# Immunedeconv

Kevin Ryan

2023-05-08

## R Markdown

```
library(biomaRt)
library(immunedeconv)
library(corrplot)
library(forcats)
library(stringr)
library(here)
library(dplyr)
library(tidyr)
library(ggplot2)
library(RColorBrewer)
```

```
annotate_genes <- function(df){

    df$hgnc_symbol <- df$gene
    mart <- useMart(biomart = "ensembl", dataset = "hsapiens_gene_ensembl") #, host="uswest.ensembl.org
    info <- getBM(attributes=c("hgnc_symbol",
                               "ensembl_gene_id_version",
                               "chromosome_name",
                               "start_position",
                               "end_position",
                               "strand",
                               "entrezgene_description"),
                  filters = c("hgnc_symbol"),
                  values = df$gene,
                  mart = mart,
                  useCache=FALSE)

    tmp <- merge(df, info, by="hgnc_symbol")
    tmp$strand <- gsub("-1", "-", tmp$strand)
    tmp$strand <- gsub("1", "+", tmp$strand)
    tmp <- tmp[!grepl("CHR", tmp$chromosome_name),]

    return(tmp)

}

convert_ensg_version_to_hgnc_df_rownames <- function(df, ensdb = EnsDb.Hsapiens.v86){
  library(AnnotationDbi)
```

```
  library(ensembldb)
  library(EnsDb.Hsapiens.v86)
  genes_base <- str_split_fixed(string = rownames(df), pattern = "\\.", n = 2)[,1]
  newnames_original <- suppressWarnings(mapIds(EnsDb.Hsapiens.v86,
    keys = genes_base,
    column = 'SYMBOL',
    keytype = 'GENEID'))

# keep ensg version of newnames is na or is duplicated
newnames <- ifelse(is.na(newnames_original) | duplicated(newnames_original),
    rownames(df), newnames_original)
rownames(df) <- newnames
return(df)
}
```

```
salmon_tpm_path <- "/home/rstudio/Documents/PhD/CAF_data/nfcore_results/inhouse_data_nfcore_results_vers
salmon_tpm_in <- read.table(salmon_tpm_path, header = T)
colnames(salmon_tpm_in) <- gsub("X", "", colnames(salmon_tpm_in))
tx2gene_path = "~/Documents/PhD/subtypes/caf-subtype-analysis/nf-subpop/outdir/tx2gene/tx2gene.txt"
tx2gene <- read.table(tx2gene_path, header = T)
```

```
salmon_tpm_in[1:5,1:5]
```

```
##              gene_id gene_name      3532      3533      3536
## 1 ENSG00000000003.14    TSPAN6 15.867374 18.118736 17.147614
## 2  ENSG00000000005.6      TNMD  0.000000  0.000000  0.041895
## 3 ENSG00000000419.12      DPM1 76.108329 73.852506 73.578070
## 4 ENSG00000000457.14     SCYL3  4.126492  4.089532  3.484089
## 5 ENSG00000000460.17  C1orf112 16.378414 12.994835 12.456041
```

```
salmon_tpm_hgnc <- convert_ensg_version_to_hgnc_df_rownames(salmon_tpm_in)
```

```
## Loading required package: stats4
```

```
## Loading required package: BiocGenerics
```

```
##
## Attaching package: 'BiocGenerics'
```

```
## The following objects are masked from 'package:dplyr':
##
##     combine, intersect, setdiff, union
```

```
## The following objects are masked from 'package:stats':
##
##     IQR, mad, sd, var, xtabs
```

```
## The following objects are masked from 'package:base':
##
##     anyDuplicated, aperm, append, as.data.frame, basename, cbind,
```

```
##     colnames, dirname, do.call, duplicated, eval, evalq, Filter, Find,
##     get, grep, grepl, intersect, is.unsorted, lapply, Map, mapply,
##     match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,
##     Position, rank, rbind, Reduce, rownames, sapply, setdiff, sort,
##     table, tapply, union, unique, unsplit, which.max, which.min

## Loading required package: Biobase

## Welcome to Bioconductor
##
##     Vignettes contain introductory material; view with
##     'browseVignettes()'. To cite Bioconductor, see
##     'citation("Biobase")', and for packages 'citation("pkgname")'.

## Loading required package: IRanges

## Loading required package: S4Vectors

##
## Attaching package: 'S4Vectors'

## The following object is masked from 'package:tidyr':
##
##     expand

## The following objects are masked from 'package:dplyr':
##
##     first, rename

## The following object is masked from 'package:utils':
##
##     findMatches

## The following objects are masked from 'package:base':
##
##     expand.grid, I, unname

##
## Attaching package: 'IRanges'

## The following objects are masked from 'package:dplyr':
##
##     collapse, desc, slice

##
## Attaching package: 'AnnotationDbi'

## The following object is masked from 'package:dplyr':
##
##     select
```

```
## Loading required package: GenomicRanges

## Loading required package: GenomeInfoDb

## Loading required package: GenomicFeatures

## Loading required package: AnnotationFilter

##
## Attaching package: 'ensembldb'

## The following object is masked from 'package:dplyr':
##
##     filter

## The following object is masked from 'package:stats':
##
##     filter
```

```
salmon_tpm_hgnc_not_duplicated <- salmon_tpm_hgnc[!duplicated(salmon_tpm_hgnc$gene_name),]
salmon_tpm_hgnc_not_duplicated_hgnc_col <-  subset(salmon_tpm_hgnc_not_duplicated, select = -c(gene_id)
colnames(salmon_tpm_hgnc_not_duplicated_hgnc_col) <- gsub("X", "Sample_", colnames(salmon_tpm_hgnc_not_c
#write.table(salmon_tpm_hgnc_not_duplicated_hgnc_col, file = "/home/rstudio/Downloads/salmon_tpm_hgnc_c
rownames(salmon_tpm_hgnc_not_duplicated) <- salmon_tpm_hgnc_not_duplicated$gene_name
salmon_tpm_hgnc_not_duplicated <-  subset(salmon_tpm_hgnc_not_duplicated, select = -c(gene_name, gene_id
```

```
metadata <- read.csv(file = here("~/Documents/PhD/subtypes/caf-subtype-analysis/intermediate_files/metad

colnames(metadata)[1] <- "sample"
metadata$sample <- as.character(metadata$sample)
metadata$Condition <- ifelse(metadata$Condition == "Tumour", "CAF", "TAN")
```

```
metadata
```

```
##    sample Patient Condition Age Size   Grade Histology          ER          PR
## 1    4033       1       CAF  46   45 Grade_2   Lobular ER_positive PR_positive
## 2    4034       1       TAN  46   45 Grade_2   Lobular ER_positive PR_positive
## 3    4027       2       CAF  77   40 Grade_3    Ductal ER_negative PR_negative
## 4    4028       2       TAN  77   40 Grade_3    Ductal ER_negative PR_negative
## 5    4112       3       CAF  62   12 Grade_3    Ductal ER_positive PR_negative
## 6    4113       3       TAN  62    8 Grade_3    Ductal ER_positive PR_negative
## 7    4116       4       CAF  45   35 Grade_2   Lobular ER_positive PR_positive
## 8    4117       4       TAN  45   13 Grade_2   Lobular ER_positive PR_positive
## 9    4214       5       CAF  78   90 Grade_2   Lobular ER_positive PR_negative
## 10   4215       5       TAN  78   90 Grade_2   Lobular ER_positive PR_negative
## 11   4315       6       CAF  84   30 Grade_2    Ductal ER_positive PR_positive
## 12   4316       6       TAN  84   22 Grade_2    Ductal ER_positive PR_positive
## 13   4340       7       CAF  62  100 Grade_2   Lobular ER_positive PR_positive
## 14   4341       7       TAN  62  100 Grade_2   Lobular ER_positive PR_positive
## 15   4344       8       CAF  50   28 Grade_2    Ductal ER_positive PR_positive
## 16   4345       8       TAN  50   28 Grade_2    Ductal ER_positive PR_positive
```

```
## 17   3532       9       CAF  48   16 Grade_2    Ductal ER_positive PR_positive
## 18   3533       9       TAN  48   16 Grade_2    Ductal ER_positive PR_positive
## 19   3536      10       CAF  50   52 Grade_3   Lobular ER_positive PR_positive
## 20   3537      10       TAN  50   52 Grade_3   Lobular ER_positive PR_positive
## 21   4299      11       CAF  84   40 Grade_3    Ductal ER_positive PR_positive
## 22   4300      11       TAN  84   40 Grade_3    Ductal ER_positive PR_positive
## 23   4722      12       CAF  81   52 Grade_2   Lobular ER_positive PR_negative
## 24   4723      12       TAN  81   52 Grade_2   Lobular ER_positive PR_negative
##             Her2  Subtype          LVI
## 1  Her2_negative LuminalA LVI_negative
## 2  Her2_negative LuminalA LVI_negative
## 3  Her2_negative     TNBC LVI_positive
## 4  Her2_negative     TNBC LVI_positive
## 5  Her2_negative LuminalA LVI_negative
## 6  Her2_negative LuminalA LVI_negative
## 7  Her2_negative LuminalA LVI_positive
## 8  Her2_negative LuminalA LVI_positive
## 9  Her2_negative LuminalA LVI_positive
## 10 Her2_negative LuminalA LVI_positive
## 11 Her2_negative LuminalA LVI_positive
## 12 Her2_negative LuminalA LVI_positive
## 13 Her2_negative LuminalA LVI_negative
## 14 Her2_negative LuminalA LVI_negative
## 15 Her2_negative LuminalA LVI_negative
## 16 Her2_negative LuminalA LVI_negative
## 17 Her2_negative LuminalA LVI_negative
## 18 Her2_negative LuminalA LVI_negative
## 19 Her2_negative LuminalA LVI_negative
## 20 Her2_negative LuminalA LVI_negative
## 21 Her2_negative LuminalA LVI_positive
## 22 Her2_negative LuminalA LVI_positive
## 23 Her2_negative LuminalA LVI_negative
## 24 Her2_negative LuminalA LVI_negative
```

```
# includes cancer-associated fibroblast
deconvolution_mcp <- immunedeconv::deconvolute(salmon_tpm_hgnc_not_duplicated, "mcp_counter")
```

```
##
## >>> Running mcp_counter
```

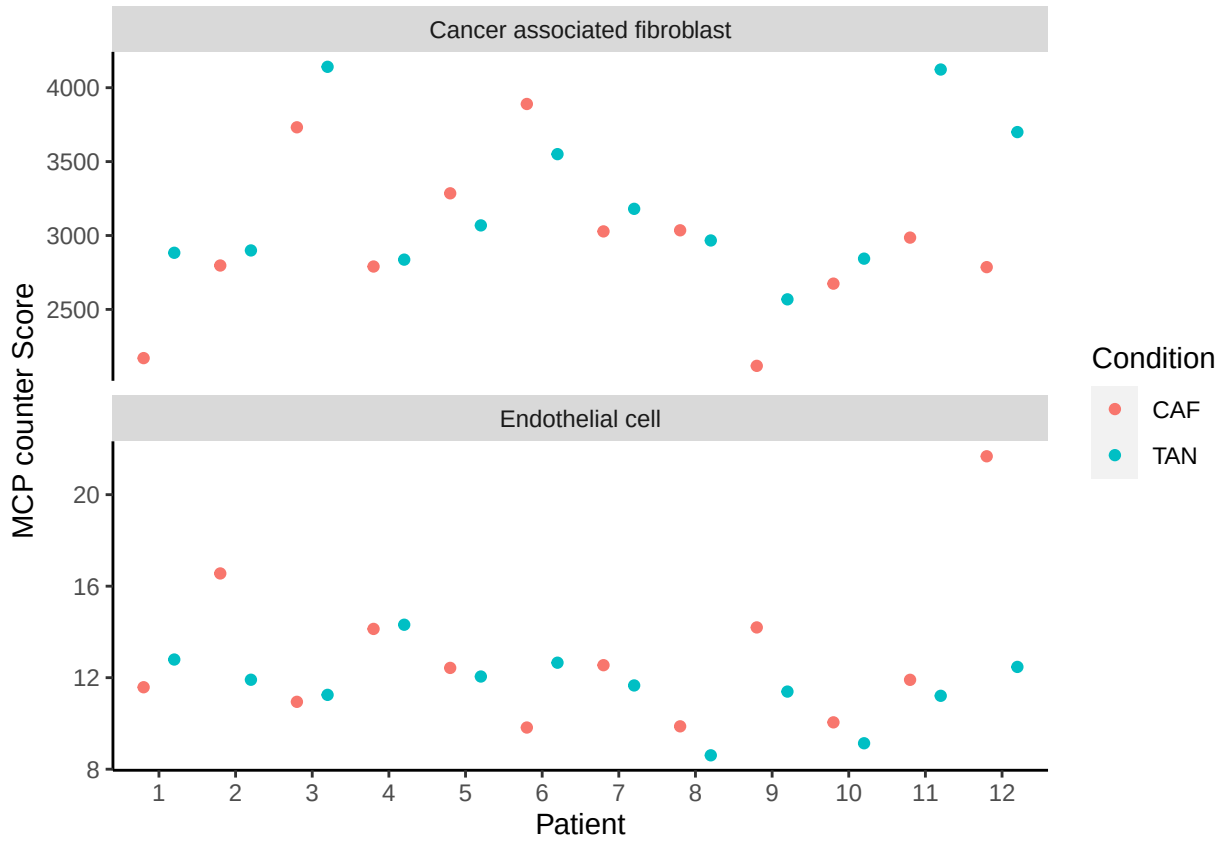
```
deconvolution_mcp_filtered_long <- deconvolution_mcp %>% dplyr::filter(cell_type %in% c("Cancer associa
deconvolution_mcp_filtered_long_metadata <- full_join(deconvolution_mcp_filtered_long, metadata)
```

