

Normal tissues CTA expression

Léa ROGUE

2025-03-24

List of Figures

1	Heatmap of CTAs expression in normal tissues	5
2	Heatmap of CTAs expression that impact survival in normal tissues	8
3	Scatter plot of CTAs intensities and their expression in normal tissues with clustering colors from n = 63 (log10 scale)	10
4	Scatter plot of CTAs intensities and their expression in normal tissues with clustering colors from n = 63 (linear scale)	11
5	Scatter plot of CTAs intensities and their expression in normal tissues with clustering colors from n = 82 (linear scale)	13
6	Scatter plot of CTAs intensities and their expression in normal tissues with clustering colors from n = 82 (linear scale)	14

Contents

Load libraries	4
Load data	4
I. Expression of all the CTAs in normal tissues	4
II. Expression of CTAs that impact survival probabilities in conventional chondrosarcoma in normal tissues	7
III. Scatter plot of CTAs intensities and their expression in normal tissues.	9

This script generates plots to see the expression of CTAs in normal tissues from GTEx project. This allow us to visualize the most tumor specific CTAs. We added EGFR to compare with an housekeeping genes.

Load libraries

```
library(ComplexHeatmap)
library(dplyr)
library(ggplot2)
library(ggiraph)
library(htmltools)
```

Load data

We use expression matrix generated by script 7, the list of CTA that impact survival probabilities in conventional chondrosarcomas and the intensities data from (E-MTAB-7462).

```
# Load mean expression of CTA in each tissue
df_expr <- read.table("../results/matrix_expr_cta_tissues_gtex.tsv",
  sep = "\t", header = TRUE, row.names = 1)

# Read list of CTA that impact survival probabilities
l_CTA_conv <- read.table("../data/CTA_signif_coxph_conv_indiv.txt",
  sep = "\t", header = FALSE)$V1

# Load intensities data
df_int <- read.table("../results/whole_gene_int_CTA_sign_imm_clean.tsv",
  sep = "\t", row.names = 1, header = TRUE)

# List of housekeeping genes
housekp_genes <- c("EGFR", "ERBB2", "FOLH1", "SSTR1", "SSTR2",
  "SSTR3", "SSTR4", "SSTR5")
```

I. Expression of all the CTAs in normal tissues

This part show that these genes are CTAs are expressed in the testis.

```
# Log10 to scale the data
df_log10 <- log10(t(df_expr) + 1)

# Create heatmap
colors = c("black", "red")
Heatmap(as.matrix(df_log10), cluster_rows = FALSE, cluster_columns = FALSE,
  col = colors, border = NA, show_column_names = TRUE, show_row_names = TRUE,
  column_names_gp = gpar(fontsize = 6), row_names_gp = gpar(fontsize = 2),
  heatmap_legend_param = list(title = "Expression Level"))
```

