Normal tissues and osteosarcoma (GSE21257) CTA expression

Léa ROGUE

2025-09-05

List of Figures

Contents

This script analyzes CTA expression in osteosarcoma and compares it to normal expression levels.

Load libraries

```r
library(dplyr)
library(tidyr)
library(ComplexHeatmap)
library(circlize)
library(colorRamp2)
library(RColorBrewer)
library(ggplot2)
library(plotly)
```

Load data

```r
# Read HR
df_hr <- read.table("../results/results_coxph_osteo_cta_zscore_signif.tsv",
    sep = "\t", header = T)
df_hr$SYMBOL <- df_hr$Variable

# Read CTA
df_expr_complete <- read.table("../results/expr_matrix_int_norm_osteo_GSE21257.tsv",
    sep = "\t", header = T)
rownames(df_expr_complete) <- df_expr_complete$SYMBOL
df_cta_osteo <- df_expr_complete %>%
    filter(CTA != "NA") %>%
    select(-SYMBOL, -CTA, -Signature)
```

```r
# Read CTA
l_cta <- read.table("../../data/CTA_list_clean.txt", header = F)$V1
housekp_genes <- c("EGFR", "ERBB2", "FOLH1", "SSTR1", "SSTR2",
    "SSTR3", "SSTR4", "SSTR5")
l_cta_validated <- read.table("../../data/CTA_validated_list.txt",
    sep = "\t", header = FALSE)$V1

# Read HPA data consensus and select CTA
df_hpa_raw <- read.table("../../data/rna_tissue_consensus.tsv",
    sep = "\t", header = T)
df_hpa_cta <- df_hpa_raw[df_hpa_raw$Gene.name %in% l_cta | df_hpa_raw$Gene.name %in%
    housekp_genes, ]

# any(is.na(df_hpa_cta))

df_hpa_cta <- as.data.frame(df_hpa_cta %>%
    pivot_wider(id_cols = Tissue, names_from = Gene.name, values_from = nTPM,
        values_fn = first, values_fill = 0))

rownames(df_hpa_cta) <- df_hpa_cta$Tissue
```

```r
df_hpa_cta <- df_hpa_cta[, -1]

# any(is.na(df_hpa_cta))
```

I. CTA expression in healthy tissues

```r
# Log10 to scale the data
df_log10 <- log10(df_hpa_cta + 1)

# CTA annotation
df_cta_validated <- df_log10[, colnames(df_log10) %in% l_cta_validated]
df_cta_putatives <- df_log10[, !colnames(df_log10) %in% l_cta_validated]

# Create heatmaps
colors = colorRamp2(c(min(df_log10), max(df_log10)), c("black",
    "red"))
ht_validated <- Heatmap(as.matrix(df_cta_validated), cluster_rows = FALSE,
    cluster_columns = FALSE, col = colors, border = NA, show_column_names = TRUE,
    show_row_names = TRUE, column_title = "Validated CTAs", column_names_gp = gpar(fontsize = 1),
    row_names_gp = gpar(fontsize = 10), heatmap_legend_param = list(title = "Expression"),
    rect_gp = gpar(col = NA))

ht_putatives <- Heatmap(as.matrix(df_cta_putatives), cluster_rows = FALSE,
    cluster_columns = FALSE, col = colors, show_column_names = TRUE,
    show_row_names = TRUE, column_title = "Putatives CTAs", column_names_gp = gpar(fontsize = 1),
    row_names_gp = gpar(fontsize = 10), rect_gp = gpar(col = NA))
# png('../results/figures/heatmaps/heatmap_cta_normal_tissues.png',
# width = 3000, height = 1800, res = 300)
draw(ht_validated + ht_putatives, merge_legend = TRUE)
```

```r
# dev.off()
```