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

```
ord <- as.character(seq(1,12,1))
# Plotting
deconvolution_mcp_filtered_long_metadata_ordered <- deconvolution_mcp_filtered_long_metadata[order(deco
mcp_plot <- ggplot(deconvolution_mcp_filtered_long_metadata_ordered, aes(x = as.character(Patient), y =
  geom_point(position = position_dodge(width = 0.8)) +
    scale_x_discrete(limits = ord) +
    theme(#panel.grid.major = element_blank(),
```

```
        #panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        axis.line = element_line(colour = "black")) +
  facet_wrap(~ cell_type, nrow = 2, scales = "free_y") +
  xlab("Patient") +
  ylab("MCP counter Score")
```

```
mcp_plot
```




```
# no fibroblast cell type
quantiseq <- immunedeconv::deconvolute(salmon_tpm_hgnc_not_duplicated, method= "quantiseq")
```

```
##
## >>> Running quantiseq

##
## Running quanTIseq deconvolution module

## Gene expression normalization and re-annotation (arrays: FALSE)

## Removing 17 noisy genes

## Removing 15 genes with high expression in tumors
```

```
## Signature genes found in data set: 135/138 (97.83%)

## Mixture deconvolution (method: lsei)

## Deconvolution successful!
```

```
set_cibersort_binary("/home/rstudio/sw/CIBERSORT/CIBERSORT.R")
#set_cibersort_binary("/home/rstudio/sw/CIBERSORT/CIBERSORT_1.4.R")
set_cibersort_mat("/home/rstudio/sw/CIBERSORT/LM22.txt")
#cibersort_out <- immunedeconv::deconvolute(salmon_tpm_hgnc_not_duplicated, "cibersort")   # or 'cibers
```

```
# includes cancer-associated fibroblast
deconvolution_epic <- immunedeconv::deconvolute(salmon_tpm_hgnc_not_duplicated, "epic")
```

```
##
## >>> Running epic

## Warning in (function (bulk, reference = NULL, mRNA_cell = NULL, mRNA_cell_sub = NULL, : The optimiza
## 4723
##  - check fit.gof for the convergeCode and convergeMessage

## Warning in (function (bulk, reference = NULL, mRNA_cell = NULL, mRNA_cell_sub =
## NULL, : mRNA_cell value unknown for some cell types: CAFs, Endothelial - using
## the default value of 0.4 for these but this might bias the true cell
## proportions from all cell types.
```

```
#cibersort_signature_matrix <- read.table("/home/rstudio/Documents/PhD/")
```

```
deconvolution_timer <- deconvolute(salmon_tpm_hgnc_not_duplicated, "timer",
            indications=rep("BRCA", 24))
```

```
##
## >>> Running timer

## ## Enter batch mode

## ## Loading immune gene expression

## [1] "Outlier genes: ACTB ACTG1 COL1A1 COL1A2 EEF1A1 FN1 GAPDH MT-ATP8 MT-CO1 MT2A SPARC TGFBI TIMP1 T

## ## Removing the batch effect of /tmp/RtmpaoXgxF/fileeba5249320a

## Found 1033 genes with uniform expression within a single batch (all zeros); these will not be adjust

## Found2batches

## Adjusting for0covariate(s) or covariate level(s)

## Standardizing Data across genes
```

```
## Fitting L/S model and finding priors

## Finding parametric adjustments

## Adjusting the Data
```

```
# timer default is for immune cells - no cafs
deconvolution_timer
```

```
## # A tibble: 6 x 25
##   cell_type       `3532` `3533` `3536` `3537` `4027` `4028` `4033` `4034` `4112`
##   <chr>            <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>
## 1 B cell           0.151  0.139  0.146  0.141  0.151  0.141  0.144  0.132  0.144
## 2 T cell CD4+      0.167  0.162  0.161  0.163  0.163  0.164  0.164  0.165  0.159
## 3 T cell CD8+      0.162  0.149  0.146  0.147  0.150  0.157  0.146  0.159  0.148
## 4 Neutrophil       0.115  0.127  0.134  0.124  0.120  0.120  0.126  0.119  0.116
## 5 Macrophage       0.113  0.115  0.112  0.124  0.108  0.115  0.129  0.131  0.133
## 6 Myeloid dendri~  0.356  0.331  0.329  0.328  0.365  0.333  0.325  0.318  0.323
## # i 15 more variables: `4113` <dbl>, `4116` <dbl>, `4117` <dbl>, `4214` <dbl>,
## #   `4215` <dbl>, `4299` <dbl>, `4300` <dbl>, `4315` <dbl>, `4316` <dbl>,
## #   `4340` <dbl>, `4341` <dbl>, `4344` <dbl>, `4345` <dbl>, `4722` <dbl>,
## #   `4723` <dbl>
```

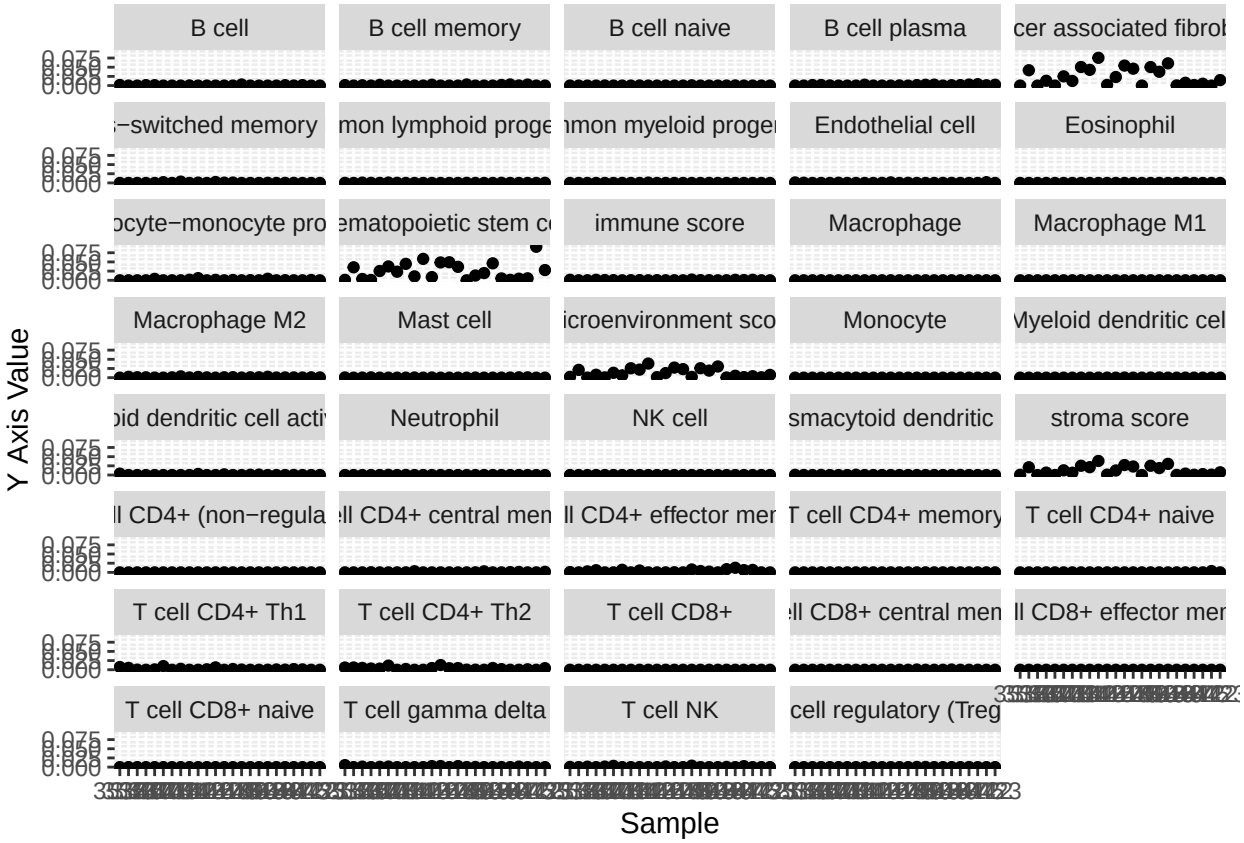
```
deconvolution_xcell <- immunedeconv::deconvolute(salmon_tpm_hgnc_not_duplicated, "xcell")
```

```
##
## >>> Running xcell
```

```
colnames(deconvolution_xcell) <- gsub("X", "", colnames(deconvolution_xcell))
```

```
# cancer-associated fibroblast, hematopoeitic stem cell, microenvironment score, stroma score
deconvolution_xcell_long <- deconvolution_xcell %>% pivot_longer(!cell_type, names_to = "sample", values

# ggplot(data = deconvolution_xcell, aes(x = cell_type, y = xcell_score)) +
#   geom_bar(stat = "identity", alpha = 0.7) +
#   facet_grid(. ~sample)  +
#   ylim(0,800) +
#   geom_text(aes(label = Freq), fontface = "bold", vjust = 1.5, colour = "white", size = 4) +
#   labs(x = "\n Coin Flip Outcome", y = "Frequency\n", title = "\n Coin Flip Results \n") +
#   theme(plot.title = element_text(hjust = 0.5),
#         axis.title.x = element_text(face="bold", colour="darkgreen", size = 12),
#         axis.title.y = element_text(face="bold", colour="darkgreen", size = 12),
#         legend.title = element_text(face="bold", size = 10),
#         strip.background = element_rect(fill="lightblue", colour="black", size=1),
#         strip.text = element_text(face="bold", size=rel(1.2)))
# Create the ggplot object
ggplot(deconvolution_xcell_long, aes(x = sample, y = xcell_score)) +
  # Add the facet wrap
  facet_wrap(~ cell_type, ncol = 5) +
  # Add the points
  geom_point() +
```

```
  # Add the x-axis label
  xlab("Sample") +
  # Add the y-axis label
  ylab("Y Axis Value")
```

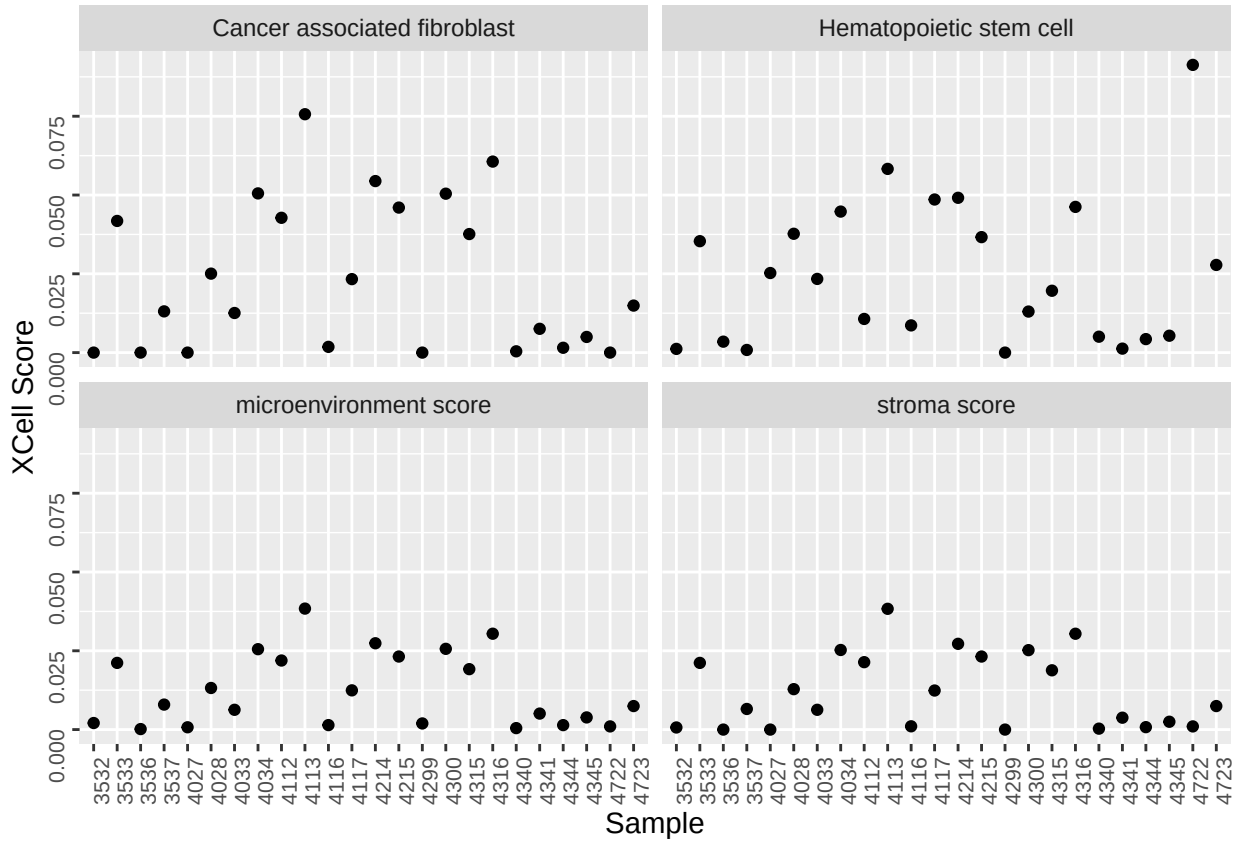



```
deconvolution_xcell_filtered_long <- deconvolution_xcell %>% dplyr::filter(cell_type %in% c("Cancer ass
deconvolution_xcell_filtered_long_metadata <- full_join(deconvolution_xcell_filtered_long, metadata)
```

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

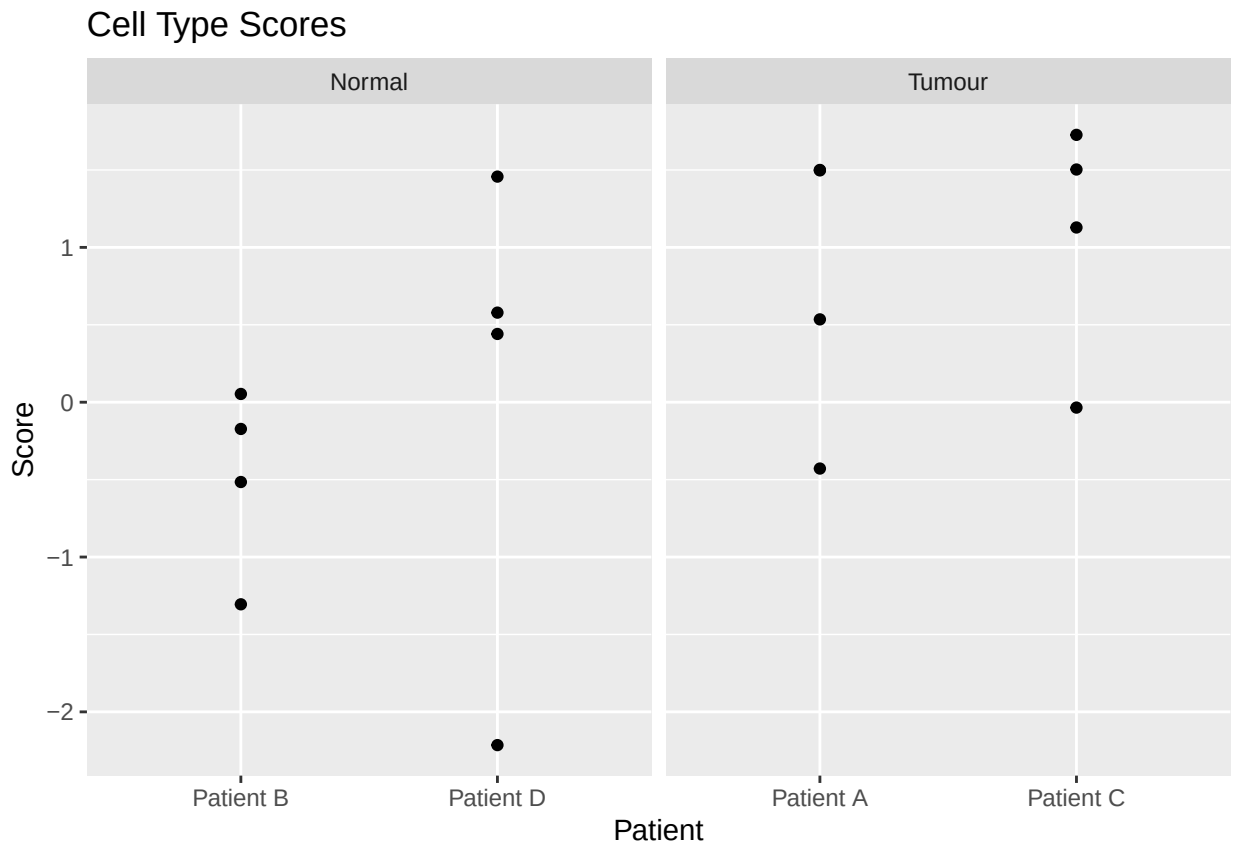
```
#deconvolution_xcell %>% pivot_longer(!cell_type, names_to = "sample", values_to = "xcell_score")
```

```
ggplot(deconvolution_xcell_filtered_long, aes(x = sample, y = xcell_score)) +
  # Add the facet wrap
  facet_wrap(~ cell_type, ncol = 2) +
  # Add the points
  geom_point() +
  # Add the x-axis label
  xlab("Sample") +
  # Add the y-axis label
  ylab("XCell Score") +
  theme(axis.text = element_text(size = 8, angle = 90))
```




```
xcell_plot <- deconvolution_xcell_filtered_long_metadata %>%
  mutate(Condition = fct_relevel(Condition, "TAN", "CAF")) %>%
  ggplot(aes(x = Patient, y = xcell_score)) +
  #geom_col() +
  facet_wrap(~Patient, nrow=1, scales = "free_x", strip.position = "bottom") +
    facet_wrap(~ cell_type, ncol = 2) +
    #theme(plot.title = element_text(hjust = 0.5, size = 14, face = "bold"),
          #axis.text.y = element_text(size = 12, face = "bold"),
          #axis.text.x = element_text(size = 8),
          #strip.text.y = element_text(size = rel(100))
          #) +
  #ggtitle("CAF subpopulation proportions\n determined by CIBERSORTx") +
    scale_fill_manual(name=NULL,
                      values = c(brewer.pal(3, "Dark2"), "gray")
                      ) +
  xlab(label = "Patient") +
  ylab("XCell score") +
  theme(#panel.grid.major = element_blank(),
        #panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        axis.line = element_line(colour = "black")) +
     # scale_y_continuous(expand = c(0,0)) +
  #geom_hline(yintercept = c(25, 50, 75), color = "gray", linetype = "dashed") +
  theme(legend.text=element_text(size=rel(1.2)),
        axis.text.x = element_text(size = 8, angle = 90))
```

