Comparison of Hot and Cold Tumors

Gepoliano Chaves (Ph. D.)

December 16th, 2023

To prepare material for this notebook I used notebooks ADRN-MES-Sengupta-Final.Rmd and "Survial MES.Rmd".

1) Library installation

```
library(ggplot2)
library(GSVA)
library(GSVAdata)
library(data.table)
library(stringr)
library(tibble)
library(viridis)
library(expss)
library(ggpubr)
```

2) Gene Expression and GSVA

Load data-frames

```
GSE49711 <- readRDS("../../../../../Hypoxia_Paper/for loop MES/data/GSE49711.rds")
GSE49711_GeneMatrix <- as.matrix(GSE49711)
```

Load gene sets

```
cfDNA_PCA_gene_list <- getGmt("~/Desktop/Gepoliano/Analysis/cfDNA_genes_PCA.txt")
```

```
## Warning in readLines(con, ...): incomplete final line found on
## '~/Desktop/Gepoliano/Analysis/cfDNA_genes_PCA.txt'
```

```
## Warning in getGmt("~/Desktop/Gepoliano/Analysis/cfDNA_genes_PCA.txt"): 323
## record(s) contain duplicate ids: ADRN_Gene_List_373, ADRN_Gronigen, ...,
## WP_GLYCOLYSIS_AND_GLUCONEOGENESIS_7, WP_GLYCOLYSIS_IN_SENESCENCE_Genes_4
```

Construct GSVA data-frame

```
# Needs to re-run this step if there are new gene sets
GSE49711_GSVA <- gsva(GSE49711_GeneMatrix,
                      cfDNA_PCA_gene_list,
                      min.sz=1, max.sz=Inf,
                      verbose=TRUE)
```

```
## Warning: useNames = NA is deprecated. Instead, specify either useNames = TRUE
## or useNames = TRUE.
```

```
## Warning in .filterFeatures(expr, method): 755 genes with constant expression
## values throuhgout the samples.
```

```
## Warning in .filterFeatures(expr, method): Since argument method!="ssgsea",
## genes with constant expression values are discarded.
```

```
## Warning in .gsva(expr, mapped.gset.idx.list, method, kcdf, rnaseq, abs.ranking,
## : Some gene sets have size one. Consider setting 'min.sz > 1'.
```

```
## Estimating GSVA scores for 314 gene sets.
## Estimating ECDFs with Gaussian kernels
##   |                                                                    |
```

```
GSVA_GeneExpression_Matrix <- rbind(GSE49711_GSVA, GSE49711_GeneMatrix)
```

2) Process GSVA DF with Phenotype Scores

Previous chunk used was from notebook "Multiple Test Correction Kocak.Rmd". Below we will transpose and Name Column with Sample Names

```
GSVA_GeneExpression_Matrix_t <- t(GSVA_GeneExpression_Matrix)
GSVA_GeneExpression_Matrix_t_df <- as.data.frame(GSVA_GeneExpression_Matrix_t)
GSVA_GeneExpression_Matrix_t_df["Sample_ID"] <- rownames(GSVA_GeneExpression_Matrix_t_df)
```