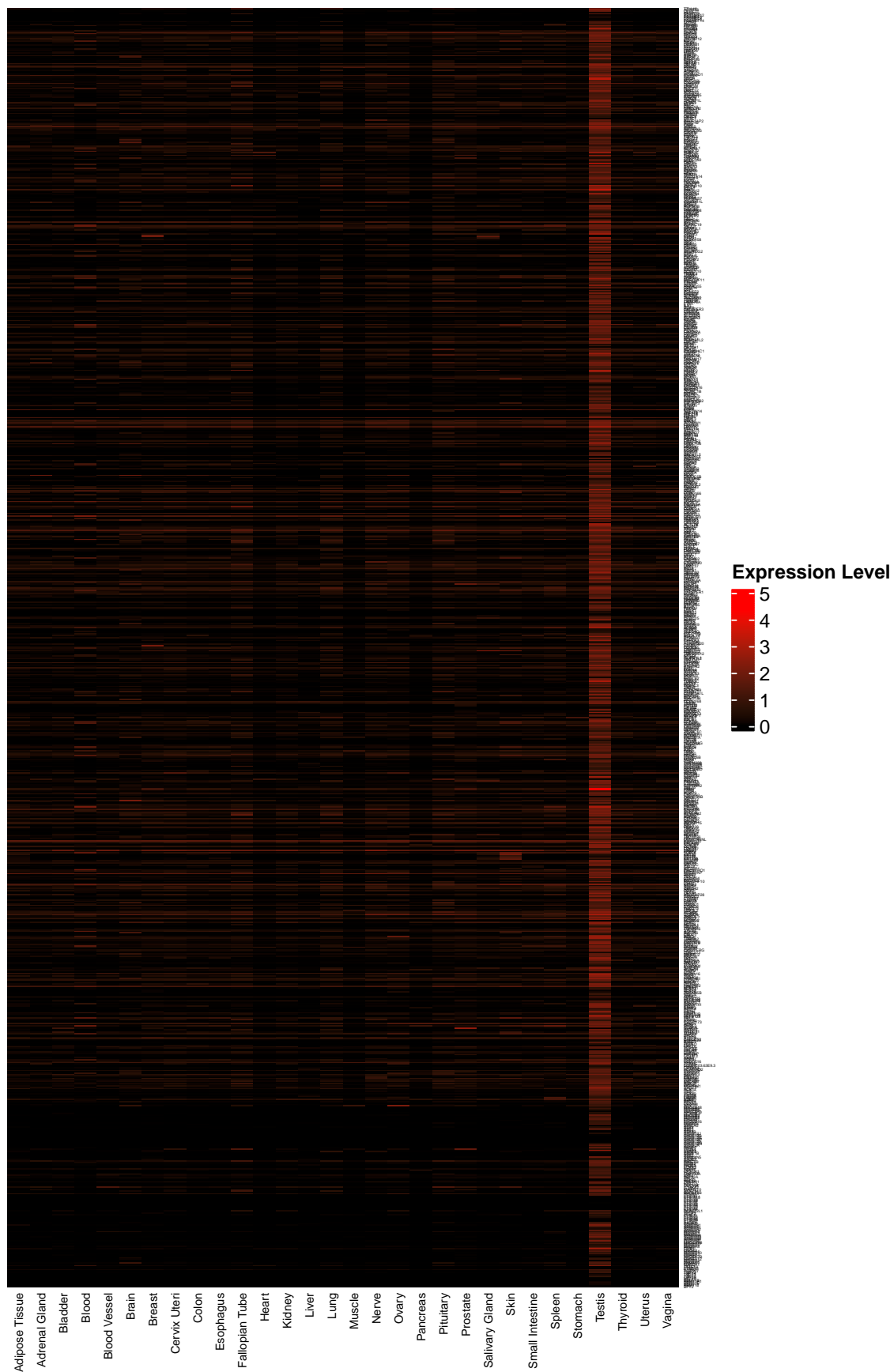


Figure 1: Heatmap of CTAs expression in normal tissues

So, this heatmap shows that genes are generally expressed in the testis and less expressed in other tissues.

II. Expression of CTAs that impact survival probabilities in conventional chondrosarcoma in normal tissues

This section shows the expression of CTAs that impact survival probabilities in conventional chondrosarcoma in normal tissues. This permit to see if these specific CTAs are expressed in normal tissues.

```
# Select CTA
df_log10_conv_CTA <- df_log10[rownames(df_log10) %in% 1_CTA_conv,
]

# Df to have genes and their correspondent color
df_clust_cta_conv <- read.table("../results/clusters_indiv/clusters_cta_signif_conv_indiv.tsv",
  sep = "\t", header = TRUE)
rownames(df_clust_cta_conv) <- df_clust_cta_conv$SYMBOL
df_clust_cta_conv <- df_clust_cta_conv[rownames(df_clust_cta_conv) %in%
  rownames(df_log10_conv_CTA), ]
df_clust_cta_conv$Cluster <- as.character(df_clust_cta_conv$Cluster)

# Colors of clusters
clust_colors <- c(`1` = "#3498DB", `2` = "#E74C3C", `3` = "#2ECC71")

# Heatmap
# pdf('../results/figures/heatmaps/heatmap_expr_cta_conv_gtex.pdf')
Heatmap(as.matrix(df_log10_conv_CTA), cluster_rows = TRUE, cluster_columns = TRUE,
  cluster_column_slices = TRUE, clustering_distance_columns = "euclidean",
  clustering_method_columns = "complete", show_column_dend = TRUE,
  col = colors, column_gap = unit(0, "mm"), row_gap = unit(0,
    "mm"), show_column_names = TRUE, show_row_names = TRUE,
  column_names_gp = gpar(fontsize = 7), row_names_gp = gpar(fontsize = 3),
  right_annotation = rowAnnotation(Cluster = df_clust_cta_conv$Cluster,
    col = list(Cluster = clust_colors)), heatmap_legend_param = list(title = "Expression Level"),
  rect_gp = gpar(col = "transparent", lwd = 0), )

# dev.off()
```

We see that there is some groups that are not really expressed in normal tissues so they could be good biomarkers in chondrosarcomas.