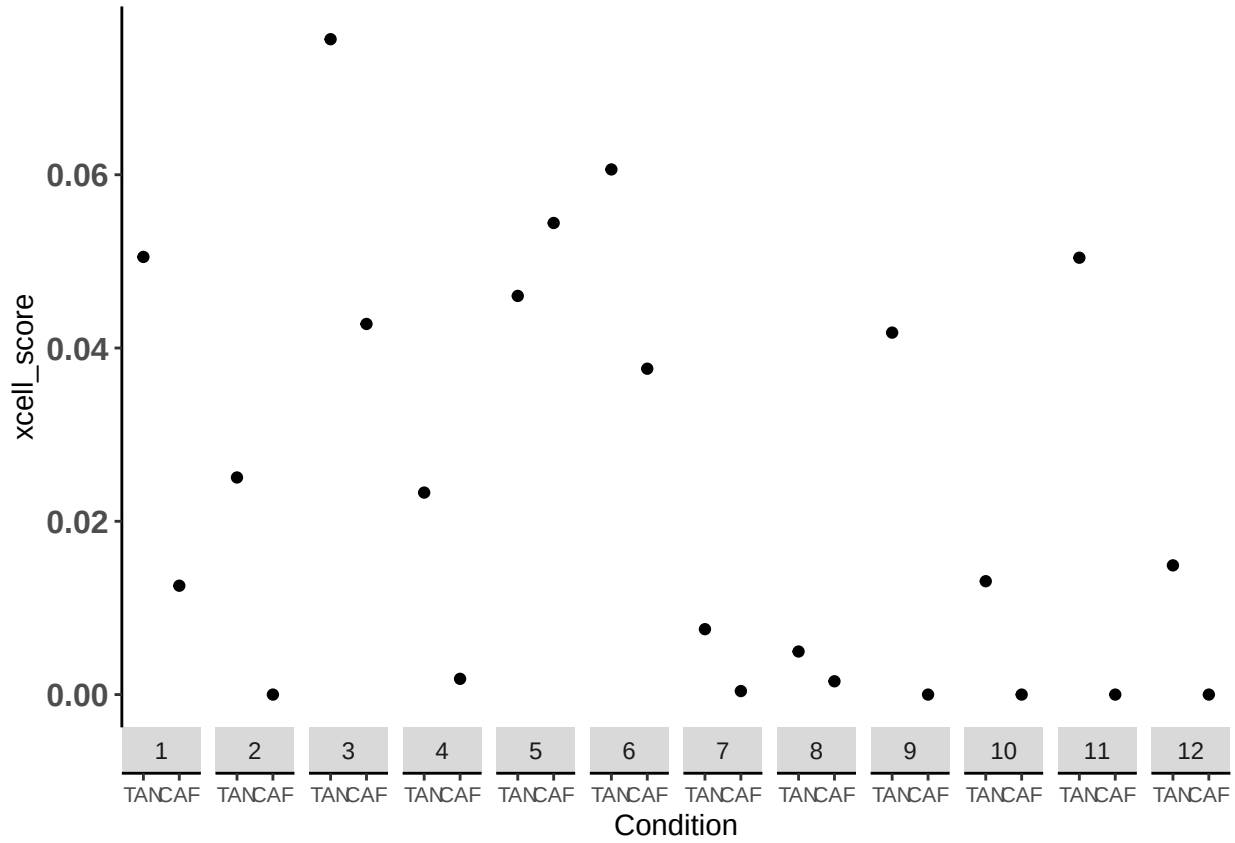
```r
# Example data frame
df <- data.frame(
  Patient = rep(c("Patient A", "Patient B", "Patient C", "Patient D"), 4),
  CellType = rep(c("Cell Type 1", "Cell Type 2", "Cell Type 3", "Cell Type 4"), each = 4),
  Status = rep(c("Tumour", "Normal"), times = 8),
  Score = rnorm(16)
)

# Plotting
ggplot(df, aes(x = Patient, y = Score)) +
  geom_point() +
  facet_wrap(~ CellType, scales = "free_y") +
  facet_grid(~ Status, scales = "free_x") +
  xlab("Patient") +
  ylab("Score") +
  ggtitle("Cell Type Scores")
```




```r
xcell_plot_test <- deconvolution_xcell_filtered_long_metadata %>%
  dplyr::filter(cell_type == "Cancer associated fibroblast") %>%
  mutate(Condition = fct_relevel(Condition, "TAN", "CAF")) %>%
  ggplot(aes(x = Condition, y = xcell_score)) +
  geom_point() +
 facet_wrap(~Patient, nrow=1, scales = "free_x", strip.position = "bottom") +
    theme(plot.title = element_text(hjust = 0.5, size = 14, face = "bold"),
          axis.text.y = element_text(size = 12, face = "bold"),
```

```
          axis.text.x = element_text(size = 8),
          strip.text.y = element_text(size = rel(100))
          ) +
  #ggtitle("CAF subpopulation proportions\n determined by CIBERSORTx") +
  #   scale_fill_manual(name=NULL,
  #                     values = c(brewer.pal(3, "Dark2"), "gray")
  #                     ) +
  #xlab(label = "Patient") +
  theme(#panel.grid.major = element_blank(),
        #panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        axis.line = element_line(colour = "black")) +
      # scale_y_continuous(expand = c(0,0.08)) +
  #geom_hline(yintercept = c(25, 50, 75), color = "gray", linetype = "dashed") +
  theme(legend.text=element_text(size=rel(1.2)),
        axis.text.x = element_text(size = 8))
```

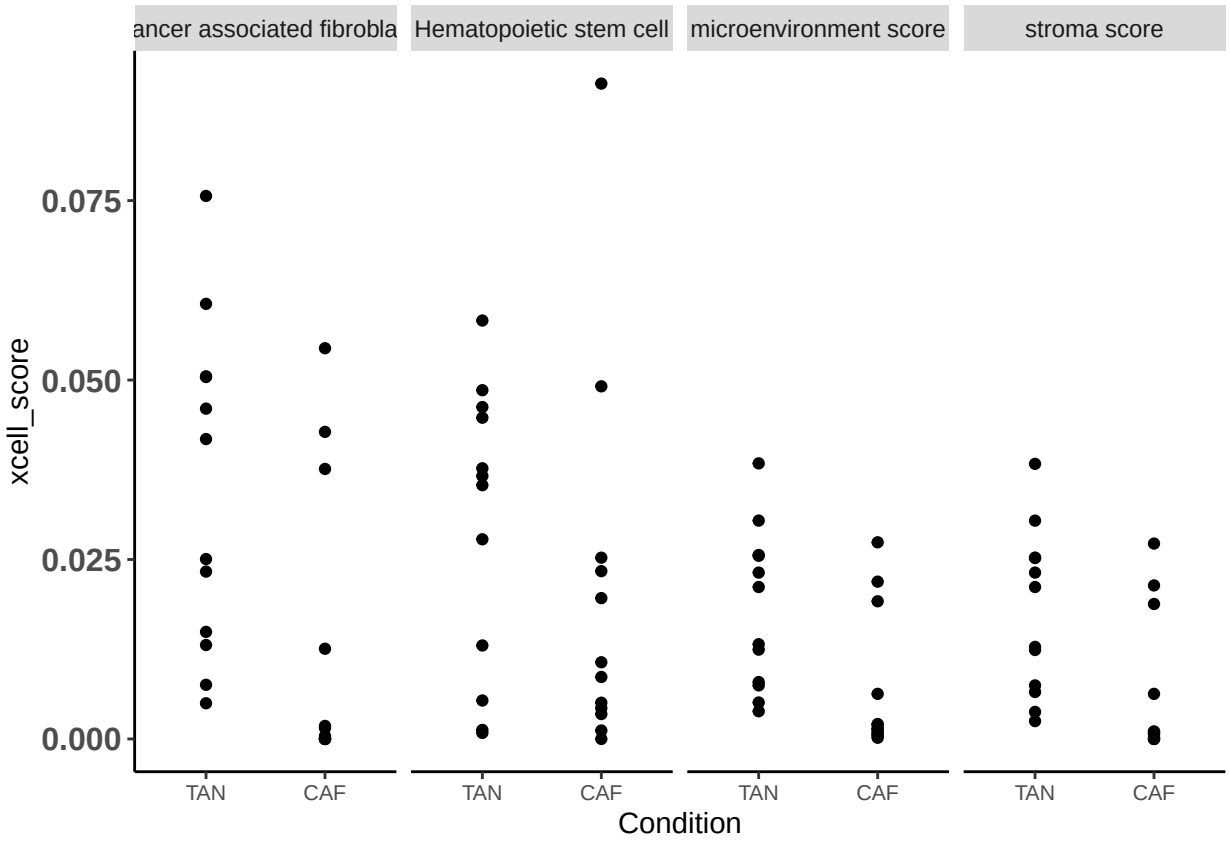
```
xcell_plot_test
```




```
xcell_plot_test2 <- deconvolution_xcell_filtered_long_metadata %>%
  #dplyr::filter(cell_type == "Cancer associated fibroblast") %>%
  mutate(Condition = fct_relevel(Condition, "TAN", "CAF")) %>%
  ggplot(aes(x = Condition, y = xcell_score)) +
  geom_point() +
 facet_wrap(~Patient, nrow=1, scales = "free_x", strip.position = "bottom") +
```

```
    theme(plot.title = element_text(hjust = 0.5, size = 14, face = "bold"),
          axis.text.y = element_text(size = 12, face = "bold"),
          axis.text.x = element_text(size = 8),
          strip.text.y = element_text(size = rel(100))
          ) +
    facet_grid(~ cell_type, scales = "free_y") +
  #ggtitle("CAF subpopulation proportions\n determined by CIBERSORTx") +
  #   scale_fill_manual(name=NULL,
  #                     values = c(brewer.pal(3, "Dark2"), "gray")
  #                     ) +
  #xlab(label = "Patient") +
  theme(#panel.grid.major = element_blank(),
        #panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        axis.line = element_line(colour = "black")) +
     # scale_y_continuous(expand = c(0,0.08)) +
  #geom_hline(yintercept = c(25, 50, 75), color = "gray", linetype = "dashed") +
  theme(legend.text=element_text(size=rel(1.2)),
        axis.text.x = element_text(size = 8))
```

```
xcell_plot_test2
```




```
# Example data frame
ord <- as.character(seq(1,12,1))
```

```r
# Plotting
deconvolution_xcell_filtered_long_metadata_ordered <- deconvolution_xcell_filtered_long_metadata[order(
xcell_plot <- ggplot(deconvolution_xcell_filtered_long_metadata_ordered, aes(x = as.character(Patient),
  geom_point(position = position_dodge(width = 0.8)) +
    scale_x_discrete(limits = ord) +
    theme(#panel.grid.major = element_blank(),
          #panel.grid.minor = element_blank(),
          panel.background = element_blank(),
          axis.line = element_line(colour = "black")) +
  facet_wrap(~ cell_type, nrow = 2, scales = "free_y") +
  xlab("Patient") +
  ylab("xCell Score")
```

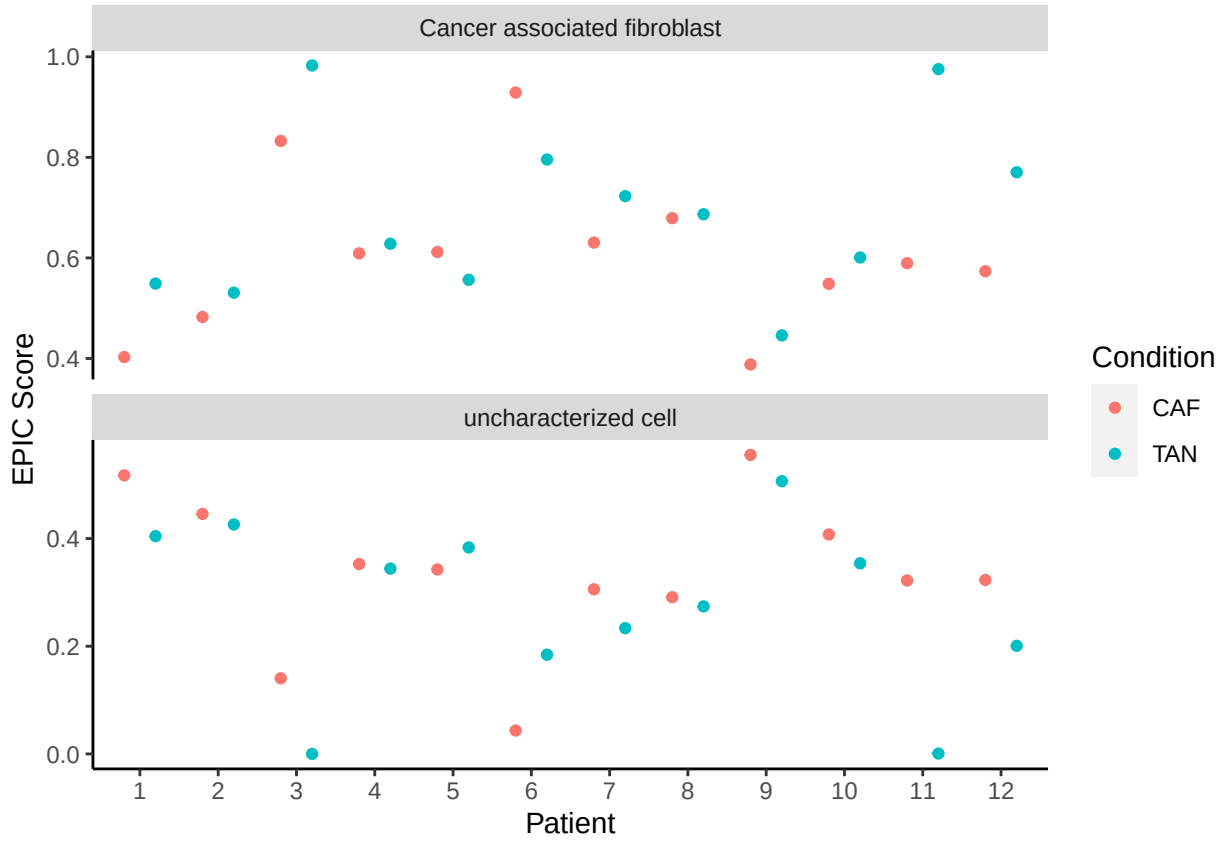
```r
colnames(deconvolution_epic) <- gsub("X", "", colnames(deconvolution_epic))
deconvolution_epic_long <- deconvolution_epic %>% pivot_longer(!cell_type, names_to = "sample", values_
```

```r
deconvolution_epic_filtered_long <- deconvolution_epic %>% dplyr::filter(cell_type %in% c("Cancer assoc
deconvolution_epic_filtered_long_metadata <- full_join(deconvolution_epic_filtered_long, metadata)
```