II. CTA expression in normal tissues and osteosarcoma

```r
# Read immunophenotype
df_clust_cta <- read.table("../results/clusters_cta_pearson_53_k2.tsv",
    sep = "\t", header = T)
df_clust_cta$SYMBOL <- rownames(df_clust_cta)

# Compute means
df_mean_int_all <- as.data.frame(rowMeans(df_cta_osteo))
colnames(df_mean_int_all) <- "Intensity"
df_mean_int_all$SYMBOL <- rownames(df_mean_int_all)

df <- as.data.frame(t(df_hpa_cta[!rownames(df_hpa_cta) == "testis",
    ]))
df_mean_tissue <- as.data.frame(rowMeans(df))
```
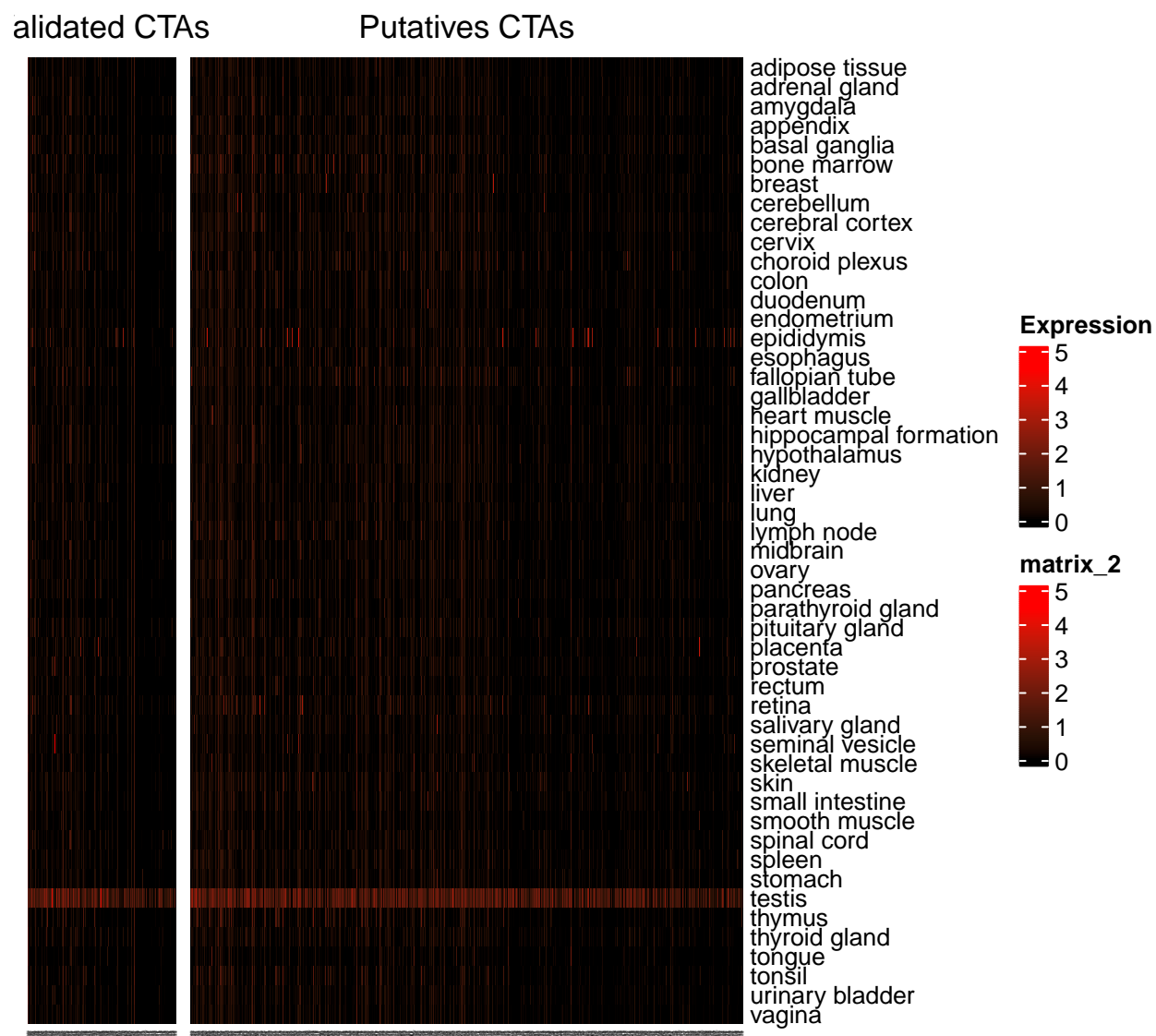
Figure 1: Heatmap of CTAs expression in normal tissues

```r
df_mean_tissue$SYMBOL <- rownames(df_mean_tissue)

# Merge data
df_scatter <- merge(df_mean_int_all, df_mean_tissue, by = "SYMBOL")
df_scatter <- merge(df_scatter, df_clust_cta, by = "SYMBOL",
    all.x = TRUE)

# Clean and rename
rownames(df_scatter) <- df_scatter$SYMBOL
colnames(df_scatter) <- c("SYMBOL", "Intensity", "Mean_expression_tissues",
    "num_Cluster", "Associated_immunophenotype")

# Add CTA cluster
df_cta_ht <- read.table("../results/clusters_cta_heatmap_signif_coxph_osteo_53.tsv",
    header = TRUE)
df_cta_ht$SYMBOL <- rownames(df_cta_ht)
df_scatter <- merge(df_scatter, df_cta_ht, by = "SYMBOL", all.x = TRUE)

# Ajouter HR proprement (une seule fois)
df_scatter <- merge(df_scatter, df_hr[, c("SYMBOL", "HR")], by = "SYMBOL",