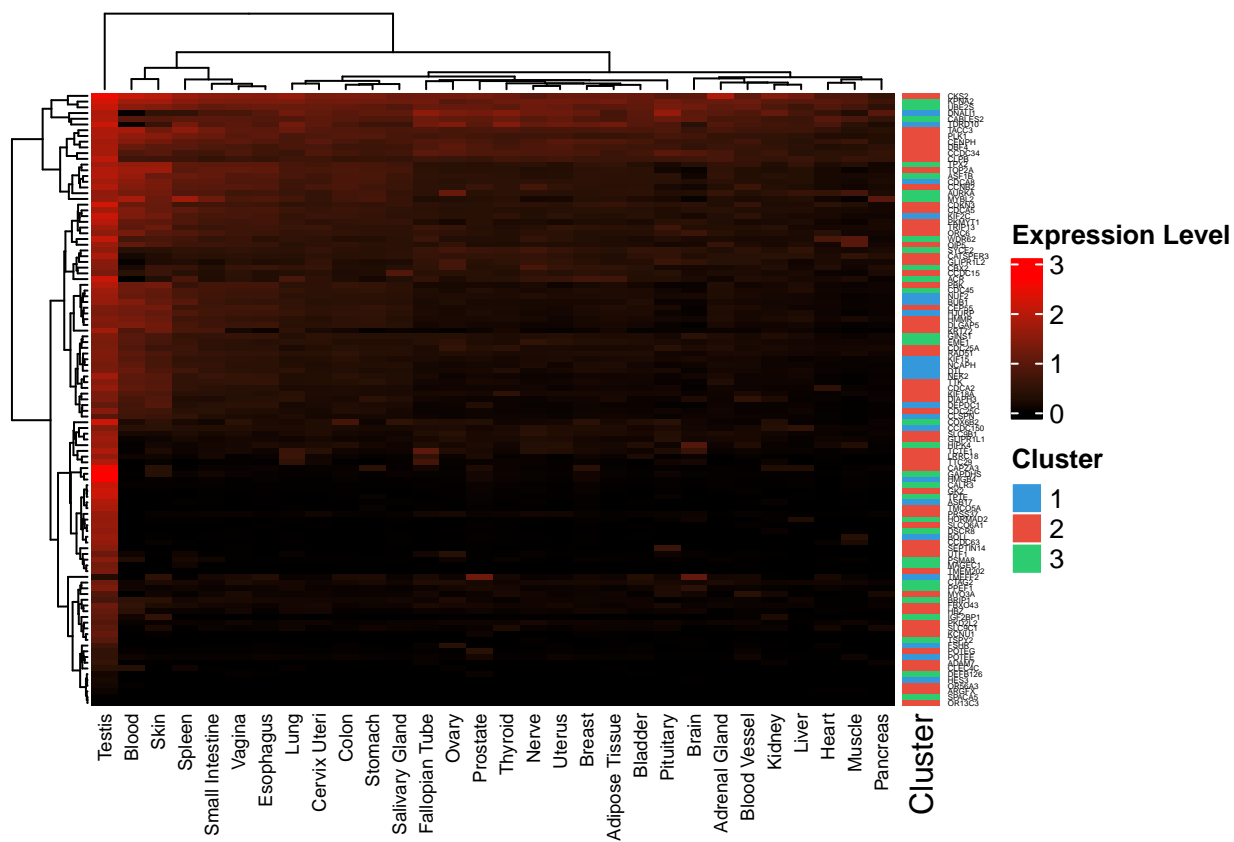


Figure 2: Heatmap of CTAs expression that impact survival in normal tissues

III. Scatter plot of CTAs intensities and their expression in normal tissues.

In this part, we want to see the less expressed CTAs in normal tissues and the more expressed CTAs in chondrosarcomas.

```
# add cluster info
df_clust_cta_all <- read.table("../results/clusters_indiv/clusters_cta_signif_coxph_all_indiv.tsv",
  sep = "\t", header = TRUE)
rownames(df_clust_cta_all) <- df_clust_cta_all$SYMBOL
df_clust_cta_all <- df_clust_cta_all[rownames(df_clust_cta_all) %in%
  rownames(df_log10), ]
df_clust_cta_all$Cluster <- as.character(df_clust_cta_all$Cluster)
colnames(df_clust_cta_all) <- c("Cluster_all", "SYMBOL")

# Select CTA intensities
df_int_cta <- df_int %>%
  filter(CTA != "NA") %>%
  select(-Signature, -CTA)
df_housekp_genes <- df_int[rownames(df_int) %in% housekp_genes,
  ]
df_int_cta <- rbind(df_int_cta, df_housekp_genes[, -c(1, 2)])

# Compute mean for each CTA for all patients
df_mean_int_all <- as.data.frame(rowMeans(df_int_cta))
df_mean_int_all$SYMBOL <- rownames(df_mean_int_all)

# Compute mean for each CTA in all tissues excluding Testis
df <- as.data.frame(t(df_expr[!rownames(df_expr) == "Testis",
  ]))
df_mean_tissue <- as.data.frame(rowMeans(df))