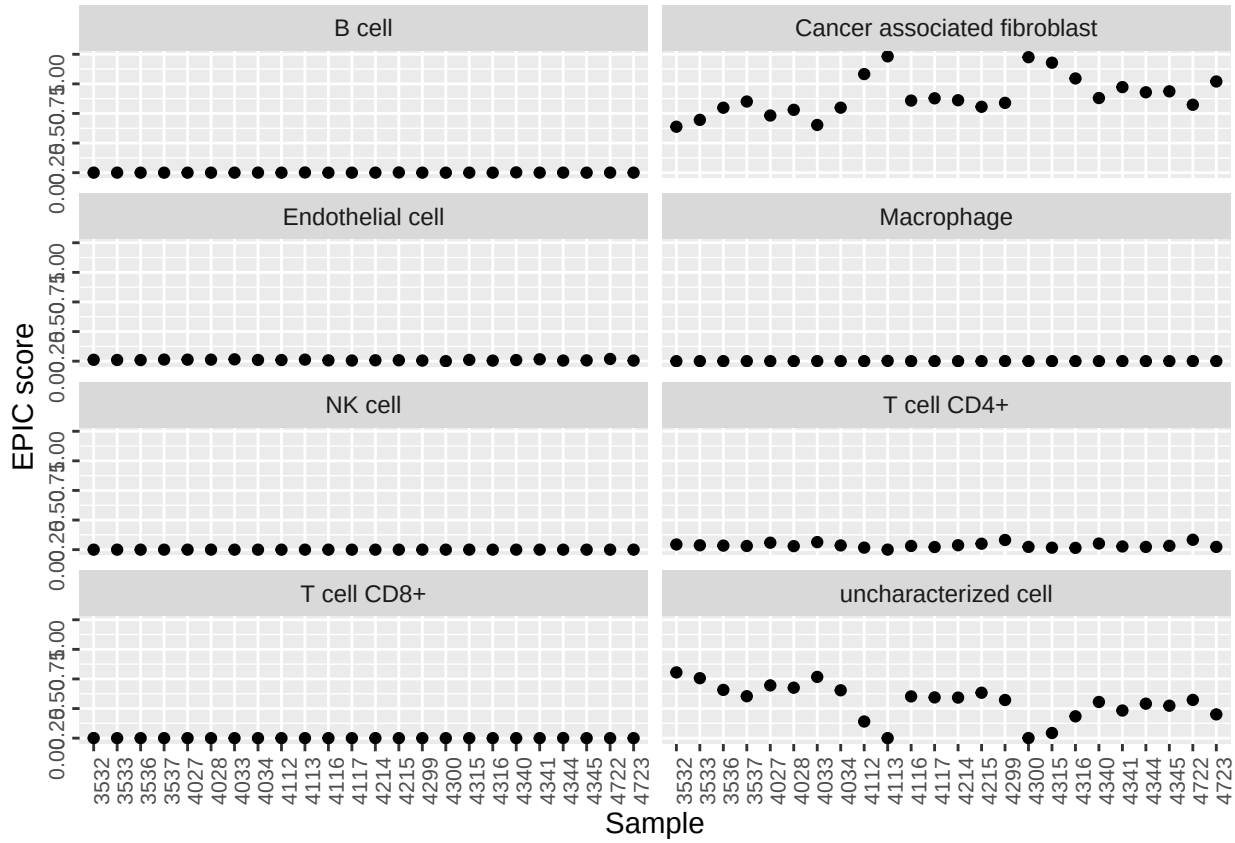
```
## Joining with `by = join_by(sample)`
```

```r
# Plotting
deconvolution_epic_filtered_long_metadata_ordered <- deconvolution_epic_filtered_long_metadata[order(de
epic_plot <- ggplot(deconvolution_epic_filtered_long_metadata_ordered, aes(x = as.character(Patient), y
  geom_point(position = position_dodge(width = 0.8)) +
    scale_x_discrete(limits = ord) +
    theme(#panel.grid.major = element_blank(),
          #panel.grid.minor = element_blank(),
          panel.background = element_blank(),
          axis.line = element_line(colour = "black")) +
  facet_wrap(~ cell_type, nrow = 2, scales = "free_y") +
  xlab("Patient") +
  ylab("EPIC Score")
```

```r
epic_plot
```




```
ggplot(deconvolution_epic_long, aes(x = sample, y = epic_score)) +
  # Add the facet wrap
  facet_wrap(~ cell_type, ncol = 2) +
  # Add the points
  geom_point() +
  # Add the x-axis label
  xlab("Sample") +
  # Add the y-axis label
  ylab("EPIC score") +
  theme(axis.text = element_text(size = 8, angle = 90))
```



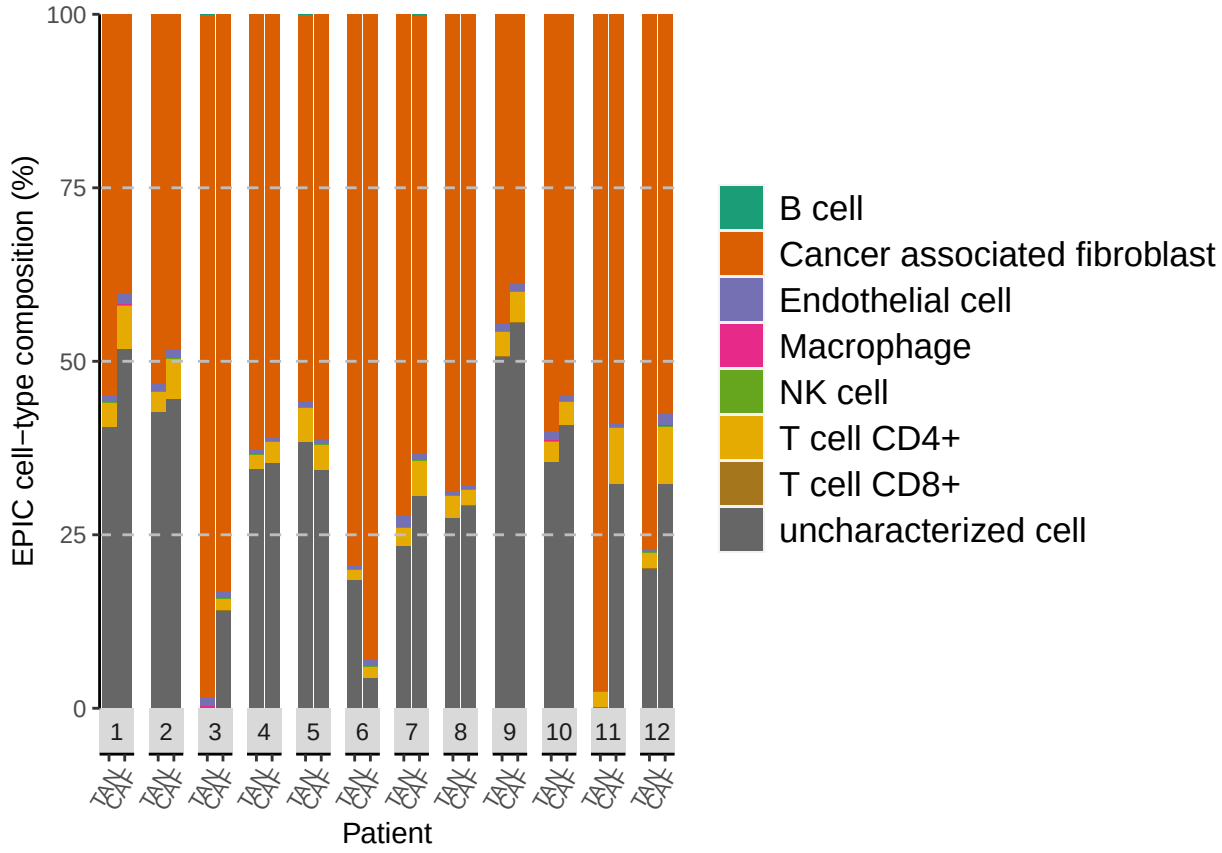

```
deconvolution_epic_long_metadata <- full_join(deconvolution_epic_long, metadata)
```

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

```
#deconvolution_epic_filtered_long_metadata$cell_type <- relevel(deconvolution_epic_filtered_long_metada

epic_plot_stack <- deconvolution_epic_long_metadata %>%
  mutate(Condition = fct_relevel(Condition, "TAN", "CAF")) %>%
  mutate(epic_score = 100*epic_score) %>%
  ggplot(aes(x = Condition, y = epic_score, fill = cell_type)) +
  geom_col() +
  facet_wrap(~Patient, nrow=1, scales = "free_x", strip.position = "bottom") +
    #theme(plot.title = element_text(hjust = 0.5, size = 14, face = "bold"),
          #axis.text.y = element_text(size = 12, face = "bold"),
          #axis.text.x = element_text(size = 8),
          #strip.text.y = element_text(size = rel(100))
          #) +
  #ggtitle("CAF subpopulation proportions\n determined by CIBERSORTx") +
    scale_fill_manual(name=NULL,
                      values = c(brewer.pal(8, "Dark2"))
                      ) +
  xlab(label = "Patient") +
  ylab("EPIC cell-type composition (%)") +
  theme(#panel.grid.major = element_blank(),
        #panel.grid.minor = element_blank(),
```

```
        panel.background = element_blank(),
        axis.line = element_line(colour = "black")) +
      scale_y_continuous(expand = c(0,0)) +
  geom_hline(yintercept = c(25, 50, 75), color = "gray", linetype = "dashed") +
  theme(legend.text=element_text(size=rel(1.2)),
        axis.text.x = element_text(size = 8))


 # labs(tag = "TAN = Tumour-associated normal\nCAF=Cancer-associated fibroblast") +
  #  theme(plot.tag.position = c(0.9, 0.3),)

  #+
     # scale_y_continuous(expand = c(0, 0), limits = c(0, 1.0000001))
```

```
epic_plot_stack <- epic_plot_stack + theme(axis.text.x=element_text(angle=60, hjust=1))
epic_plot_stack
```




```
cor.test(as.numeric(deconvolution_epic[which(deconvolution_epic$cell_type == "Cancer associated fibrobla
```

```
##
##  Pearson's product-moment correlation
##
## data:  as.numeric(deconvolution_epic[which(deconvolution_epic$cell_type == "Cancer associated fibrobl
## t = -42.528, df = 22, p-value < 2.2e-16
```

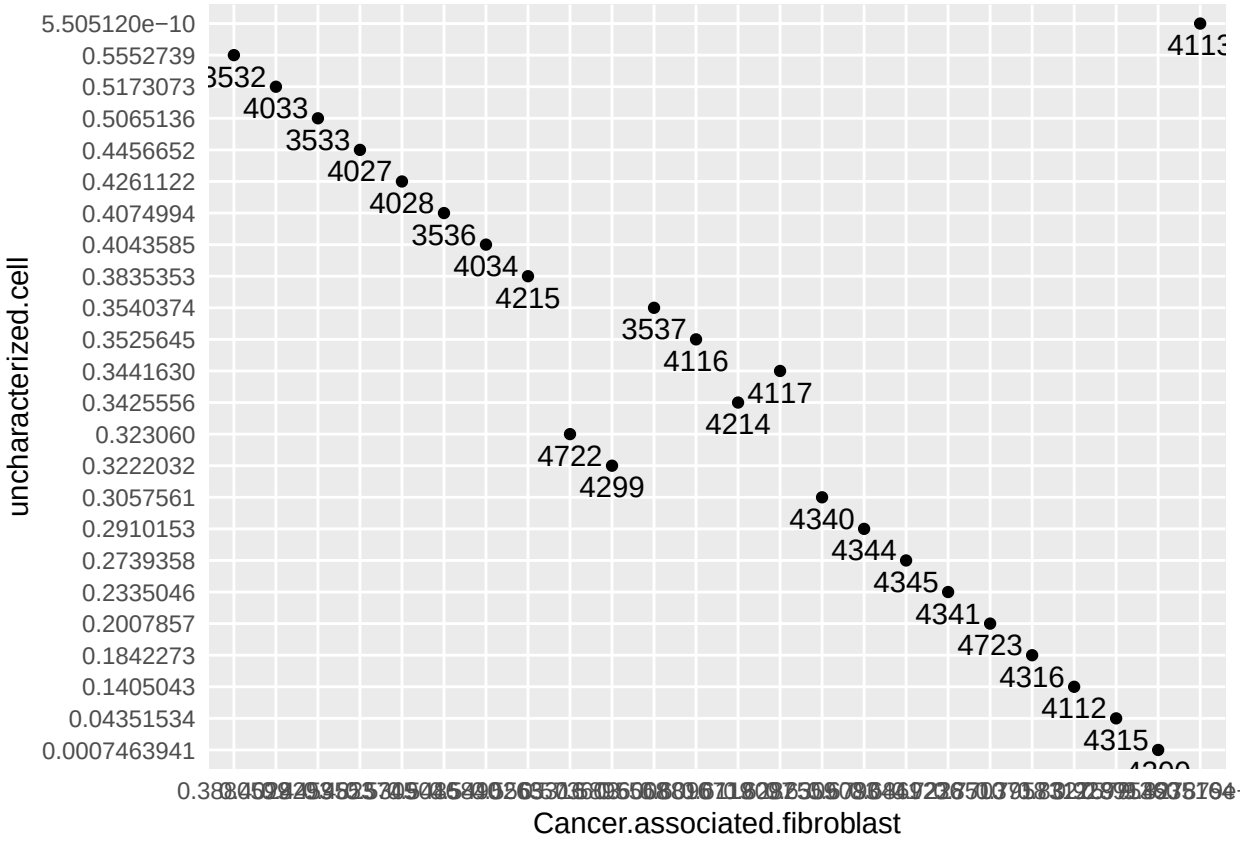
```
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.9974334 -0.9858799
## sample estimates:
##       cor
## -0.9939729
```

```
as.numeric(deconvolution_epic[which(deconvolution_epic$cell_type == "Cancer associated fibroblast"),c(2
```

```
##  [1] 0.3880599 0.4459553 0.5485840 0.6008816 0.4825745 0.5309046 0.4028293
##  [8] 0.5490263 0.8327579 0.9827879 0.6090719 0.6282559 0.6118097 0.5565316
## [15] 0.5895568 0.9754935 0.9289136 0.7957312 0.6305083 0.7228510 0.6790441
## [22] 0.6869236 0.5736030 0.7703718
```

```
#cor.test(cibersort_caf_subpopulation_results$S1, as.numeric(deconvolution_epic[which(deconvolution_epi
```