    all.x = TRUE)

# Add HR value to add point size
df_scatter$HR <- ifelse(is.na(df_scatter$HR), 0.1, df_scatter$HR)

# Prepare the data
rownames(df_scatter) <- df_scatter$SYMBOL
df_scatter$row_clusters <- as.character(df_scatter$row_clusters)
folh1_expr <- mean(df_hpa_cta[, colnames(df_hpa_cta) %in% "FOLH1"])

# Generate plot
ggplot(df_scatter, aes(x = Mean_expression_tissues, y = Intensity,
    color = row_clusters)) + geom_point(size = 1) + geom_vline(xintercept = folh1_expr,
    linetype = "dashed", color = "black") + geom_hline(yintercept = median(df_scatter$Intensity),
    linetype = "dashed", color = "black") + annotate("text",
    x = folh1_expr - 3.1, y = max(df_scatter$Intensity), label = "FOLH1 expression",
    size = 2.5) + annotate("text", x = max(df_scatter$Mean_expression_tissues) -
    2.6, y = median(df_scatter$Intensity) + 0.4, label = "Median\nexpression",
    size = 2.5) + scale_color_manual(values = c(`1` = "#dadada",
    `2` = "#dadada"), na.value = "#dadada") + theme_minimal() +
    scale_x_log10() + labs(title = "Scatter plot of tumor specificity of CTAs (n = 53)",
    x = "Normal tissues expression (Log10 TPM mean)", y = "Chondrosarcoma CTA expression (intensities)",
    color = "Clustering gene\nimpacting survival")
```

```
## Warning in scale_x_log10(): log-10 transformation introduced infinite values.
```