2.1) Inflammatory Response vs Hot
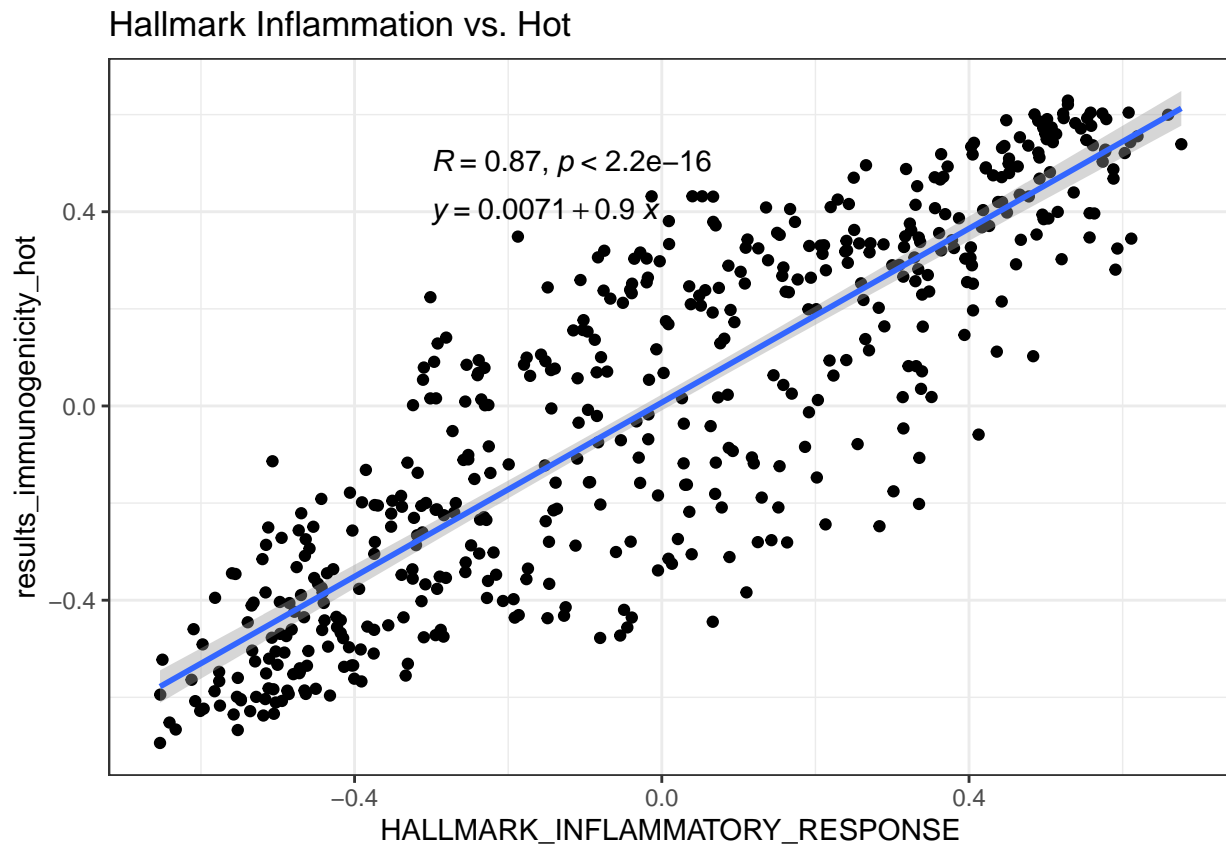
```
qplot(data = GSVA_GeneExpression_Matrix_t_df,
      HALLMARK_INFLAMMATORY_RESPONSE,
      results_immunogenicity_hot,
      xlab = "HALLMARK_INFLAMMATORY_RESPONSE",
      ylab = "results_immunogenicity_hot",
     main = "Hallmark Inflammation vs. Hot")+
  geom_smooth(method=lm)+
  stat_cor(label.x = (-.3), label.y = 0.5) +
  stat_regline_equation(label.x = (-.3), label.y = 0.4)+
  theme_bw()
```

```
## Warning: 'qplot()' was deprecated in ggplot2 3.4.0.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```
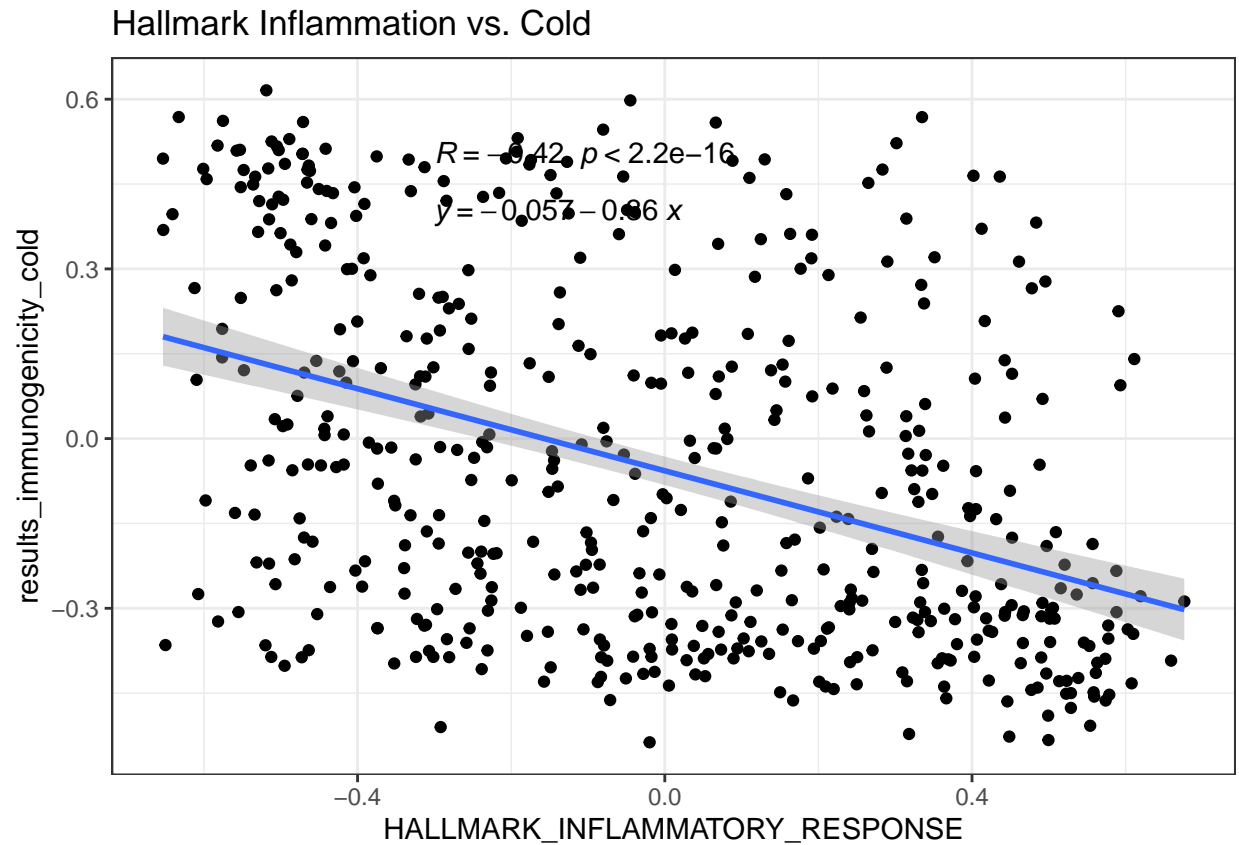
```
## 'geom_smooth()' using formula = 'y ~ x'
```

Hallmark Inflammation vs. Hot

$R = 0.87, p < 2.2e-16$

$y = 0.0071 + 0.9\,x$

HALLMARK_INFLAMMATORY_RESPONSE

2.2) Inflammatory Response vs Cold