df_mean_tissue$SYMBOL <- rownames(df_mean_tissue)

# Merge data and add cluster information from heatmap from
# script 5
df_scatter <- merge(df_mean_int_all, df_mean_tissue, by = "SYMBOL")
df_scatter <- merge(df_scatter, df_clust_cta_conv, by = "SYMBOL",
  all.x = TRUE)
df_scatter <- merge(df_scatter, df_clust_cta_all, by = "SYMBOL",
  all.x = TRUE)
rownames(df_scatter) <- df_scatter$SYMBOL
df_scatter <- df_scatter[, -1]
colnames(df_scatter) <- c("Intensity", "Mean_expression_tissues",
  "Cluster_conv", "Cluster_all")

# Generate plot
p_log10 <- ggplot(df_scatter, aes(x = Mean_expression_tissues,
  y = Intensity, color = Cluster_conv)) + scale_x_log10() +
  geom_point_interactive(size = 1, hover_nearest = TRUE) +
  scale_color_manual(values = c(`1` = "#3498DB", `2` = "#E74C3C",
    `3` = "#2ECC71"), na.value = "#dadada") + labs(x = "Normal tissues expression (Log10 TPM mean)",
  y = "Chondrosarcoma CTA expression (intensities)", color = "Clustering gene\nimpacting survival\n(cancer)") +
  geom_label_repel_interactive(data = df_scatter[rownames(df_scatter) %in%
```

```
housekp_genes, ], aes(label = rownames(df_scatter)[rownames(df_scatter) %in%
housekp_genes]), nudge_x = 0.8, nudge_y = -0.3, size = 3,
color = "black") + theme_minimal()
```