```r
# Generate plot
ggplot(df_scatter, aes(x = Mean_expression_tissues, y = Intensity,
    color = row_clusters, size = HR, shape = Associated_immunophenotype)) +
    geom_point(alpha = 0.8) + geom_vline(xintercept = 4.139,
    linetype = "dashed", color = "black") + geom_hline(yintercept = median(df_scatter$Intensity),
    linetype = "dashed", color = "black") + annotate("text",
```
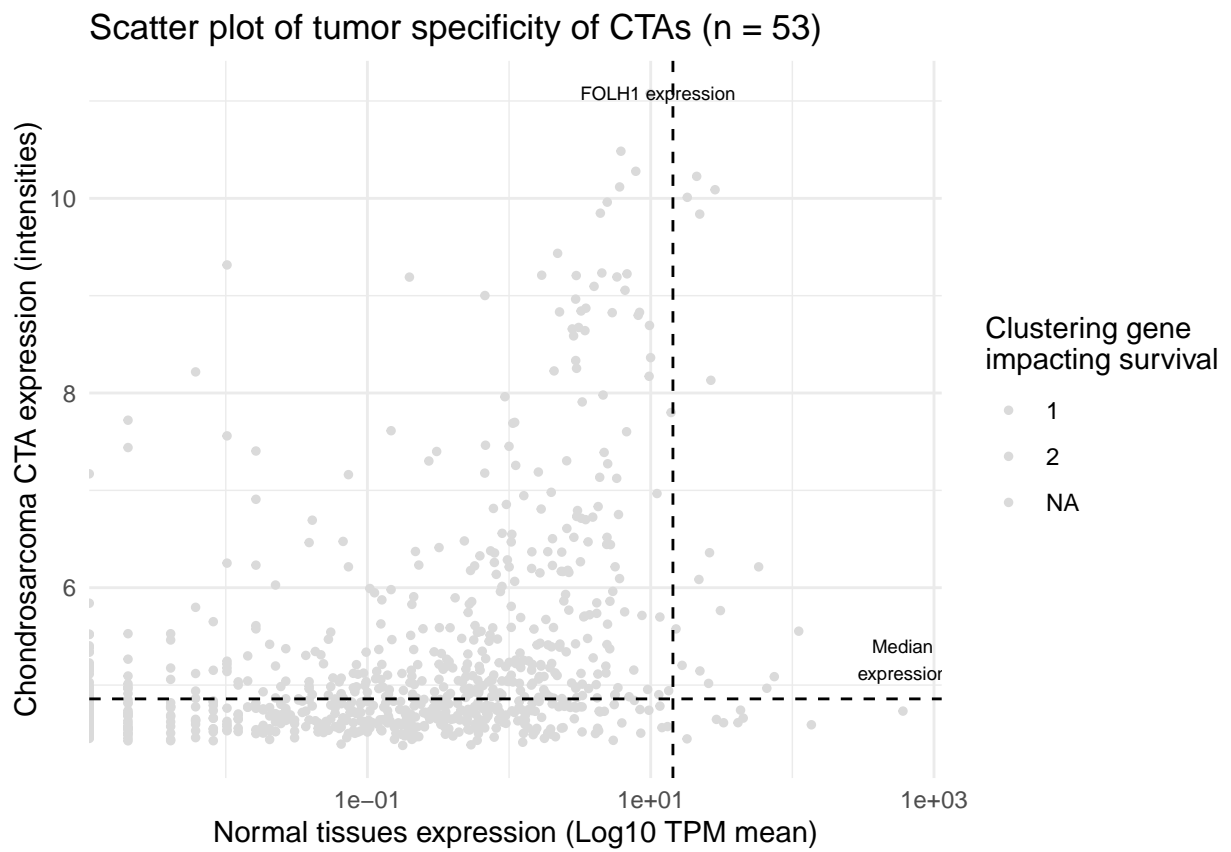
Figure 2: Scatter plot of CTAs expression in normal tissues and osteosarcoma (without color)

```
    x = folh1_expr - 3.1, y = max(df_scatter$Intensity), label = "FOLH1 expression",
    size = 2.5) + annotate("text", x = max(df_scatter$Mean_expression_tissues) -
    2.6, y = median(df_scatter$Intensity) + 0.4, label = "Median\nexpression",
    size = 2.5) + scale_shape_manual(values = c(COLD = 15, HOT = 16)) +
    scale_size_continuous(range = c(1, 3)) + scale_color_manual(values = c(`1` = "#3498DB",
    `2` = "#E74C3C"), na.value = "#dadada") + theme_minimal() +
    scale_x_log10() + labs(title = "Scatter plot of tumor specificity of CTAs (n = 53)",
    x = "Normal tissues expression (Log10 TPM mean)", y = "Chondrosarcoma CTA expression (intensities)"
    color = "Clustering gene\nimpacting survival\n(conventional CHS)",
    shape = "Associated Immunophenotype", size = "Signficant HR")
```