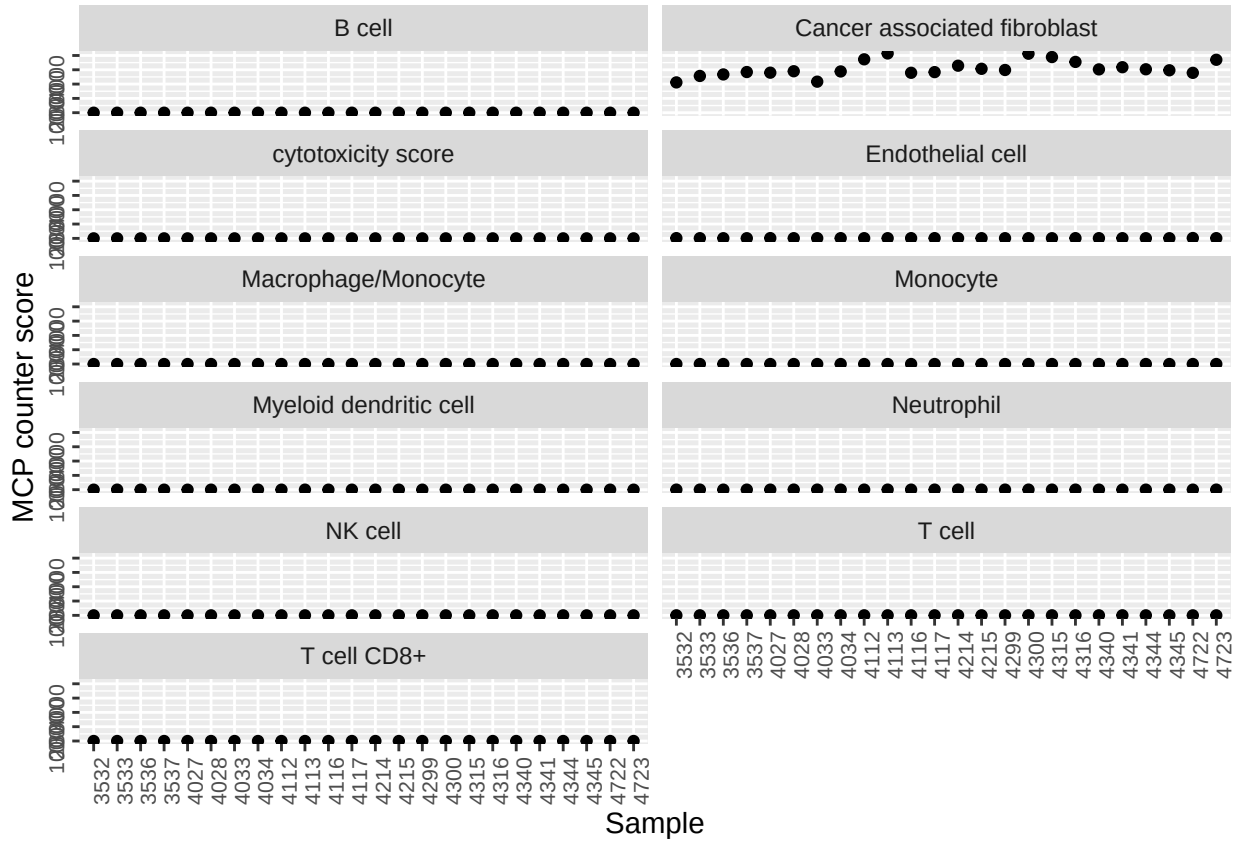
```
#deconvolution_epic_filter_long <- deconvolution_epic %>% dplyr::filter(cell_type %in% c("Cancer associ
deconvolution_epic_filter <- deconvolution_epic %>% dplyr::filter(cell_type %in% c("Cancer associated f
deconvolution_epic_filter <- t(deconvolution_epic_filter)
colnames(deconvolution_epic_filter) <- deconvolution_epic_filter[1,]
deconvolution_epic_filter <- data.frame(deconvolution_epic_filter[-c(1),])
deconvolution_epic_filter <- deconvolution_epic_filter %>% mutate(Sample = rownames(deconvolution_epic_f
ggplot(data = deconvolution_epic_filter) +
  geom_point(mapping = aes(x = Cancer.associated.fibroblast, y = uncharacterized.cell)) +
   theme(#axis.text.x=element_blank(), #remove x axis labels
        axis.ticks.x=element_blank(), #remove x axis ticks
        #axis.text.y=element_blank(),  #remove y axis labels
        axis.ticks.y=element_blank()  #remove y axis ticks
        ) +
  #geom_text(data=subset(deconvolution_epic_filter, Sample == "4113"), aes(x = Cancer.associated.fibrob
  geom_text(data=deconvolution_epic_filter, aes(x = Cancer.associated.fibroblast, y = uncharacterized.ce
```




```
cibersort_caf_subpopulation_results <- read.csv("/home/rstudio/Documents/PhD/subtypes/caf-subtype-analys
```

```
cibersort_caf_subpopulation_results$Mixture <- gsub("X", "",cibersort_caf_subpopulation_results$Mixture)
colnames(cibersort_caf_subpopulation_results)[1] <- "Sample"
cibersort_caf_subpopulation_results <- cibersort_caf_subpopulation_results %>% dplyr::select(-c("P.value
```

```
deconvolution_mcp_long <- deconvolution_mcp %>% pivot_longer(!cell_type, names_to = "sample", values_to
deconvolution_mcp_long$sample <- gsub("X", replacement = "", deconvolution_mcp_long$sample)
ggplot(deconvolution_mcp_long, aes(x = sample, y = mcp_counter_score)) +
  # Add the facet wrap
  facet_wrap(~ cell_type, ncol = 2) +
  # Add the points
  geom_point() +
  # Add the x-axis label
  xlab("Sample") +
  # Add the y-axis label
  ylab("MCP counter score") +
  theme(axis.text = element_text(size = 8, angle = 90))
```




```
deconvolution_mcp_filter <- data.frame(t(deconvolution_mcp[which(deconvolution_mcp$cell_type =="Cancer a
colnames(deconvolution_mcp_filter) <- deconvolution_mcp_filter[1,]
deconvolution_mcp_filter <- data.frame(deconvolution_mcp_filter[-1,])
deconvolution_mcp_filter$Sample <- colnames(deconvolution_mcp)[2:ncol(deconvolution_mcp)]
colnames(deconvolution_mcp_filter)[1] <- "mcp_counter_score_CAF"
```

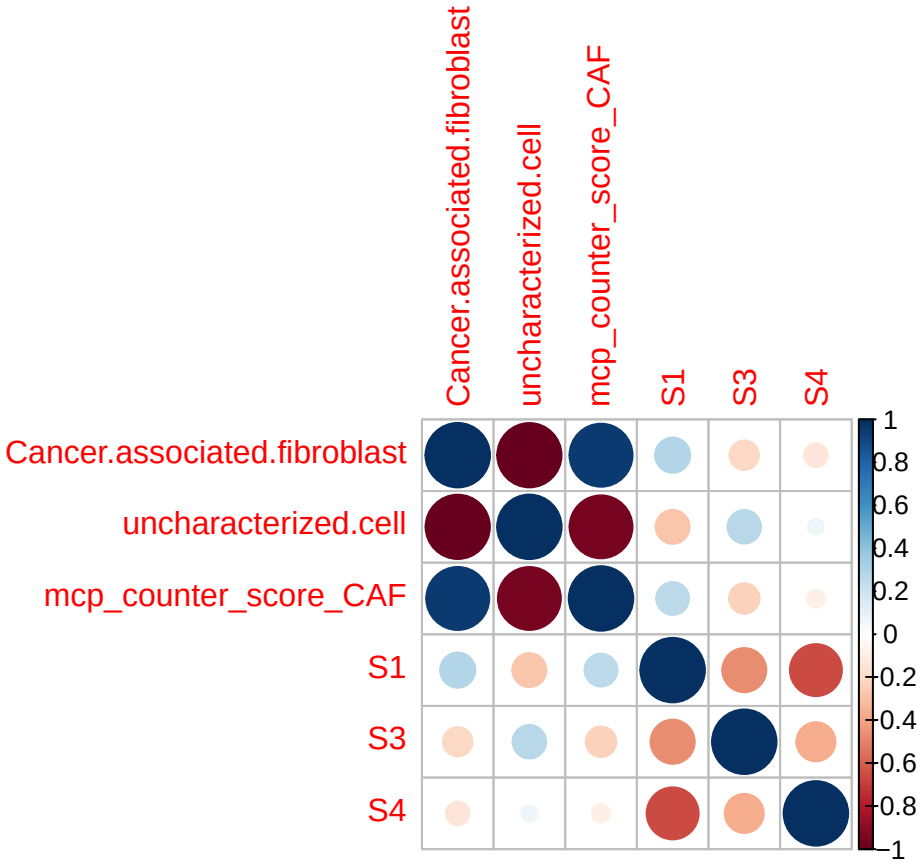
```
deconvolution_mcp_filter
```

```
##    mcp_counter_score_CAF Sample
## 1                2118.187   3532
## 2                2568.089   3533
## 3                2674.509   3536
## 4                2843.439   3537
## 5                2797.128   4027
## 6                2899.125   4028
## 7                2170.477   4033
## 8                2883.104   4034
## 9                3732.086   4112
## 10               4141.333   4113
## 11                2790.22   4116
## 12                2836.76   4117
## 13               3285.481   4214
## 14               3068.483   4215
## 15               2985.898   4299
## 16               4122.946   4300
```

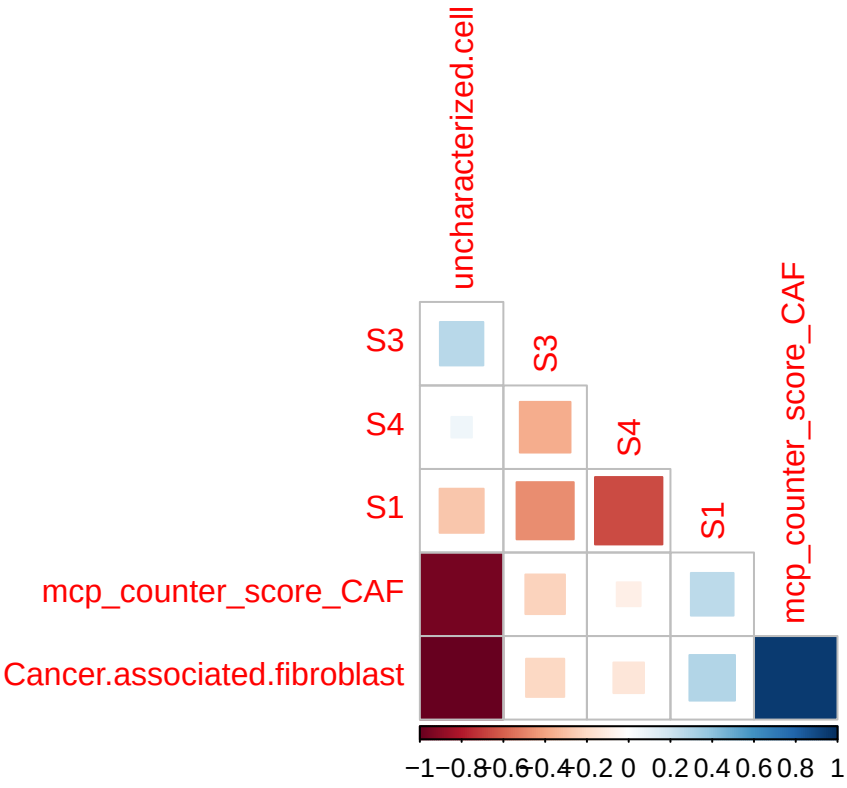
```
## 17              3889.782   4315
## 18              3550.484   4316
## 19              3027.929   4340
## 20              3180.459   4341
## 21              3035.335   4344
## 22              2966.457   4345
## 23              2786.021   4722
## 24              3699.696   4723
```

```
deconvolution_epic_cibersort_mcp_combined <- full_join(deconvolution_epic_filter, deconvolution_mcp_filt
deconvolution_epic_cibersort_mcp_combined <- full_join(deconvolution_epic_cibersort_mcp_combined, ciberso
rownames(deconvolution_epic_cibersort_mcp_combined) <- deconvolution_epic_cibersort_mcp_combined$Sample
deconvolution_epic_cibersort_mcp_combined <- deconvolution_epic_cibersort_mcp_combined %>% dplyr::select
```

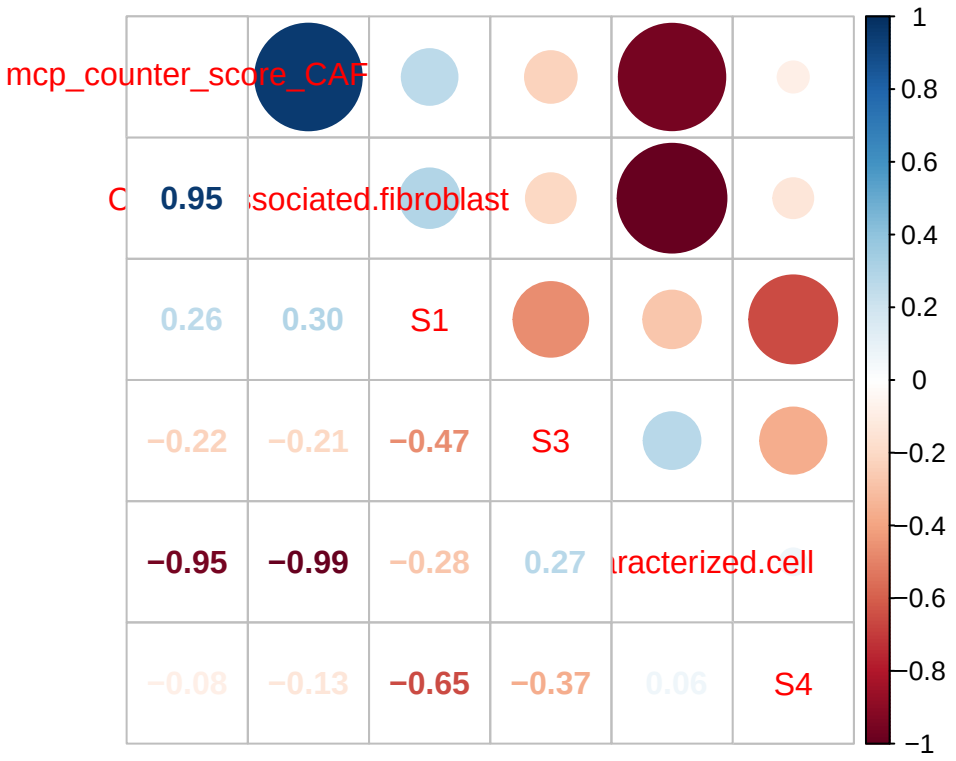
```
corrplot(cor(deconvolution_epic_cibersort_mcp_combined))
```




```
corrplot(cor(deconvolution_epic_cibersort_mcp_combined), method = 'square', order = 'FPC', type = 'lowe
```




```
corrplot.mixed(cor(deconvolution_epic_cibersort_mcp_combined), order = 'AOE')
```




```
deconvolution_estimate <- immunedeconv::deconvolute_estimate(salmon_tpm_hgnc_not_duplicated)
```

```
## [1] "Merged dataset includes 9963 genes (449 mismatched)."
## [1] "1 gene set: StromalSignature  overlap= 136"
## [1] "2 gene set: ImmuneSignature  overlap= 140"
```

```
deconvolution_estimate
```

```
##                        3532          3533        3536        3537          4027
## StromalScore   -398.9779264   -77.0901415  -367.80715  -234.41848  -421.6918121
## ImmuneScore   -1357.1717989 -1241.8458068 -1266.27719 -1320.35708 -1187.1899556
## ESTIMATEScore -1756.1497253 -1318.9359483 -1634.08434 -1554.77556 -1608.8817677
## TumorPurity       0.9403275     0.9165688     0.93408     0.92986     0.9327527
##                        4028          4033          4034          4112
## StromalScore   -310.5104970  -145.7250766   -10.3216652   -20.8256033
## ImmuneScore   -1313.4536825 -1220.6344024 -1155.0647891 -1235.5174420
## ESTIMATEScore -1623.9641796 -1366.3594790 -1165.3864543 -1256.3430453
## TumorPurity       0.9335485     0.9193302     0.9073238     0.9128561
##                        4113          4116          4117          4214
## StromalScore    152.0644980  -300.7698916  -249.6659772   -96.2201331
## ImmuneScore   -1139.3841557 -1288.4770855 -1219.8616524 -1155.4397974
## ESTIMATEScore  -987.3196577 -1589.2469771 -1469.5276296 -1251.6599305
## TumorPurity       0.8960258     0.9317097     0.9251835     0.9125752
##                        4215          4299          4300          4315
## StromalScore   -172.1903869  -236.2956163   -71.9668837    21.6951624
```

```
## ImmuneScore   -1212.7012973 -1171.9712346 -1201.7329345 -1142.4733879
## ESTIMATEScore -1384.8916842 -1408.2668509 -1273.6998182 -1120.7782255
## TumorPurity       0.9203972     0.9217333     0.9138934     0.9045514
##                        4316          4340          4341          4344
## StromalScore     30.4199104  -274.1544943  -173.9367732  -255.1282555
## ImmuneScore   -1138.2189328 -1236.3081962 -1193.6425982 -1300.0417197
## ESTIMATEScore -1107.7990224 -1510.4626906 -1367.5793713 -1555.1699753
## TumorPurity       0.9037374     0.9274472     0.9194006     0.9298813
##                        4345          4722          4723
## StromalScore   -367.5866878  -240.7848074  -165.711409
## ImmuneScore   -1326.4001046  -981.5805628 -1159.640857
## ESTIMATEScore -1693.9867924 -1222.3653702 -1325.352266
## TumorPurity       0.9371835     0.9108085     0.916945
```

```
deconvolution_xcell_filtered_long_metadata_caf <- deconvolution_xcell_filtered_long_metadata %>% dplyr:
summary(deconvolution_xcell_filtered_long_metadata_caf)
```