Le chargement a nécessité le package : ggrepel

p_log10

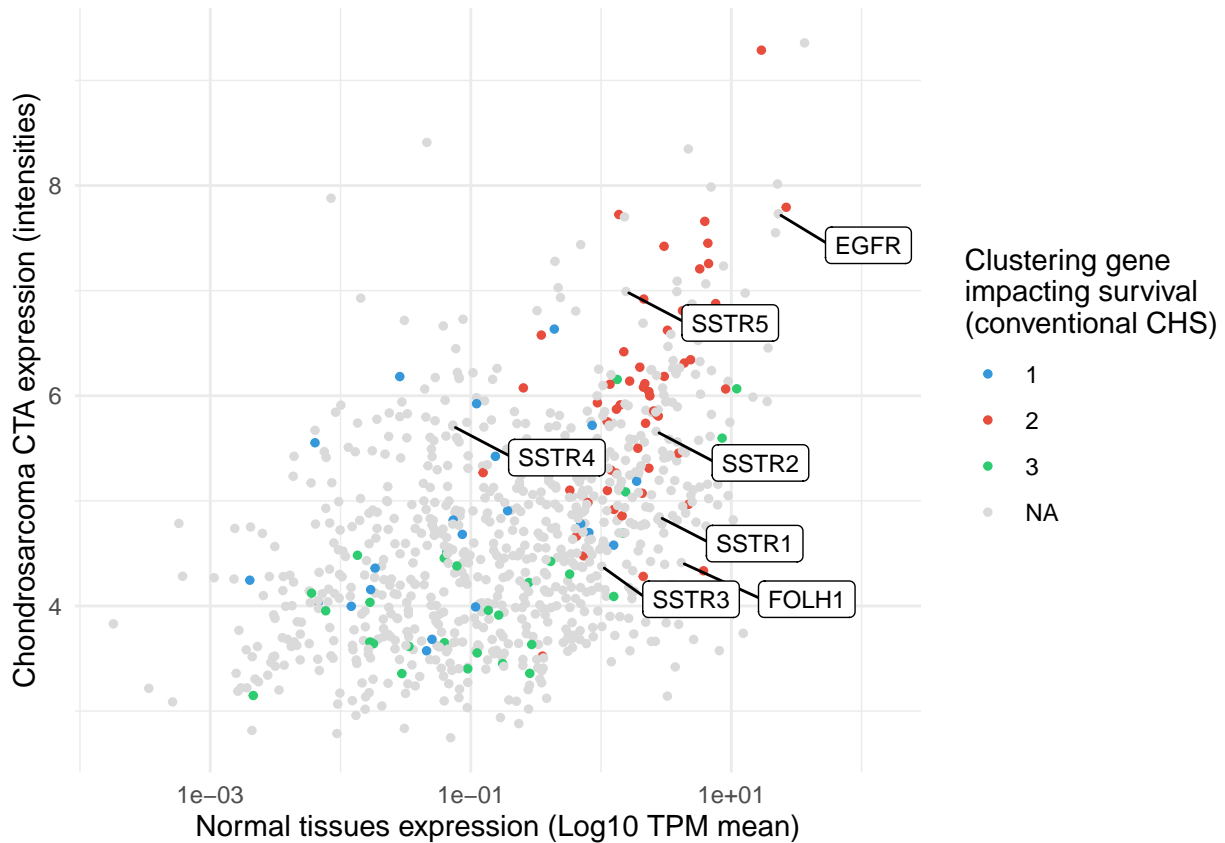


Figure 3: Scatter plot of CTAs intensities and their expression in normal tissues with clustering colors from n = 63 (log10 scale)

```
p <- ggplot(df_scatter, aes(x = Mean_expression_tissues, y = Intensity,
color = Cluster_conv)) + geom_point_interactive(size = 1,
hover_nearest = TRUE) + scale_color_manual(values = c(`1` = "#3498DB",
`2` = "#E74C3C", `3` = "#2ECC71"), na.value = "#dadada") +
labs(x = "Normal tissues expression (Log10 TPM mean)", y = "Chondrosarcoma CTA expression (intensit.
color = "Clustering gene\nimpacting survival\n(conventional CHS)") +
geom_label_repel_interactive(data = df_scatter[rownames(df_scatter) %in%
housekp_genes, ], aes(label = rownames(df_scatter)[rownames(df_scatter) %in%
housekp_genes]), nudge_x = 4, nudge_y = -0.5, size = 3,
color = "black") + theme_minimal()
```