```
qplot(data = GSVA_GeneExpression_Matrix_t_df,
      HALLMARK_INFLAMMATORY_RESPONSE,
      results_immunogenicity_cold,
      xlab = "HALLMARK_INFLAMMATORY_RESPONSE",
      ylab = "results_immunogenicity_cold",
      main = "Hallmark Inflammation vs. Cold")+
  geom_smooth(method=lm)+
  stat_cor(label.x = (-.3), label.y = 0.5) +
  stat_regline_equation(label.x = (-.3), label.y = 0.4)+
  theme_bw()
```
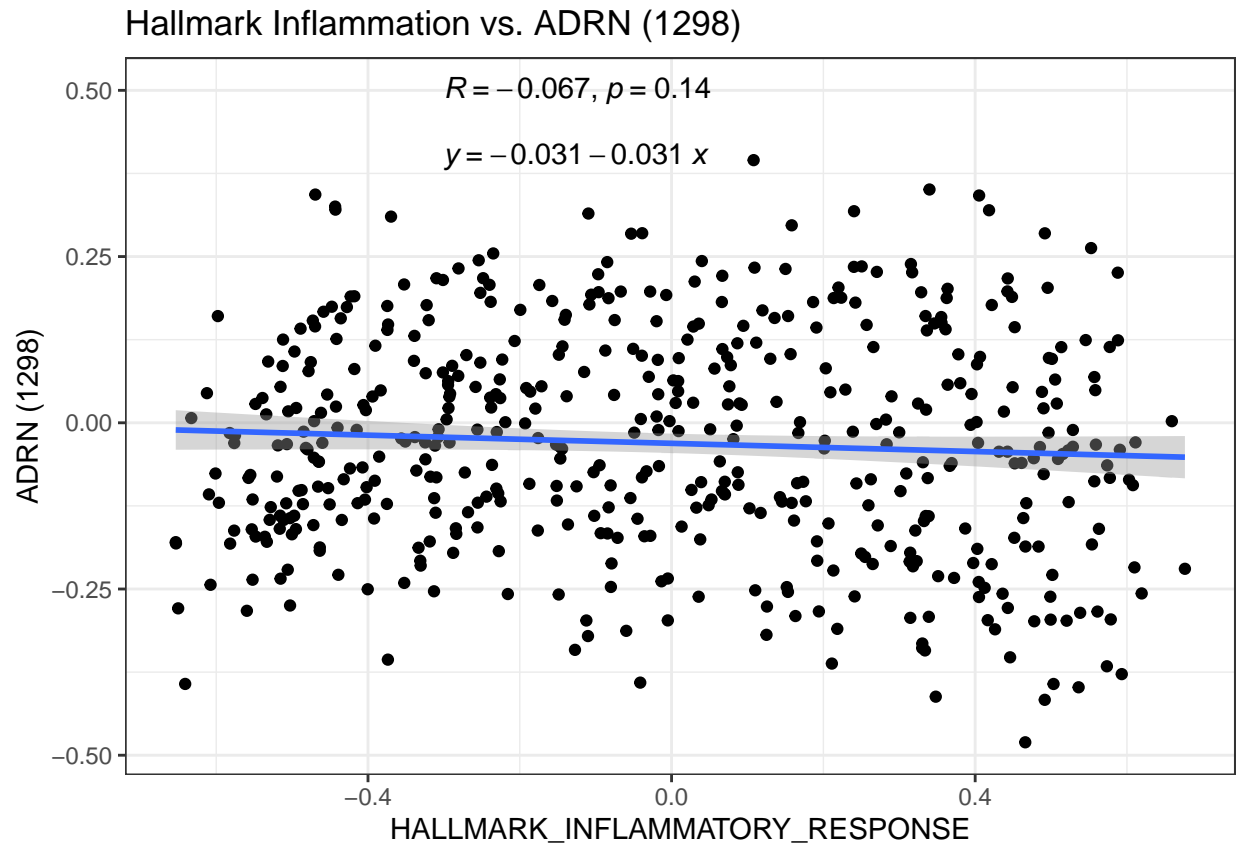
```
## 'geom_smooth()' using formula = 'y ~ x'
```

Hallmark Inflammation vs. Cold

$R = -0.42, p < 2.2e{-}16$

$y = -0.057 - 0.36\,x$

results_immunogenicity_cold

HALLMARK_INFLAMMATORY_RESPONSE

2.3) Inflammatory Response vs ADRN

```
qplot(data = GSVA_GeneExpression_Matrix_t_df,
        HALLMARK_INFLAMMATORY_RESPONSE,
        List_RNA_Normoxia_ADRN_vs_MES_Up.txt,
        xlab = "HALLMARK_INFLAMMATORY_RESPONSE",
        ylab = "ADRN (1298)",
      main = "Hallmark Inflammation vs. ADRN (1298)")+
  geom_smooth(method=lm)+
  stat_cor(label.x = (-.3), label.y = 0.5) +
  stat_regline_equation(label.x = (-.3), label.y = 0.4)+
  theme_bw()
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

Hallmark Inflammation vs. ADRN (1298)

$R = -0.067, p = 0.14$

$y = -0.031 - 0.031\,x$

ADRN (1298) — vertical axis

HALLMARK_INFLAMMATORY_RESPONSE — horizontal axis

2.4) Inflammatory Response vs MES

```
qplot(data = GSVA_GeneExpression_Matrix_t_df,
        HALLMARK_INFLAMMATORY_RESPONSE,
        List_RNA_Normoxia_ADRN_vs_MES_Down.txt,
        xlab = "HALLMARK_INFLAMMATORY_RESPONSE",
        ylab = "MES (1469)",
      main = "Hallmark Inflammation vs. MES (1469)")+
  geom_smooth(method=lm)+
  stat_cor(label.x = (-.3), label.y = 0.5) +
  stat_regline_equation(label.x = (-.3), label.y = 0.4)+
  theme_bw()
```
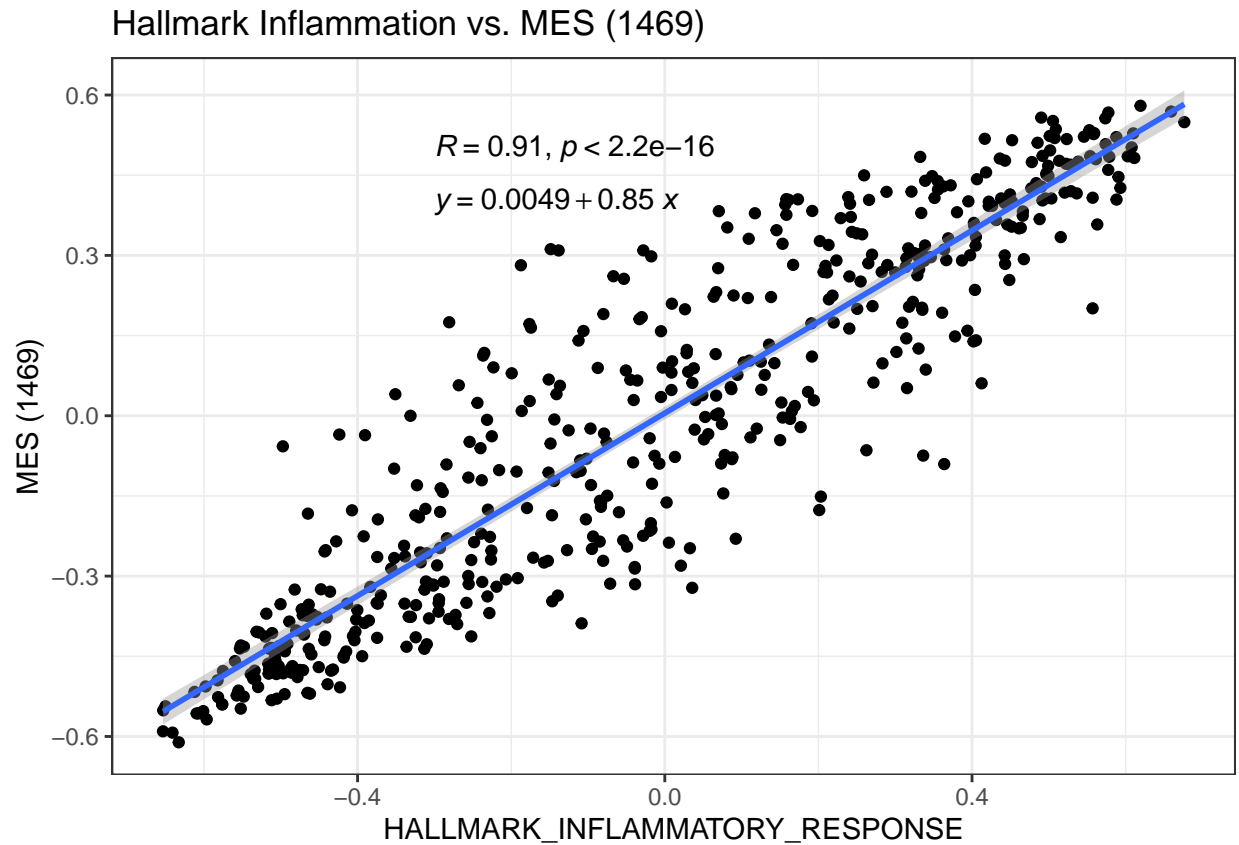
```
## `geom_smooth()` using formula = 'y ~ x'
```