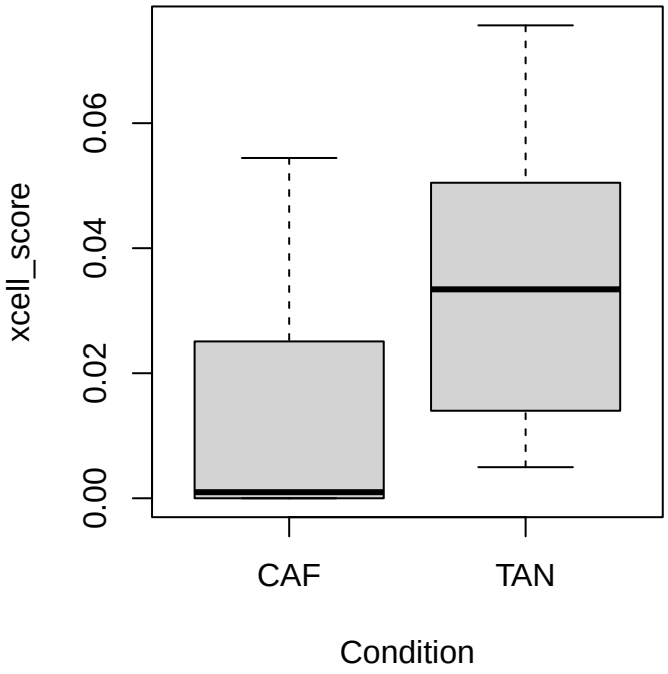
```
##   cell_type            sample           xcell_score          Patient
##  Length:24          Length:24          Min.   :0.000000   Min.   : 1.00
##  Class :character   Class :character   1st Qu.:0.001253   1st Qu.: 3.75
##  Mode  :character   Mode  :character   Median :0.014000   Median : 6.50
##                                        Mean   :0.023543   Mean   : 6.50
##                                        3rd Qu.:0.043588   3rd Qu.: 9.25
##                                        Max.   :0.075644   Max.   :12.00
##   Condition             Age              Size           Grade
##  Length:24          Min.   :45.00   Min.   :  8.00   Length:24
##  Class :character   1st Qu.:49.50   1st Qu.: 26.50   Class :character
##  Mode  :character   Median :62.00   Median : 40.00   Mode  :character
##                     Mean   :63.92   Mean   : 43.58
##                     3rd Qu.:78.75   3rd Qu.: 52.00
##                     Max.   :84.00   Max.   :100.00
##   Histology             ER                 PR                Her2
##  Length:24          Length:24          Length:24          Length:24
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##    Subtype              LVI
##  Length:24          Length:24
##  Class :character   Class :character
##  Mode  :character   Mode  :character
##
##
##
```

```
caf_xcell <- deconvolution_xcell_filtered_long_metadata_caf$xcell_score[which(deconvolution_xcell_filte
tan_xcell <- deconvolution_xcell_filtered_long_metadata_caf$xcell_score[which(deconvolution_xcell_filte
cor.test(x = caf_xcell, y = tan_xcell,
         method = c("pearson"),
         conf.level = 0.95)
```

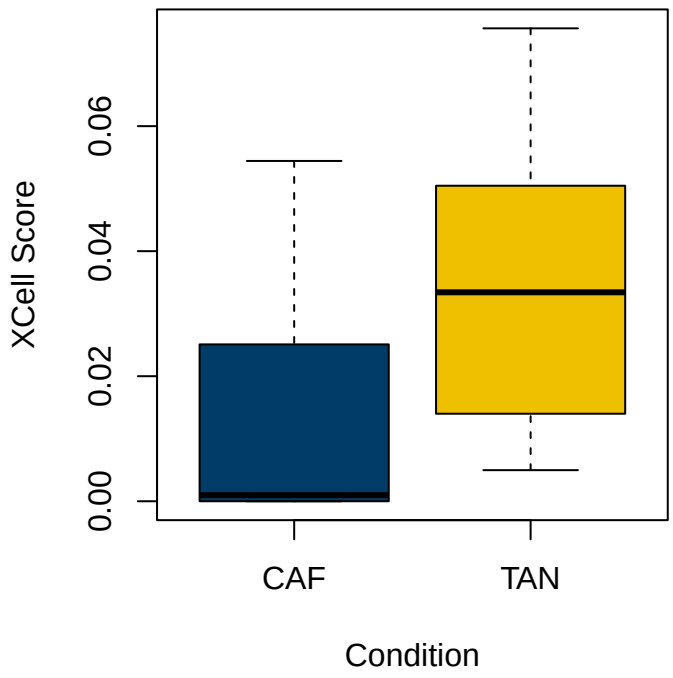
```
##
```

```
##  Pearson's product-moment correlation
##
## data:  caf_xcell and tan_xcell
## t = 3.0238, df = 10, p-value = 0.01281
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1942461 0.9057587
## sample estimates:
##       cor
## 0.6911047
```

```
par(pty = "s")
boxplot(xcell_score ~ Condition, data = deconvolution_xcell_filtered_long_metadata_caf)
```




```
boxplot(xcell_score ~ Condition,
        col = c("#003C67FF", "#EFC000FF"),
        main = "XCell CAF score InHouse CAF vs TAN",
        xlab = "Condition", ylab = "XCell Score", data = deconvolution_xcell_filtered_long_metadata_caf)
```




```
diff = caf_xcell - tan_xcell
summary(diff)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -0.050423 -0.034136 -0.022248 -0.021894 -0.011600  0.008418
```

```
bartlett.test(xcell_score ~ Condition, data = deconvolution_xcell_filtered_long_metadata_caf)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  xcell_score by Condition
## Bartlett's K-squared = 0.1614, df = 1, p-value = 0.6879
```

```
t.test(formula = xcell_score ~ Condition,
       alternative = "two.sided",
       mu = 0,
       paired = TRUE,
       var.equal = TRUE,
       conf.level = 0.95, data = deconvolution_xcell_filtered_long_metadata_caf)
```

```
##
##  Paired t-test
##
```

```
## data:  xcell_score by Condition
## t = -4.4452, df = 11, p-value = 0.0009865
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -0.03273521 -0.01105366
## sample estimates:
## mean difference
##     -0.02189443
```

```
deconvolution_epic_filtered_long_metadata
```

```
## # A tibble: 48 x 14
##    cell_type     sample epic_score Patient Condition   Age  Size Grade Histology
##    <chr>         <chr>       <dbl>   <int> <chr>     <int> <int> <chr> <chr>
##  1 Cancer assoc~ 3532        0.388       9 CAF          48    16 Grad~ Ductal
##  2 Cancer assoc~ 3533        0.446       9 TAN          48    16 Grad~ Ductal
##  3 Cancer assoc~ 3536        0.549      10 CAF          50    52 Grad~ Lobular
##  4 Cancer assoc~ 3537        0.601      10 TAN          50    52 Grad~ Lobular
##  5 Cancer assoc~ 4027        0.483       2 CAF          77    40 Grad~ Ductal
##  6 Cancer assoc~ 4028        0.531       2 TAN          77    40 Grad~ Ductal
##  7 Cancer assoc~ 4033        0.403       1 CAF          46    45 Grad~ Lobular
##  8 Cancer assoc~ 4034        0.549       1 TAN          46    45 Grad~ Lobular
##  9 Cancer assoc~ 4112        0.833       3 CAF          62    12 Grad~ Ductal
## 10 Cancer assoc~ 4113        0.983       3 TAN          62     8 Grad~ Ductal
## # i 38 more rows
## # i 5 more variables: ER <chr>, PR <chr>, Her2 <chr>, Subtype <chr>, LVI <chr>
```

```
deconvolution_epic_filtered_long_metadata_caf <- deconvolution_epic_filtered_long_metadata %>% dplyr::f
summary(deconvolution_epic_filtered_long_metadata_caf)
```

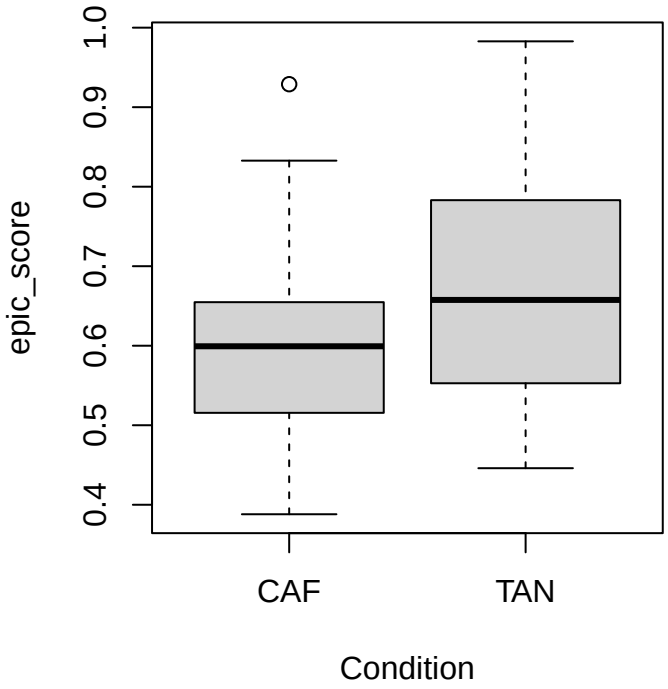
```
##   cell_type            sample            epic_score        Patient
##  Length:24          Length:24          Min.   :0.3881   Min.   : 1.00
##  Class :character   Class :character   1st Qu.:0.5489   1st Qu.: 3.75
##  Mode  :character   Mode  :character   Median :0.6104   Median : 6.50
##                                        Mean   :0.6468   Mean   : 6.50
##                                        3rd Qu.:0.7347   3rd Qu.: 9.25
##                                        Max.   :0.9828   Max.   :12.00
##   Condition             Age              Size           Grade
##  Length:24          Min.   :45.00   Min.   :  8.00   Length:24
##  Class :character   1st Qu.:49.50   1st Qu.: 26.50   Class :character
##  Mode  :character   Median :62.00   Median : 40.00   Mode  :character
##                     Mean   :63.92   Mean   : 43.58
##                     3rd Qu.:78.75   3rd Qu.: 52.00
##                     Max.   :84.00   Max.   :100.00
##   Histology             ER                 PR                Her2
##  Length:24          Length:24          Length:24          Length:24
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##    Subtype              LVI
##  Length:24          Length:24
```

```
##  Class :character   Class :character
##  Mode  :character   Mode  :character
##
##
##
```

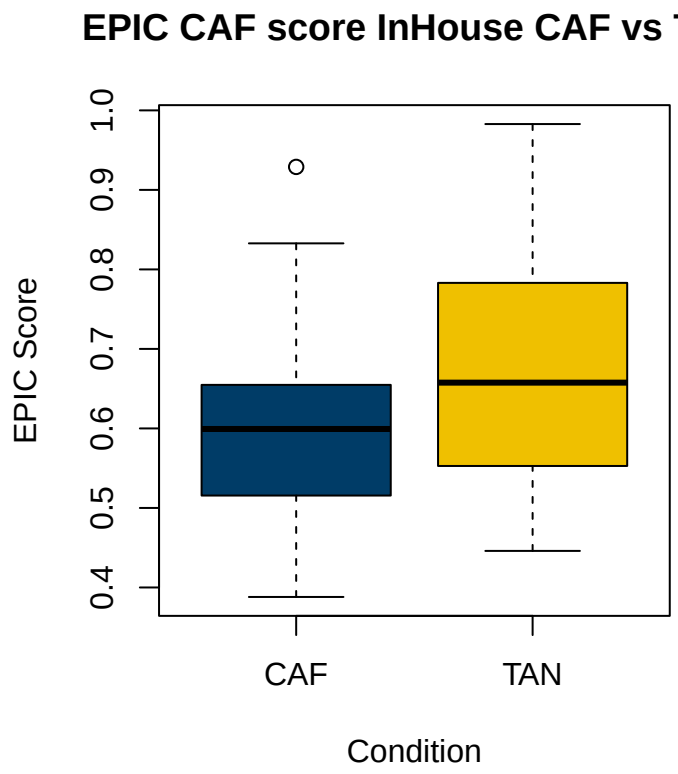
```
caf_epic <- deconvolution_epic_filtered_long_metadata_caf$epic_score[which(deconvolution_epic_filtered_l
tan_epic <- deconvolution_epic_filtered_long_metadata_caf$epic_score[which(deconvolution_epic_filtered_l
cor.test(x = caf_epic, y = tan_epic,
         method = c("pearson"),
         conf.level = 0.95)
```

```
##
##  Pearson's product-moment correlation
##
## data:  caf_epic and tan_epic
## t = 2.9167, df = 10, p-value = 0.01539
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1703742 0.9012231
## sample estimates:
##       cor
## 0.6779839
```

```
par(pty = "s")
boxplot(epic_score ~ Condition, data = deconvolution_epic_filtered_long_metadata_caf)
```



```
boxplot(epic_score ~ Condition,
        col = c("#003C67FF", "#EFC000FF"),
        main = "EPIC CAF score InHouse CAF vs TAN",
        xlab = "Condition", ylab = "EPIC Score", data = deconvolution_epic_filtered_long_metadata_caf)
```




```
bartlett.test(epic_score ~ Condition, data = deconvolution_epic_filtered_long_metadata_caf)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  epic_score by Condition
## Bartlett's K-squared = 0.075341, df = 1, p-value = 0.7837
```

```
t.test(formula = epic_score ~ Condition,
       alternative = "two.sided",
       mu = 0,
       paired = TRUE,
       var.equal = TRUE,
       conf.level = 0.95, data = deconvolution_epic_filtered_long_metadata_caf)
```

```
##
##  Paired t-test
##
## data:  epic_score by Condition
```

```
## t = -2.1183, df = 11, p-value = 0.05774
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -0.164548496  0.003148298
## sample estimates:
## mean difference
##      -0.0807001
```

```
deconvolution_mcp_filtered_long_metadata
```