p

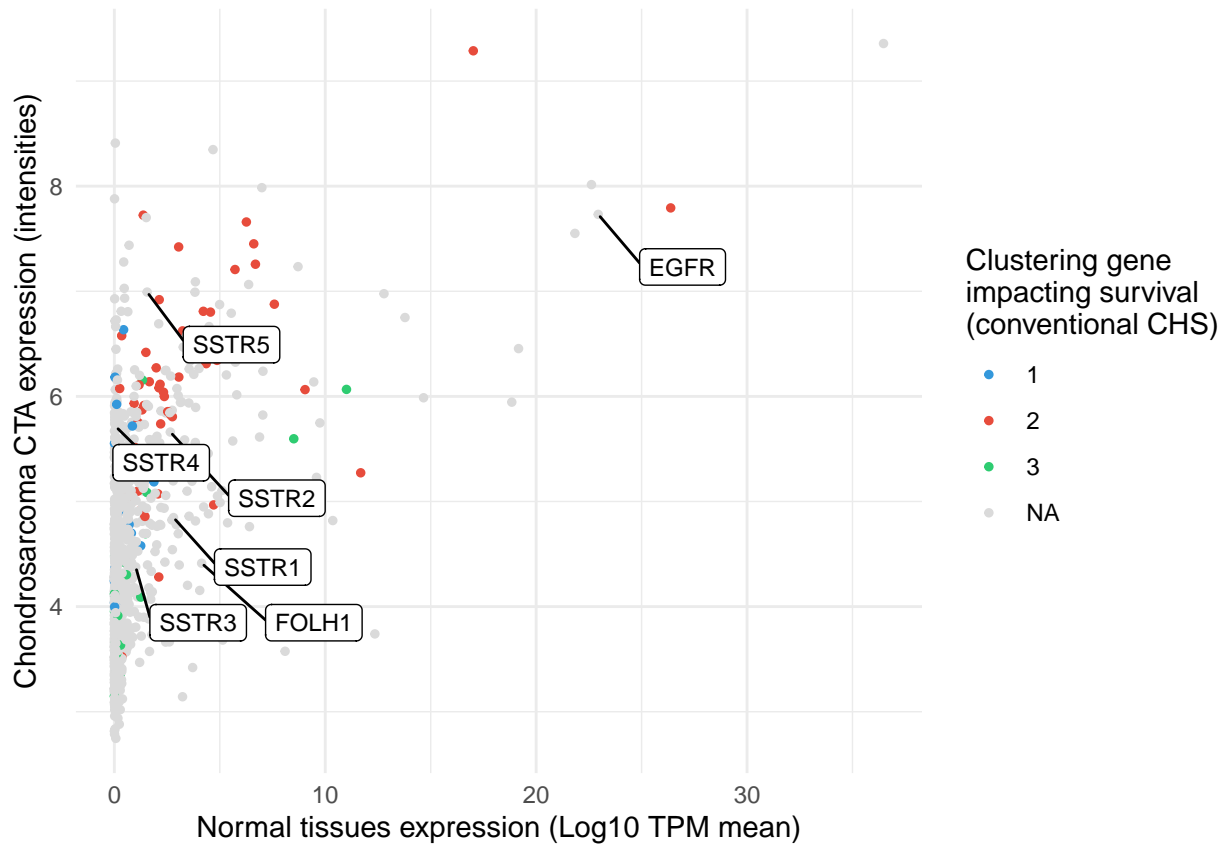


Figure 4: Scatter plot of CTAs intensities and their expression in normal tissues with clustering colors from $n = 63$ (linear scale)

```

# Generate html file with interactive plot
interactive_plot_log10 <- girafe(ggobj = p_log10)
interactive_plot_linear <- girafe(ggobj = p)
html_content <- tagList(div(style = "display: flex; justify-content: space-between;",
  div(style = "width: 48%; ", interactive_plot_log10), div(style = "width: 48%; ",
    interactive_plot_linear)))
save_html(html_content, "../results/figures/other_plots/interactive_scatter_plot_expr_cta_gtex_conv_clu

```

So, the interactive plot and the data permit us to select CTAs of interest in order to probe them in biological manipulations.

```
# Generate plot
p_log10 <- ggplot(df_scatter, aes(x = Mean_expression_tissues,
  y = Intensity, color = Cluster_all)) + scale_x_log10() +
  geom_point_interactive(size = 1, hover_nearest = TRUE) +
  scale_color_manual(values = c(`1` = "#3498DB", `2` = "#E74C3C",
    `3` = "#2ECC71"), na.value = "#dadada") + labs(x = "Normal tissues expression (Log10 TPM mean)",
  y = "Chondrosarcoma CTA expression (intensities)", color = "Clustering gene\nimpacting survival\n(a
  geom_label_repel_interactive(data = df_scatter[rownames(df_scatter) %in%
    housekp_genes, ], aes(label = rownames(df_scatter)[rownames(df_scatter) %in%
    housekp_genes]), nudge_x = 0.8, nudge_y = -0.3, size = 3,
    color = "black") + theme_minimal()
p_log10
```