Hallmark Inflammation vs. MES (1469)

$R = 0.91, p < 2.2e{-}16$

$y = 0.0049 + 0.85\,x$

MES (1469)

HALLMARK_INFLAMMATORY_RESPONSE

2.5) Inflammatory Response vs ADRN_Norm_vs_Hypo_Up_554.txt

```
qplot(data = GSVA_GeneExpression_Matrix_t_df,
      HALLMARK_INFLAMMATORY_RESPONSE,
      ADRN_Norm_vs_Hypo_Up_554.txt,
      xlab = "HALLMARK_INFLAMMATORY_RESPONSE",
      ylab = "Hypoxia Up (554)",
    main = "Hallmark Inflammation vs. ADRN Hypoxia Up (554)")+
  geom_smooth(method=lm)+
  stat_cor(label.x = (-.3), label.y = 0.5) +
  stat_regline_equation(label.x = (-.3), label.y = 0.4)+
  theme_bw()
```
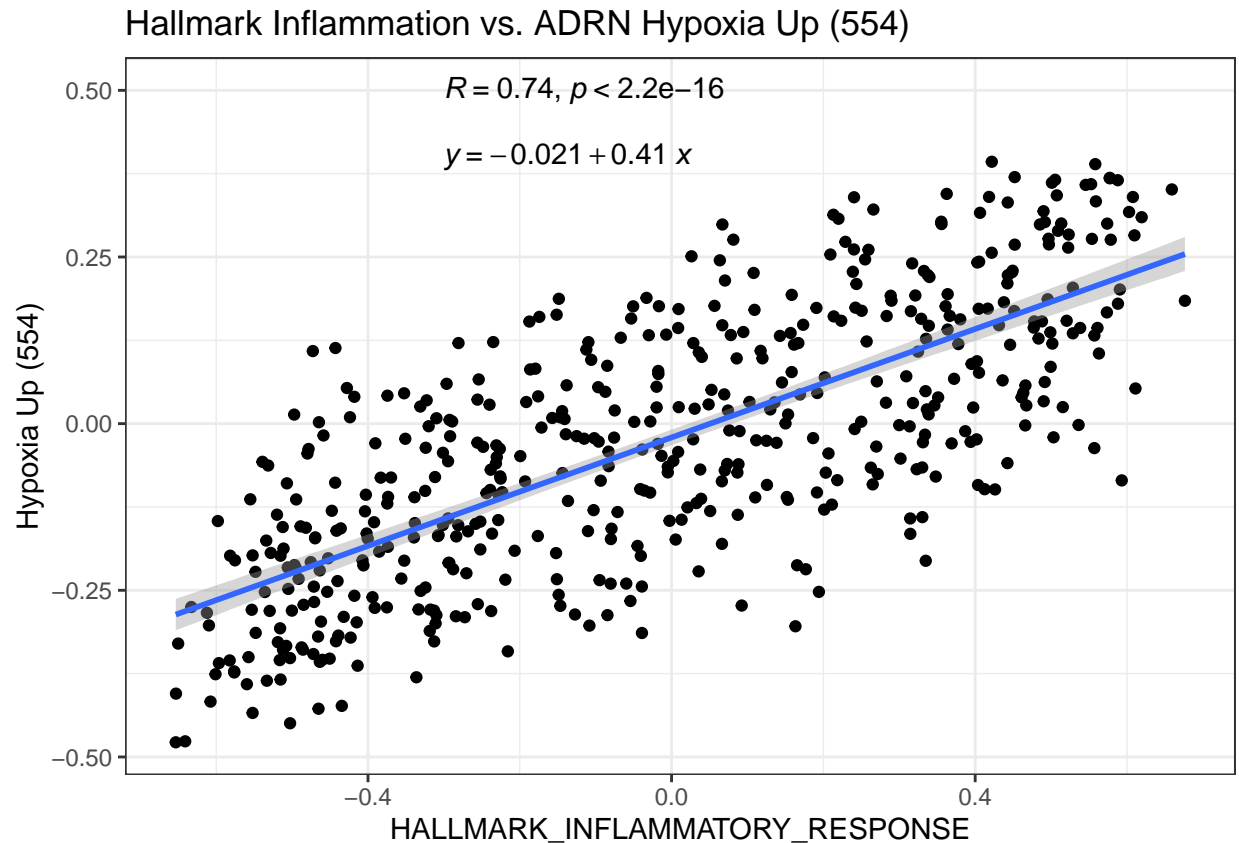
```
## 'geom_smooth()' using formula = 'y ~ x'
```