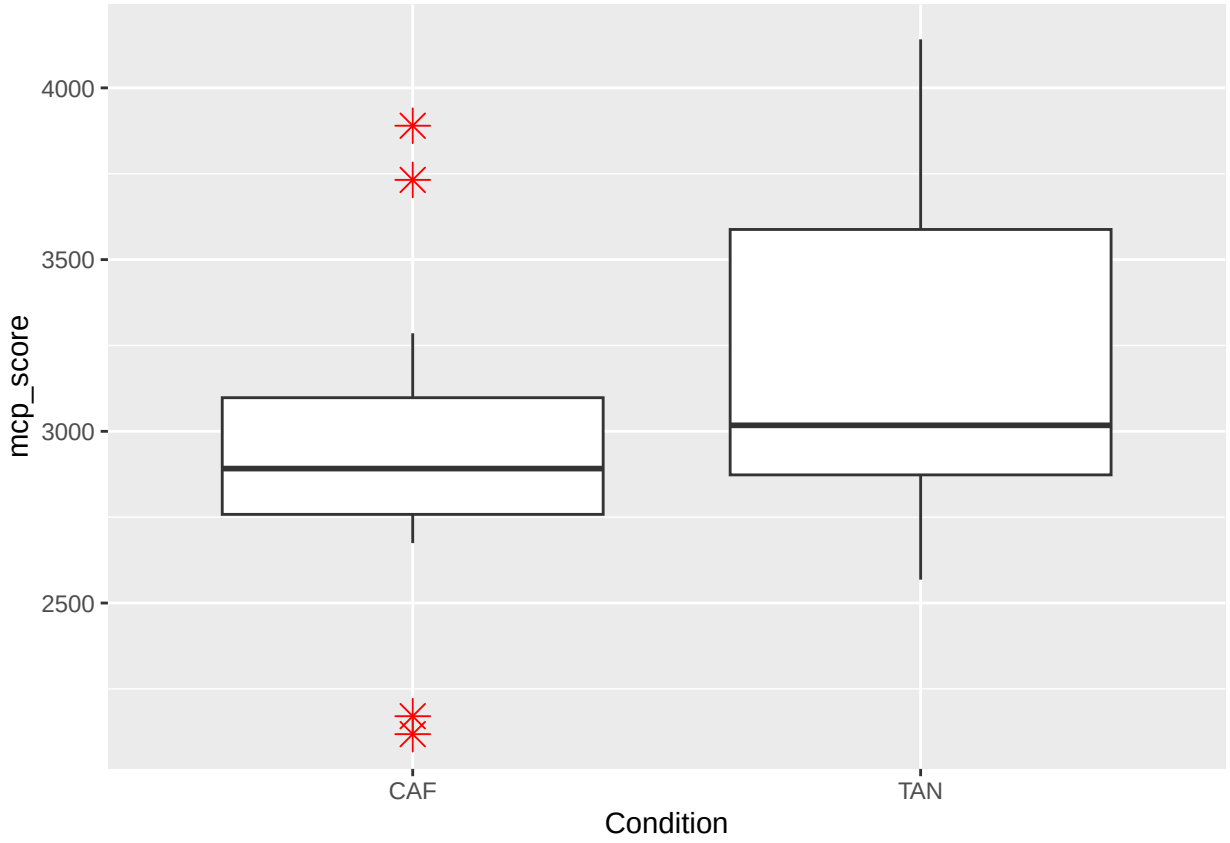
```
## # A tibble: 48 x 14
##    cell_type      sample mcp_score Patient Condition   Age  Size Grade Histology
##    <chr>          <chr>      <dbl>   <int> <chr>     <int> <int> <chr> <chr>
##  1 Endothelial c~ 3532       14.2        9 CAF          48    16 Grad~ Ductal
##  2 Endothelial c~ 3533       11.4        9 TAN          48    16 Grad~ Ductal
##  3 Endothelial c~ 3536       10.0       10 CAF          50    52 Grad~ Lobular
##  4 Endothelial c~ 3537        9.13      10 TAN          50    52 Grad~ Lobular
##  5 Endothelial c~ 4027       16.6        2 CAF          77    40 Grad~ Ductal
##  6 Endothelial c~ 4028       11.9        2 TAN          77    40 Grad~ Ductal
##  7 Endothelial c~ 4033       11.6        1 CAF          46    45 Grad~ Lobular
##  8 Endothelial c~ 4034       12.8        1 TAN          46    45 Grad~ Lobular
##  9 Endothelial c~ 4112       10.9        3 CAF          62    12 Grad~ Ductal
## 10 Endothelial c~ 4113       11.2        3 TAN          62     8 Grad~ Ductal
## # i 38 more rows
## # i 5 more variables: ER <chr>, PR <chr>, Her2 <chr>, Subtype <chr>, LVI <chr>
```

```
deconvolution_compare_fibroblast_score_caf_tan <- function(df, score_column) {
  df_caf <- df %>% dplyr::filter(cell_type == "Cancer associated fibroblast")
  caf_score <- df_caf[[score_column]][which(df_caf$Condition == "CAF")]
  tan_score <- df_caf[[score_column]][which(df_caf$Condition == "TAN")]
  print("carrying out correlation test...")
  print(cor.test(x = caf_score, y = tan_score,
         method = c("pearson"),
         conf.level = 0.95))
  plt_title <- paste(score_column, " cancer-associated fibroblast for inhouse CAF vs TAN")
  par(pty = "s")
  #boxplot(score_column ~ Condition, data = df_caf)
  #plot_out <- boxplot(df_caf[,score_column] ~ df_caf[,Condition],
   #     col = c("#003C67FF", "#EFC000FF"),
    #    main = plt_title,
     #   xlab = "Condition", ylab = score_column)
  print(df_caf)
  plot_out <- ggplot(data = df_caf, aes(Condition, .data[[score_column]])) +
              geom_boxplot(outlier.colour="red", outlier.shape=8, outlier.size=4)

  bartlet_out <- bartlett.test(df_caf[[score_column]] ~ df_caf$Condition, data = df_caf)
  if (bartlet_out$p.value > 0.05){
    print("variances equal, carry out t-test")
    t.test.out <- t.test(formula = df_caf[[score_column]] ~ df_caf$Condition,
       alternative = "two.sided",
       mu = 0,
       paired = TRUE,
       var.equal = TRUE,
```

```
      conf.level = 0.95, data = df_caf)
  }
  outputs <- list(plot_out, bartlet_out, t.test.out)

  return(outputs)
}

out_mcp <- deconvolution_compare_fibroblast_score_caf_tan(deconvolution_mcp_filtered_long_metadata, "mcp
```

```
## [1] "carrying out correlation test..."
##
##  Pearson's product-moment correlation
##
## data:  caf_score and tan_score
## t = 2.5875, df = 10, p-value = 0.02707
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.09324848 0.88538684
## sample estimates:
##       cor
## 0.6332609
##
## # A tibble: 24 x 14
##    cell_type      sample mcp_score Patient Condition   Age  Size Grade Histology
##    <chr>          <chr>      <dbl>   <int> <chr>     <int> <int> <chr> <chr>
##  1 Cancer associ~ 3532       2118.       9 CAF          48    16 Grad~ Ductal
##  2 Cancer associ~ 3533       2568.       9 TAN          48    16 Grad~ Ductal
##  3 Cancer associ~ 3536       2675.      10 CAF          50    52 Grad~ Lobular
##  4 Cancer associ~ 3537       2843.      10 TAN          50    52 Grad~ Lobular
##  5 Cancer associ~ 4027       2797.       2 CAF          77    40 Grad~ Ductal
##  6 Cancer associ~ 4028       2899.       2 TAN          77    40 Grad~ Ductal
##  7 Cancer associ~ 4033       2170.       1 CAF          46    45 Grad~ Lobular
##  8 Cancer associ~ 4034       2883.       1 TAN          46    45 Grad~ Lobular
##  9 Cancer associ~ 4112       3732.       3 CAF          62    12 Grad~ Ductal
## 10 Cancer associ~ 4113       4141.       3 TAN          62     8 Grad~ Ductal
## # i 14 more rows
## # i 5 more variables: ER <chr>, PR <chr>, Her2 <chr>, Subtype <chr>, LVI <chr>
## [1] "variances equal, carry out t-test"
```

```
out_mcp[[1]]
```




```
out_mcp[[2]]
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  df_caf[[score_column]] by df_caf$Condition
## Bartlett's K-squared = 0.00074082, df = 1, p-value = 0.9783
```

```
out_mcp[[3]]
```

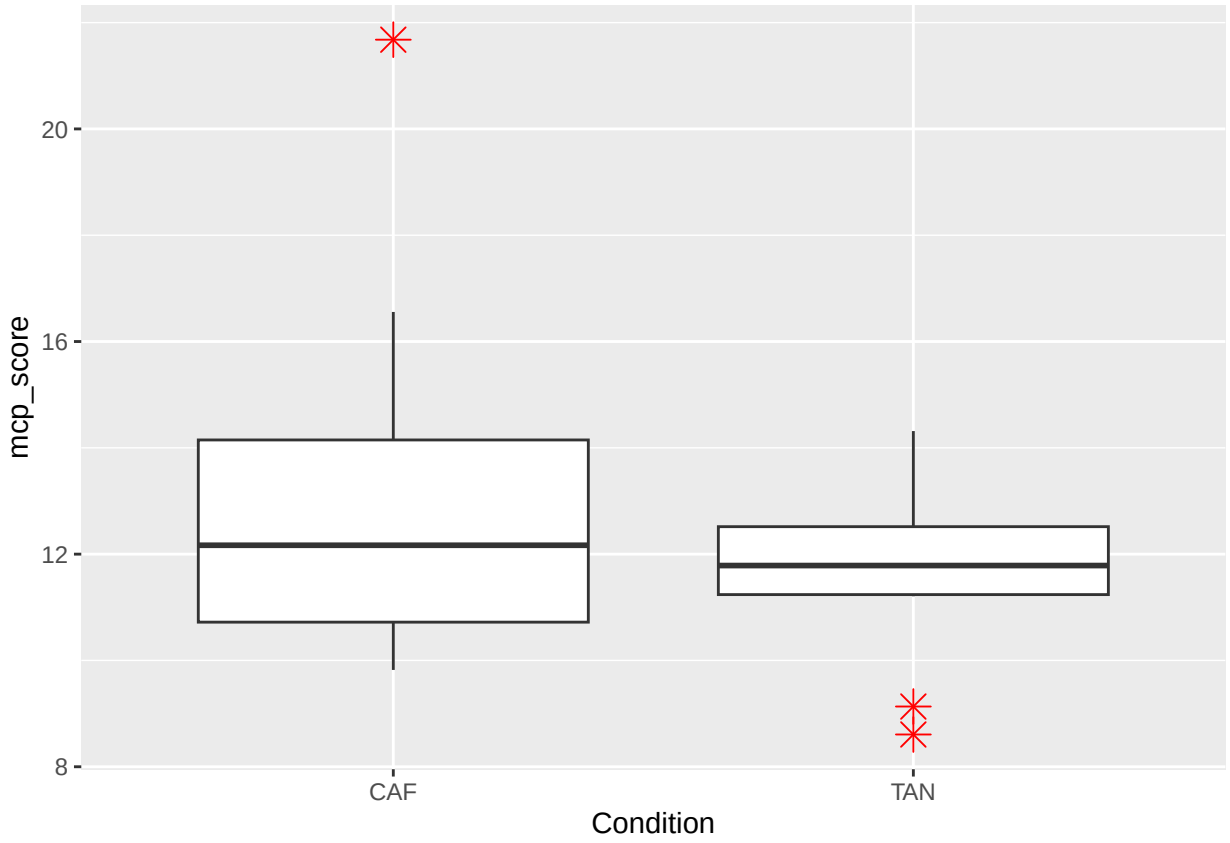
```
##
##  Paired t-test
##
## data:  df_caf[[score_column]] by df_caf$Condition
## t = -2.2206, df = 11, p-value = 0.04831
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -575.333038   -2.553832
## sample estimates:
## mean difference
##       -288.9434
```

```
deconvolution_compare_celltype_score_caf_tan <- function(df, score_column, cell_type_in) {
  df_cell_type <- df %>% dplyr::filter(cell_type == cell_type_in)
```

```
    caf_score <- df_cell_type[[score_column]][which(df_cell_type$Condition == "CAF")]
  tan_score <- df_cell_type[[score_column]][which(df_cell_type$Condition == "TAN")]
  print("carrying out correlation test...")
  print(cor.test(x = caf_score, y = tan_score,
         method = c("pearson"),
         conf.level = 0.95))
  plot_out <- ggplot(data = df_cell_type, aes(Condition, .data[[score_column]])) +
              geom_boxplot(outlier.colour="red", outlier.shape=8, outlier.size=4)

  bartlet_out <- bartlett.test(df_cell_type[[score_column]] ~ df_cell_type$Condition, data = df_cell_ty
  print(bartlet_out)
  if (bartlet_out$p.value > 0.05) {
    print("variances equal, carry out t-test equal variance")
    t.test.out <- t.test(formula = df_cell_type[[score_column]] ~ df_cell_type$Condition,
       alternative = "two.sided",
       mu = 0,
       paired = TRUE,
       var.equal = TRUE,
       conf.level = 0.95, data = df_cell_type)
   } else {
      print("variances not equal, carry out t-test unequal variance")
      t.test.out <- t.test(formula = df_cell_type[[score_column]] ~ df_cell_type$Condition,
       alternative = "two.sided",
       mu = 0,
       paired = TRUE,
       var.equal = FALSE,
       conf.level = 0.95, data = df_cell_type)
    }

  outputs <- list(plot_out, bartlet_out, t.test.out)
  return(outputs)
}
deconvolution_compare_celltype_score_caf_tan(deconvolution_mcp_filtered_long_metadata, "mcp_score", "En
```

```
## [1] "carrying out correlation test..."
##
##  Pearson's product-moment correlation
##
## data:  caf_score and tan_score
## t = 1.437, df = 10, p-value = 0.1813
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.2100771  0.7981123
## sample estimates:
##       cor
## 0.4137019
##
##
##  Bartlett test of homogeneity of variances
##
## data:  df_cell_type[[score_column]] by df_cell_type$Condition
## Bartlett's K-squared = 5.9468, df = 1, p-value = 0.01474
##
```

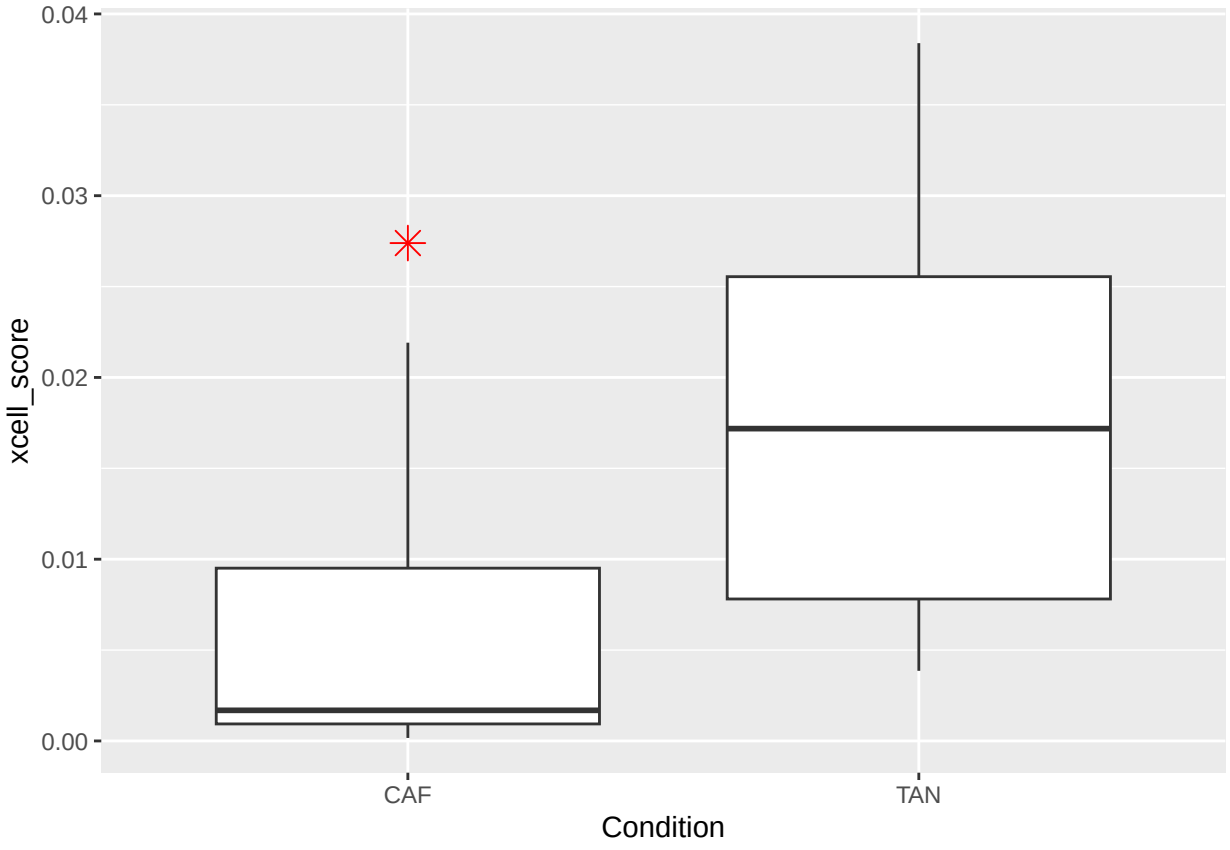
```
## [1] "variances not equal, carry out t-test unequal variance"

## [[1]]
```




```
##
## [[2]]
##
##  Bartlett test of homogeneity of variances
##
## data:  df_cell_type[[score_column]] by df_cell_type$Condition
## Bartlett's K-squared = 5.9468, df = 1, p-value = 0.01474
##
##
## [[3]]
##
##  Paired t-test
##
## data:  df_cell_type[[score_column]] by df_cell_type$Condition
## t = 1.5153, df = 11, p-value = 0.1579
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -0.6135207  3.3251092
## sample estimates:
## mean difference
##        1.355794
```

```
deconvolution_compare_celltype_score_caf_tan(deconvolution_xcell_filtered_long_metadata, "xcell_score",
```