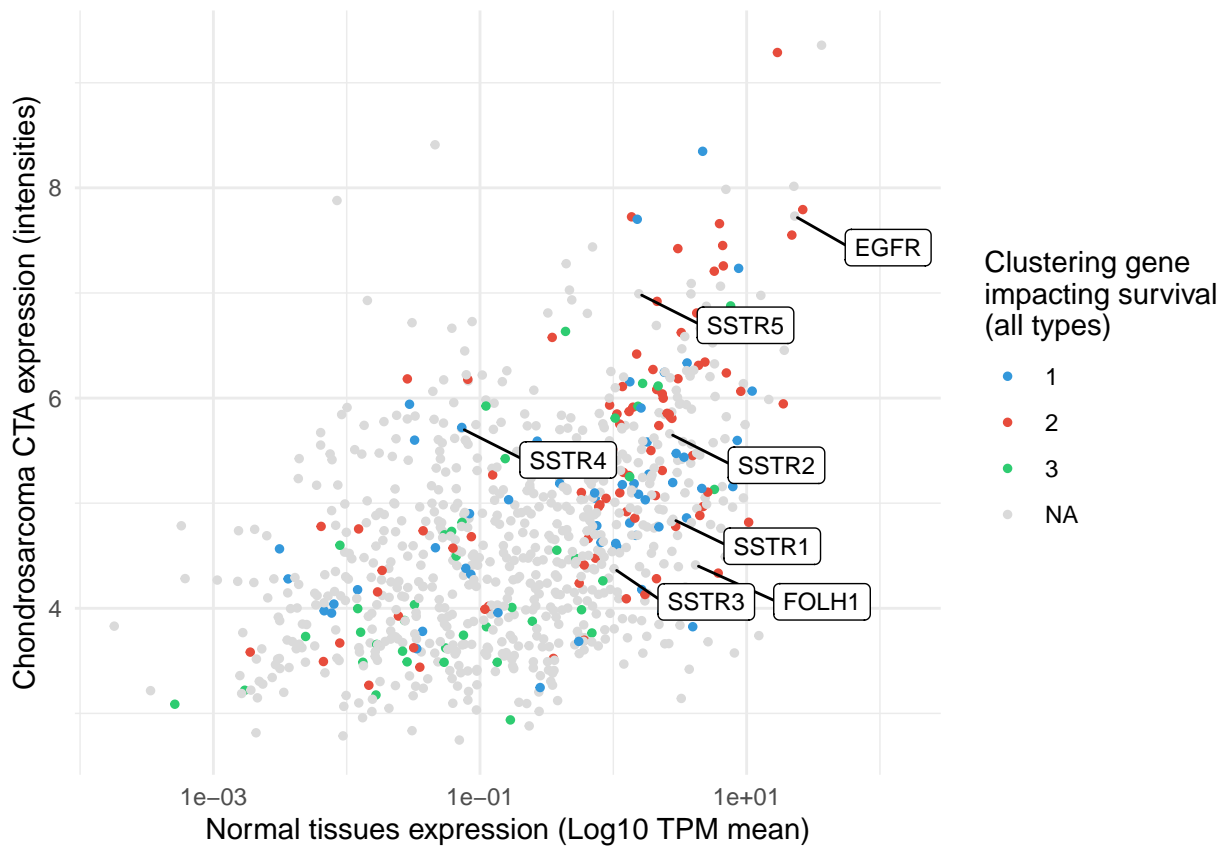


Figure 5: Scatter plot of CTAs intensities and their expression in normal tissues with clustering colors from $n = 82$ (linear scale)

```
p <- ggplot(df_scatter, aes(x = Mean_expression_tissues, y = Intensity,
  color = Cluster_all)) + geom_point_interactive(size = 1,
  hover_nearest = TRUE) + scale_color_manual(values = c(`1` = "#3498DB",
    `2` = "#E74C3C", `3` = "#2ECC71"), na.value = "#dadada") +
  labs(x = "Normal tissues expression (Log10 TPM mean)", y = "Chondrosarcoma CTA expression (intensit
    color = "Clustering gene\nimpacting survival\n(all types)") +
  geom_label_repel_interactive(data = df_scatter[rownames(df_scatter) %in%
    housekp_genes, ], aes(label = rownames(df_scatter)[rownames(df_scatter) %in%
    housekp_genes]), nudge_x = 4, nudge_y = -0.5, size = 3,
```

```
color = "black") + theme_minimal()
p
```