Hallmark Inflammation vs. ADRN Hypoxia Up (554)

$R = 0.74, p < 2.2e{-}16$

$y = -0.021 + 0.41\,x$

X-axis: HALLMARK_INFLAMMATORY_RESPONSE
Y-axis: Hypoxia Up (554)

2.6) Inflammatory Response vs ADRN_Norm_vs_Hypo_Down_635.txt

```
qplot(data = GSVA_GeneExpression_Matrix_t_df,
        HALLMARK_INFLAMMATORY_RESPONSE,
        ADRN_Norm_vs_Hypo_Down_635.txt,
        xlab = "HALLMARK_INFLAMMATORY_RESPONSE",
        ylab = "Hypoxia Down (635)",
      main = "Hallmark Inflammation vs. ADRN Hypoxia Down (635)")+
  geom_smooth(method=lm)+
  stat_cor(label.x = (-.3), label.y = 0.5) +
  stat_regline_equation(label.x = (-.3), label.y = 0.4)+
  theme_bw()
```
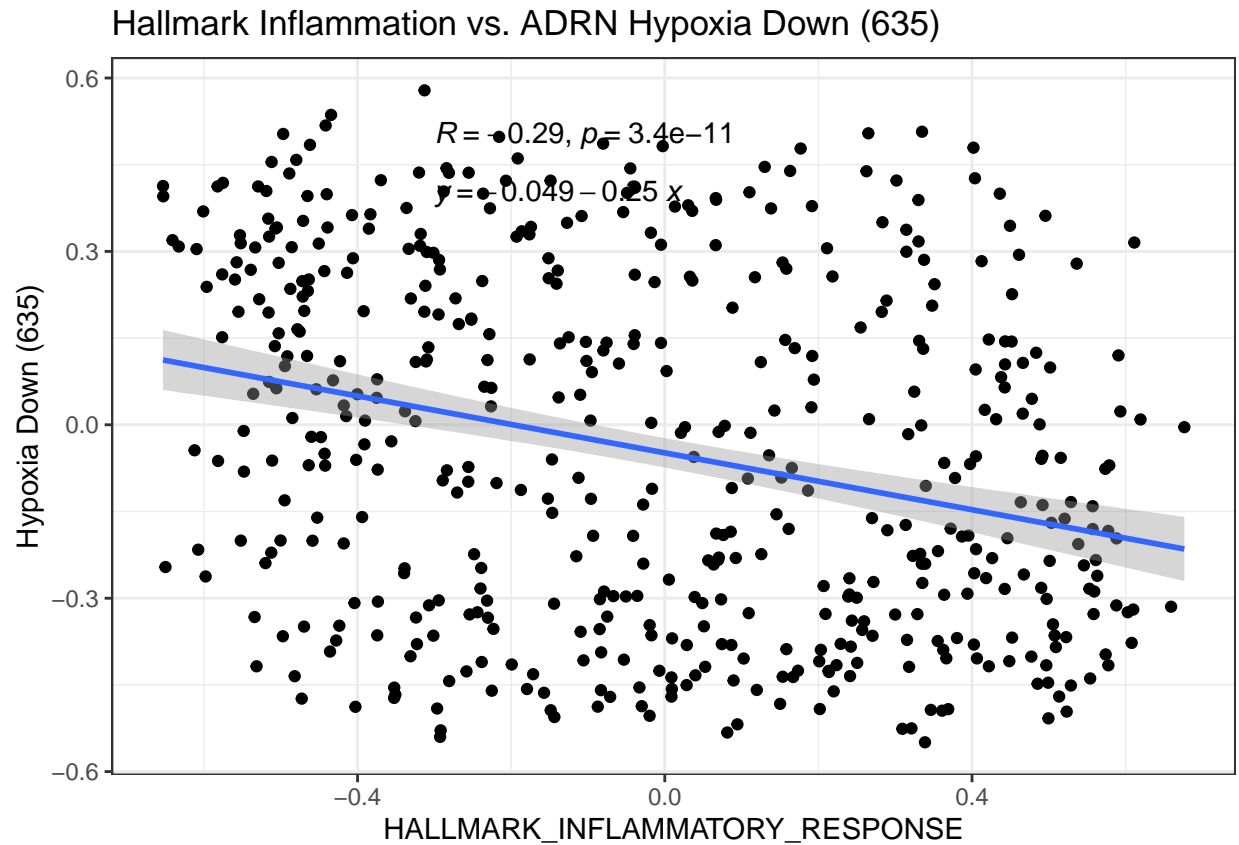
```
## 'geom_smooth()' using formula = 'y ~ x'
```

Hallmark Inflammation vs. ADRN Hypoxia Down (635)