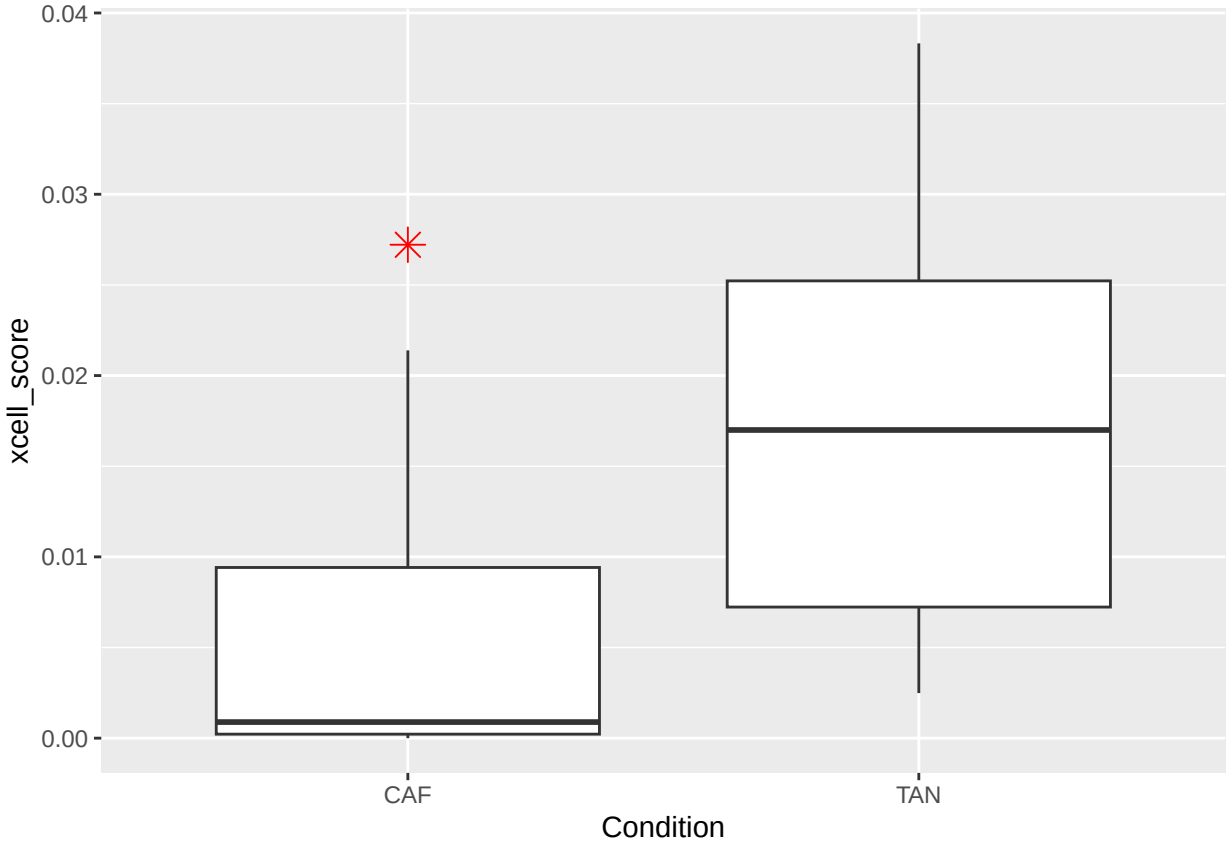
```
## [1] "carrying out correlation test..."
##
##  Pearson's product-moment correlation
##
## data:  caf_score and tan_score
## t = 3.2315, df = 10, p-value = 0.008998
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.2387499 0.9138004
## sample estimates:
##       cor
## 0.7147212
##
##
##  Bartlett test of homogeneity of variances
##
## data:  df_cell_type[[score_column]] by df_cell_type$Condition
## Bartlett's K-squared = 0.15337, df = 1, p-value = 0.6953
##
## [1] "variances equal, carry out t-test equal variance"

## [[1]]
```



```
##
## [[2]]
##
##  Bartlett test of homogeneity of variances
##
## data:  df_cell_type[[score_column]] by df_cell_type$Condition
## Bartlett's K-squared = 0.15337, df = 1, p-value = 0.6953
##
##
## [[3]]
##
##  Paired t-test
##
## data:  df_cell_type[[score_column]] by df_cell_type$Condition
## t = -4.7058, df = 11, p-value = 0.0006441
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -0.015933207 -0.005778311
## sample estimates:
## mean difference
##     -0.01085576
```

```
deconvolution_compare_celltype_score_caf_tan(deconvolution_xcell_filtered_long_metadata, "xcell_score",
```

```
## [1] "carrying out correlation test..."
##
##  Pearson's product-moment correlation
##
## data:  caf_score and tan_score
## t = 3.018, df = 10, p-value = 0.01294
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1929609 0.9055186
## sample estimates:
##       cor
## 0.6904065
##
##
##  Bartlett test of homogeneity of variances
##
## data:  df_cell_type[[score_column]] by df_cell_type$Condition
## Bartlett's K-squared = 0.20576, df = 1, p-value = 0.6501
##
## [1] "variances equal, carry out t-test equal variance"

## [[1]]
```

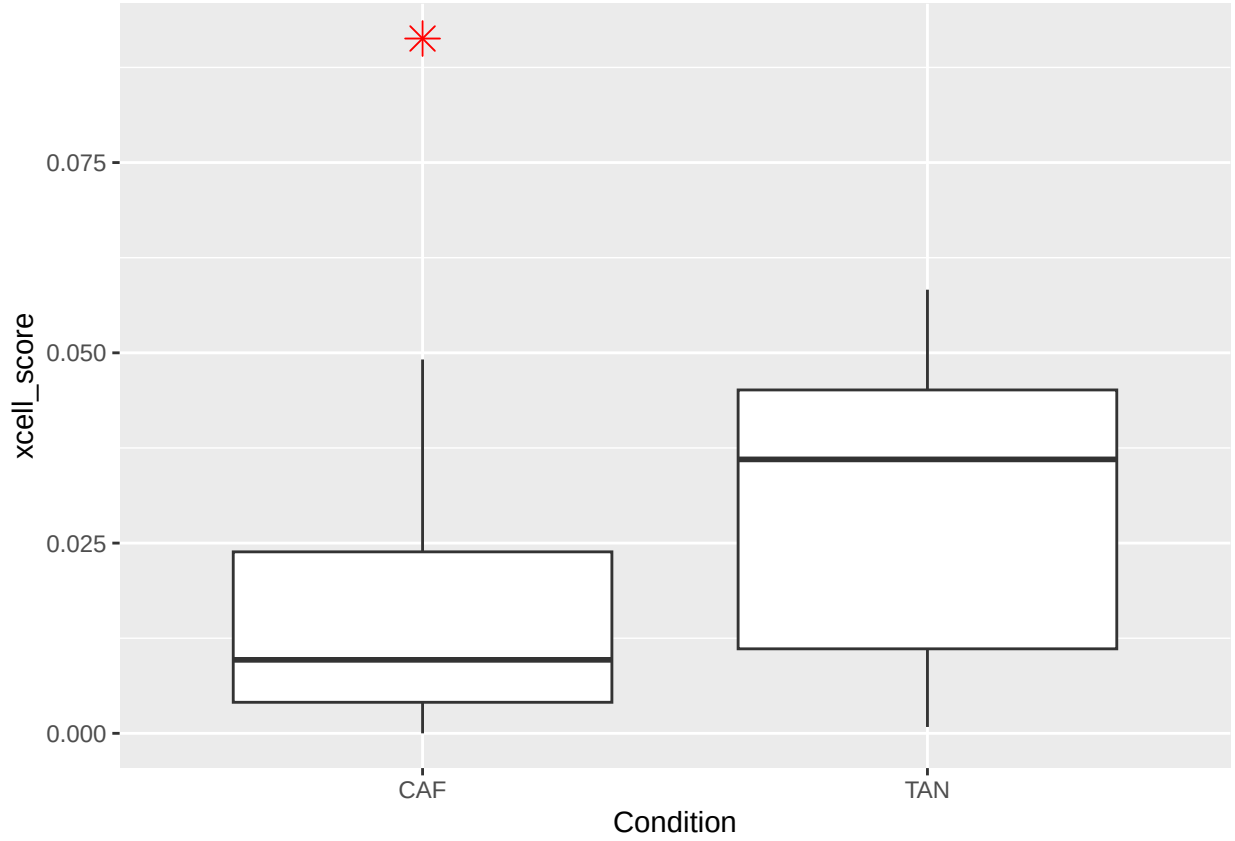



```
##
## [[2]]
##
##  Bartlett test of homogeneity of variances
##
## data:  df_cell_type[[score_column]] by df_cell_type$Condition
## Bartlett's K-squared = 0.20576, df = 1, p-value = 0.6501
##
##
## [[3]]
##
##  Paired t-test
##
## data:  df_cell_type[[score_column]] by df_cell_type$Condition
## t = -4.4415, df = 11, p-value = 0.0009926
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -0.016396379 -0.005530487
## sample estimates:
## mean difference
##     -0.01096343
```

```
deconvolution_compare_celltype_score_caf_tan(deconvolution_xcell_filtered_long_metadata, "xcell_score",
```

```
## [1] "carrying out correlation test..."
```

```
##
##  Pearson's product-moment correlation
##
## data:  caf_score and tan_score
## t = 0.66391, df = 10, p-value = 0.5218
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.4176870  0.6971612
## sample estimates:
##       cor
## 0.2054675
##
##
##  Bartlett test of homogeneity of variances
##
## data:  df_cell_type[[score_column]] by df_cell_type$Condition
## Bartlett's K-squared = 0.85096, df = 1, p-value = 0.3563
##
## [1] "variances equal, carry out t-test equal variance"

## [[1]]
```




```
##
## [[2]]
##
```

```
##  Bartlett test of homogeneity of variances
##
## data:  df_cell_type[[score_column]] by df_cell_type$Condition
## Bartlett's K-squared = 0.85096, df = 1, p-value = 0.3563
##
##
## [[3]]
##
##  Paired t-test
##
## data:  df_cell_type[[score_column]] by df_cell_type$Condition
## t = -1.1102, df = 11, p-value = 0.2906
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -0.028307954  0.009325737
## sample estimates:
## mean difference
##    -0.009491109
```

```
metadata
```

```
##    sample Patient Condition Age Size   Grade Histology          ER          PR
## 1    4033       1       CAF  46   45 Grade_2   Lobular ER_positive PR_positive
## 2    4034       1       TAN  46   45 Grade_2   Lobular ER_positive PR_positive
## 3    4027       2       CAF  77   40 Grade_3    Ductal ER_negative PR_negative
## 4    4028       2       TAN  77   40 Grade_3    Ductal ER_negative PR_negative
## 5    4112       3       CAF  62   12 Grade_3    Ductal ER_positive PR_negative
## 6    4113       3       TAN  62    8 Grade_3    Ductal ER_positive PR_negative
## 7    4116       4       CAF  45   35 Grade_2   Lobular ER_positive PR_positive
## 8    4117       4       TAN  45   13 Grade_2   Lobular ER_positive PR_positive
## 9    4214       5       CAF  78   90 Grade_2   Lobular ER_positive PR_negative
## 10   4215       5       TAN  78   90 Grade_2   Lobular ER_positive PR_negative
## 11   4315       6       CAF  84   30 Grade_2    Ductal ER_positive PR_positive
## 12   4316       6       TAN  84   22 Grade_2    Ductal ER_positive PR_positive
## 13   4340       7       CAF  62  100 Grade_2   Lobular ER_positive PR_positive
## 14   4341       7       TAN  62  100 Grade_2   Lobular ER_positive PR_positive
## 15   4344       8       CAF  50   28 Grade_2    Ductal ER_positive PR_positive
## 16   4345       8       TAN  50   28 Grade_2    Ductal ER_positive PR_positive
## 17   3532       9       CAF  48   16 Grade_2    Ductal ER_positive PR_positive
## 18   3533       9       TAN  48   16 Grade_2    Ductal ER_positive PR_positive
## 19   3536      10       CAF  50   52 Grade_3   Lobular ER_positive PR_positive
## 20   3537      10       TAN  50   52 Grade_3   Lobular ER_positive PR_positive
## 21   4299      11       CAF  84   40 Grade_3    Ductal ER_positive PR_positive
## 22   4300      11       TAN  84   40 Grade_3    Ductal ER_positive PR_positive
## 23   4722      12       CAF  81   52 Grade_2   Lobular ER_positive PR_negative
## 24   4723      12       TAN  81   52 Grade_2   Lobular ER_positive PR_negative
##             Her2  Subtype          LVI
## 1  Her2_negative LuminalA LVI_negative
## 2  Her2_negative LuminalA LVI_negative
## 3  Her2_negative     TNBC LVI_positive
## 4  Her2_negative     TNBC LVI_positive
## 5  Her2_negative LuminalA LVI_negative
## 6  Her2_negative LuminalA LVI_negative
## 7  Her2_negative LuminalA LVI_positive
```

```
## 8  Her2_negative LuminalA LVI_positive
## 9  Her2_negative LuminalA LVI_positive
## 10 Her2_negative LuminalA LVI_positive
## 11 Her2_negative LuminalA LVI_positive
## 12 Her2_negative LuminalA LVI_positive
## 13 Her2_negative LuminalA LVI_negative
## 14 Her2_negative LuminalA LVI_negative
## 15 Her2_negative LuminalA LVI_negative
## 16 Her2_negative LuminalA LVI_negative
## 17 Her2_negative LuminalA LVI_negative
## 18 Her2_negative LuminalA LVI_negative
## 19 Her2_negative LuminalA LVI_negative
## 20 Her2_negative LuminalA LVI_negative
## 21 Her2_negative LuminalA LVI_positive
## 22 Her2_negative LuminalA LVI_positive
## 23 Her2_negative LuminalA LVI_negative
## 24 Her2_negative LuminalA LVI_negative
```

```
salmon_markers <- data.frame(t(salmon_tpm_hgnc_not_duplicated[c("ACTA2","FAP"),]))
salmon_markers$sample <- rownames(salmon_markers)
metadata_markers <- full_join(metadata, salmon_markers)
```

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

```
# svg("/home/rstudio/Documents/PhD/notes/mini_viva_report/images/current_research/figure3/xcell_deconv_
# xcell_plot
# dev.off()
```

```
#svg("/home/rstudio/Documents/PhD/notes/mini_viva_report/images/current_research/figure3/epic_deconv_ca
# epic_plot_stack
# dev.off()
```

```
# svg("/home/rstudio/Documents/PhD/notes/mini_viva_report/images/current_research/figure3/mcp_deconv_ca
# mcp_plot
# dev.off()
```