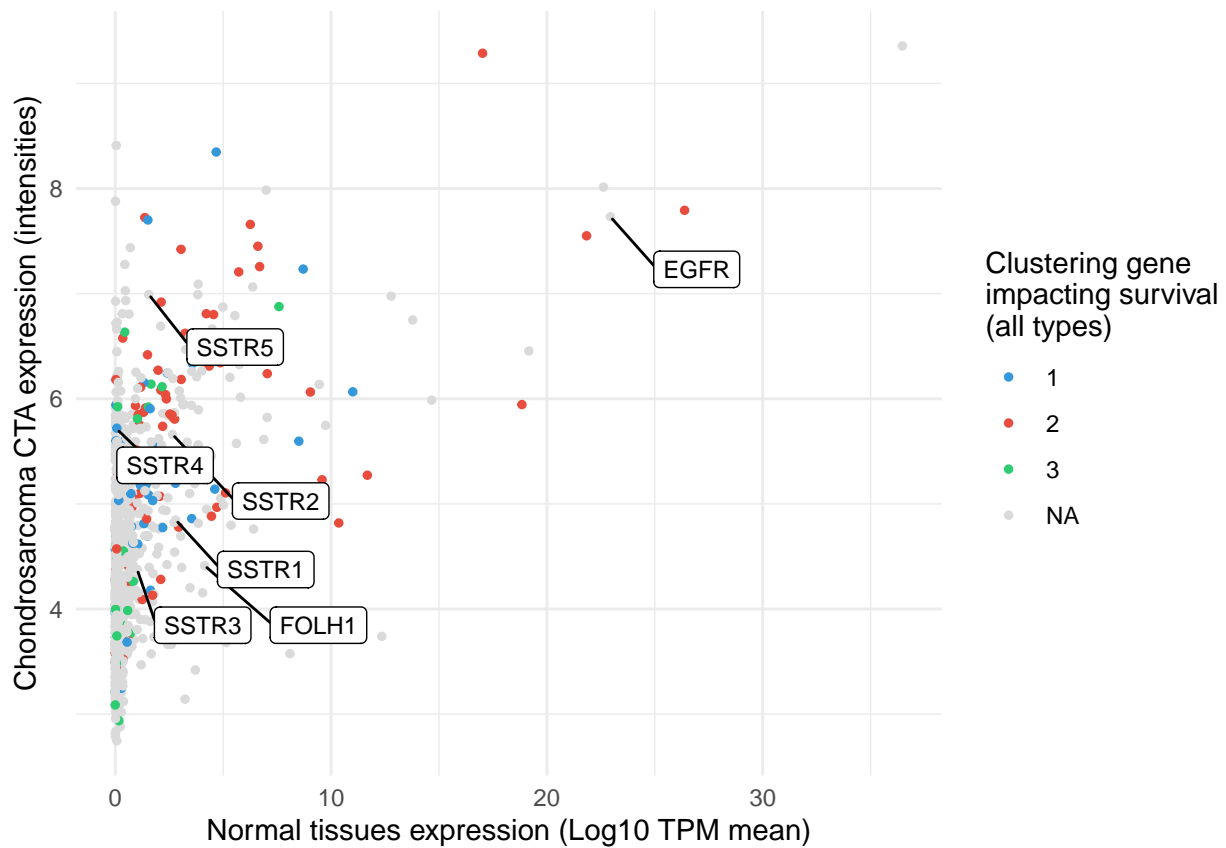


Figure 6: Scatter plot of CTAs intensities and their expression in normal tissues with clustering colors from $n = 82$ (linear scale)

```
# Generate html file with interactive plot
interactive_plot_log10 <- girafe(ggobj = p_log10)
interactive_plot_linear <- girafe(ggobj = p)
html_content <- tagList(div(style = "display: flex; justify-content: space-between;",
  div(style = "width: 48%; ", interactive_plot_log10), div(style = "width: 48%; ",
    interactive_plot_linear)))
save_html(html_content, "../results/figures/other_plots/interactive_scatter_plot_expr_cta_gtex_all_clus")
```