$R = -0.29$, $p = 3.4e{-}11$

$y = -0.049 - 0.25 \, x$

2.7) Check MES vs. Hypoxia Up

```
qplot(data = GSVA_GeneExpression_Matrix_t_df,
       List_RNA_Normoxia_ADRN_vs_MES_Down.txt,
       ADRN_Norm_vs_Hypo_Up_554.txt,
       xlab = "MES (1469)",
       ylab = "Hypoxia Up (554)",
     main = "MES (1469) vs. ADRN Hypoxia Up (554)")+
  geom_smooth(method=lm)+
  stat_cor(label.x = (-.3), label.y = 0.5) +
  stat_regline_equation(label.x = (-.3), label.y = 0.4)+
  theme_bw()
```
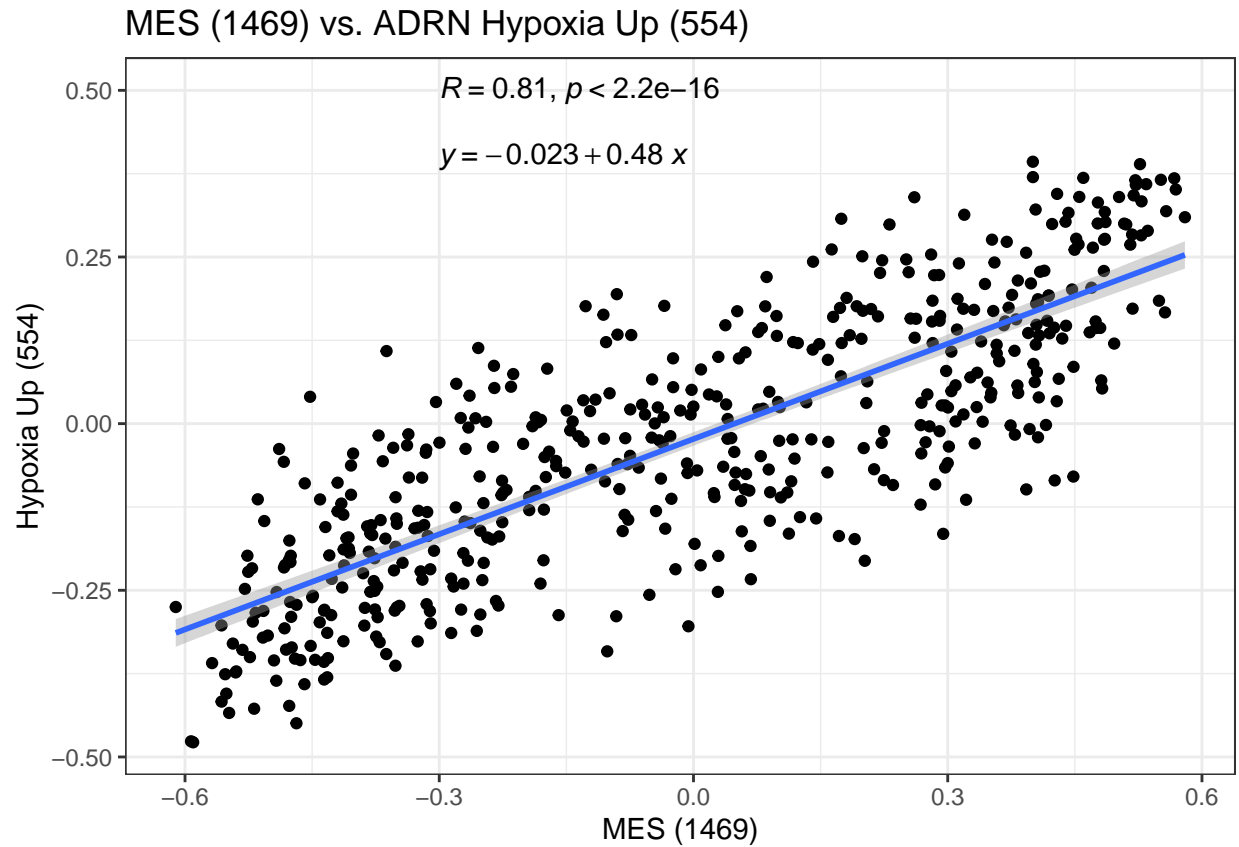
```
## 'geom_smooth()' using formula = 'y ~ x'
```

MES (1469) vs. ADRN Hypoxia Up (554)

$R = 0.81, p < 2.2e{-}16$

$y = -0.023 + 0.48\,x$

2.8) Check ADRN vs. Hypoxia Down

```
qplot(data = GSVA_GeneExpression_Matrix_t_df,
        List_RNA_Normoxia_ADRN_vs_MES_Down.txt,
        ADRN_Norm_vs_Hypo_Down_635.txt,
        xlab = "ADRN (1298)",
        ylab = "Hypoxia Down (635)",
       main = "ADRN (1298) vs. ADRN Hypoxia Down (635)")+
  geom_smooth(method=lm)+
  stat_cor(label.x = (-.3), label.y = 0.5) +
  stat_regline_equation(label.x = (-.3), label.y = 0.4)+
  theme_bw()
```
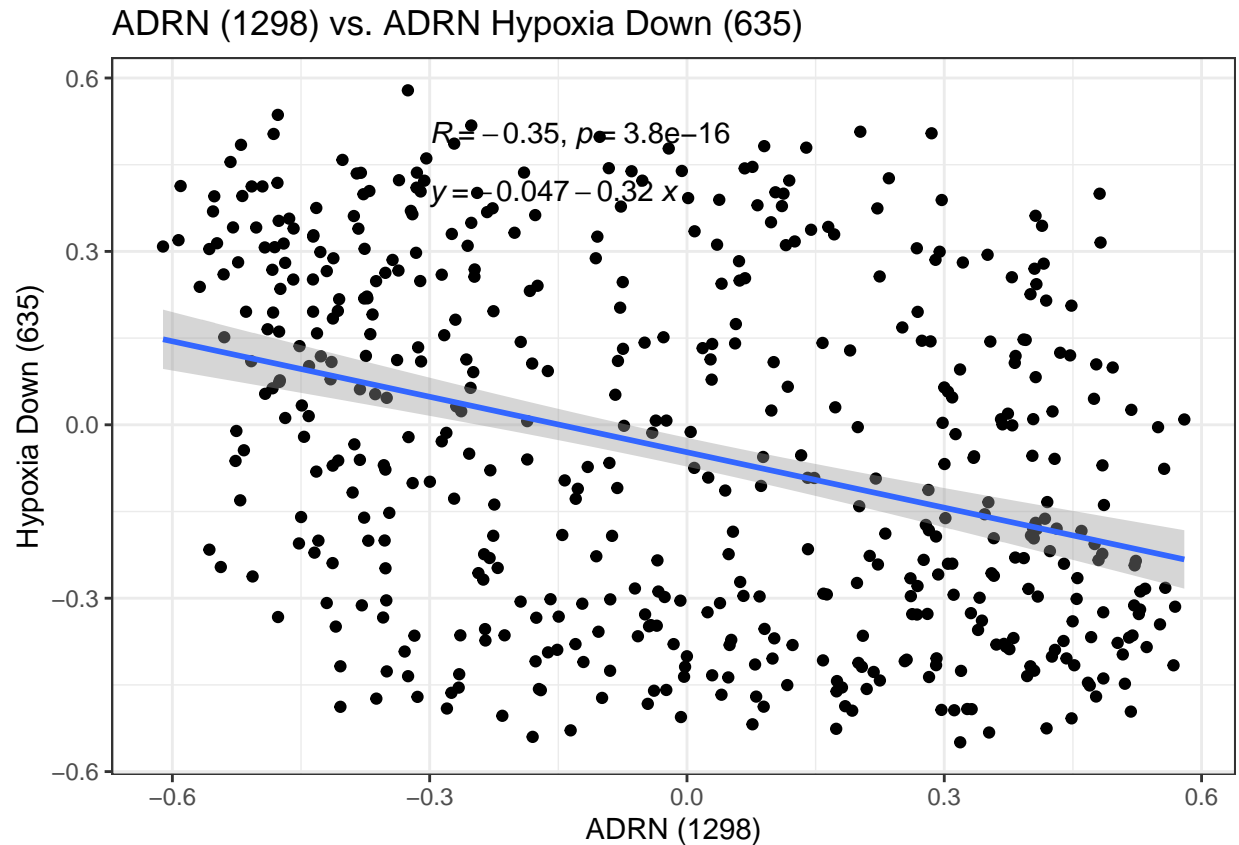
```
## `geom_smooth()` using formula = 'y ~ x'
```

