang_1.1.1        Rcpp_1.0.10         xtable_1.8-4
## [61] glue_1.6.2         httpcode_0.3.0      xml2_1.3.5
## [64] rstudioapi_0.15.0  jsonlite_1.8.5      R6_2.5.1
## [67] systemfonts_1.0.4
```