Warning in scale_x_log10(): log-10 transformation introduced infinite values.
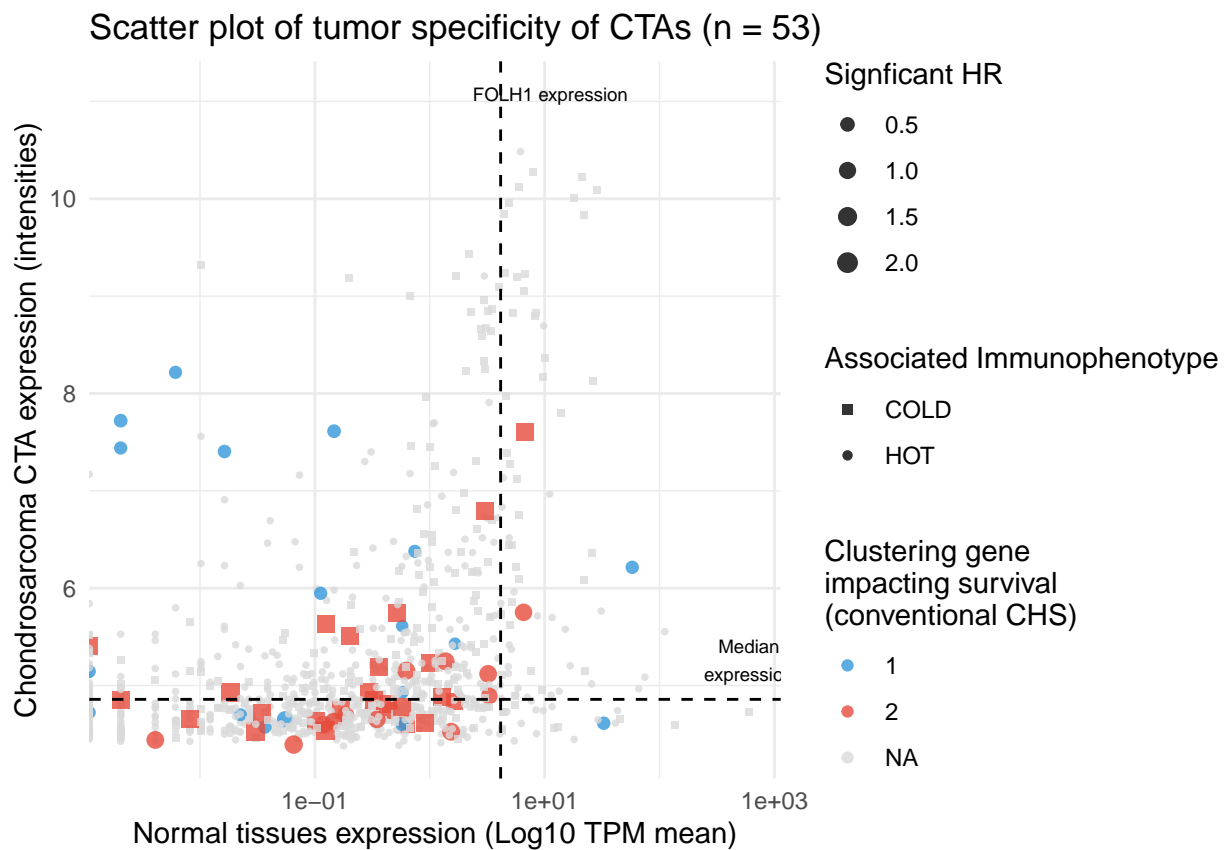

Figure 3: Scatter plot of CTAs expression in normal tissues and osteosarcoma (complete legend)