ADRN (1298) vs. ADRN Hypoxia Down (635)

$R = -0.35$, $p = 3.8e{-}16$

$y = -0.047 - 0.32\,x$

(y-axis) Hypoxia Down (635)

(x-axis) ADRN (1298)

2.9) Check Hot vs. Hypoxia Up

```
qplot(data = GSVA_GeneExpression_Matrix_t_df,
        results_immunogenicity_hot,
        ADRN_Norm_vs_Hypo_Up_554.txt,
        xlab = "results_immunogenicity_hot",
        ylab = "Hypoxia Up (554)",
      main = "Immunologicaly Hot vs. ADRN Hypoxia Up (554)")+
  geom_smooth(method=lm)+
  stat_cor(label.x = (-.3), label.y = 0.5) +
  stat_regline_equation(label.x = (-.3), label.y = 0.4)+
  theme_bw()
```
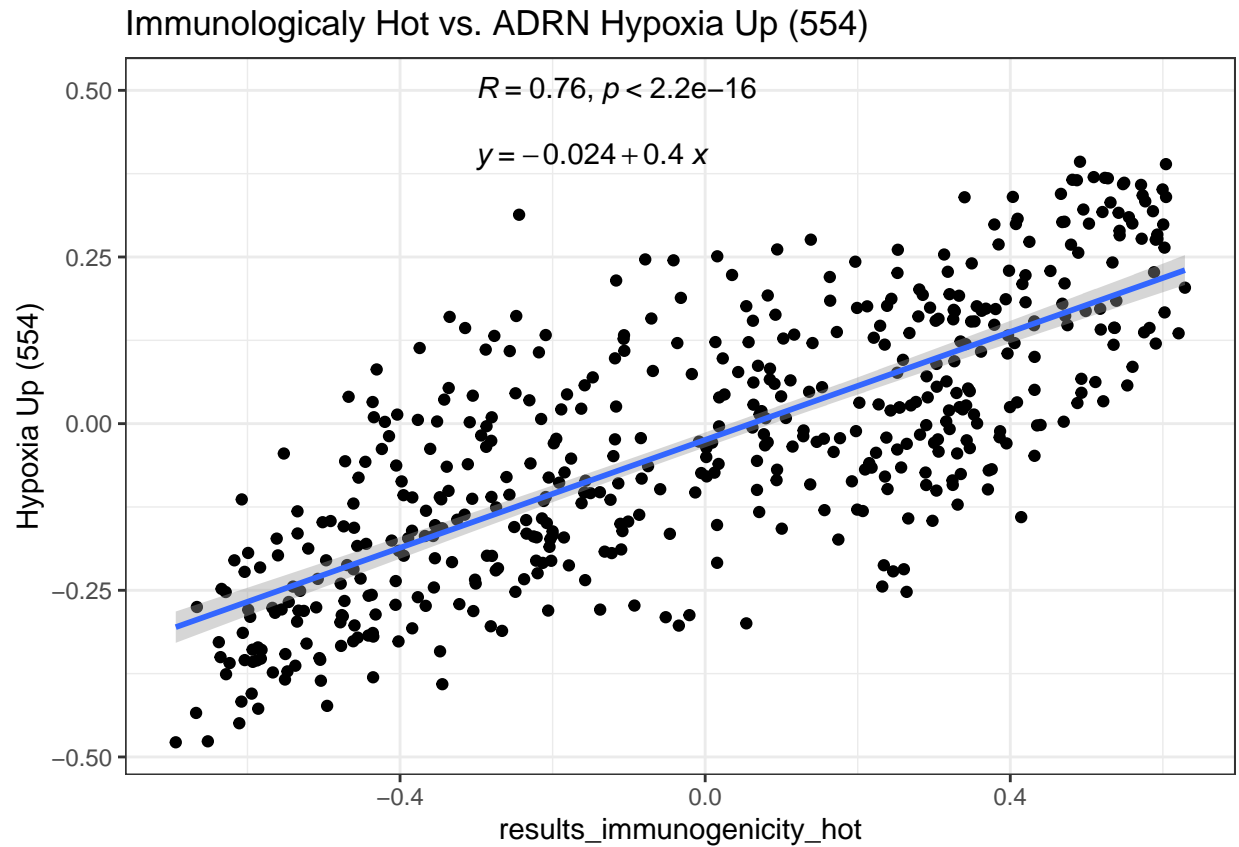
```
## 'geom_smooth()' using formula = 'y ~ x'
```

Immunologicaly Hot vs. ADRN Hypoxia Up (554)

$R = 0.76, p < 2.2e{-}16$

$y = -0.024 + 0.4\,x$

2.10) Check Cold vs. Hypoxia Down

```
qplot(data = GSVA_GeneExpression_Matrix_t_df,
        results_immunogenicity_cold,
        ADRN_Norm_vs_Hypo_Down_635.txt,
        xlab = "results_immunogenicity_cold",
        ylab = "Hypoxia Down (635)",
      main = "Immunologicaly Cold vs. ADRN Hypoxia Down (635)")+
  geom_smooth(method=lm)+
  stat_cor(label.x = (-.3), label.y = 0.5) +
  stat_regline_equation(label.x = (-.3), label.y = 0.4)+
  theme_bw()
```
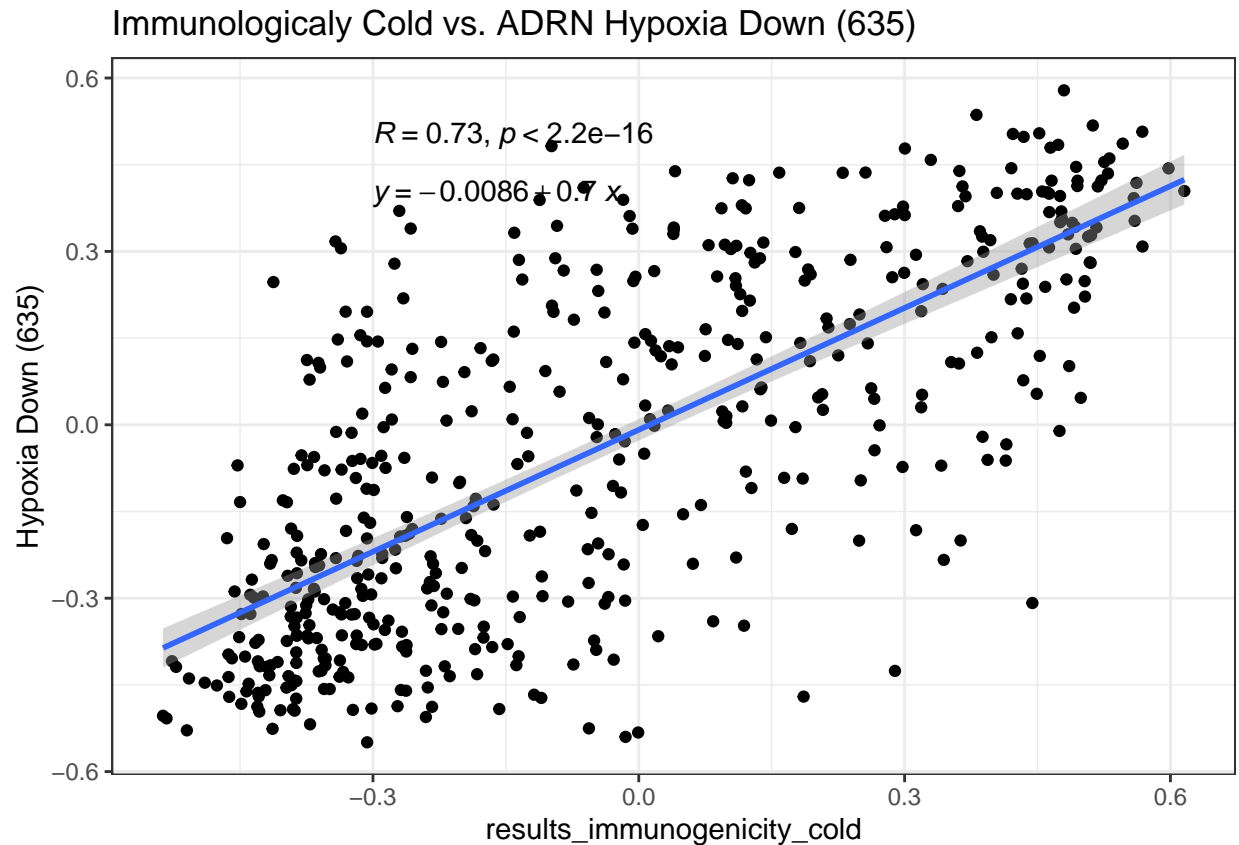
```
## 'geom_smooth()' using formula = 'y ~ x'
```

Immunologicaly Cold vs. ADRN Hypoxia Down (635)

$R = 0.73$, $p < 2.2e{-}16$

$y = -0.0086 + 0.7\,x$

Y-axis: Hypoxia Down (635)

X-axis: results_immunogenicity_cold

Correlation Matrix

```r
r2_gse62564_gsva_genes_df <- t(r2_gse62564_GSVA_genes)
```

Select terms to be plotted

```r
GSVA_GeneExpression_Matrix <- GSVA_GeneExpression_Matrix_t_df[,
                                        c("HALLMARK_INFLAMMATORY_RESPONSE",
                                          "results_immunogenicity_hot",
                                          "results_immunogenicity_cold",
                                          "List_RNA_Normoxia_ADRN_vs_MES_Up.txt",
                                          "List_RNA_Normoxia_ADRN_vs_MES_Down.txt",
                                          "ADRN_Norm_vs_Hypo_Up_554.txt",
                                          "ADRN_Norm_vs_Hypo_Down_635.txt",
                                          "MES_Norm_vs_Hypo.compCas_Up_287.txt",
                                          "MES_Norm_vs_Hypo.compCas_Down_171.txt",
                                          "MES_Norm_vs_Hypo.compCas_Up_212.txt",
                                          "MES_Norm_vs_Hypo.compCas_Down_103.txt"
                                        )]
```

```r
GSVA_GeneExpression_Matrix_df <- as.data.frame(GSVA_GeneExpression_Matrix)

GSVA_GeneExpression_Matrix_df <- GSVA_GeneExpression_Matrix_df %>% dplyr::rename(Inflam_Response = HALLI
GSVA_GeneExpression_Matrix_df <- GSVA_GeneExpression_Matrix_df %>% dplyr::rename(Hot = results_immunoger
GSVA_GeneExpression_Matrix_df <- GSVA_GeneExpression_Matrix_df %>% dplyr::rename(Cold = results_immunoge
GSVA_GeneExpression_Matrix_df <- GSVA_GeneExpression_Matrix_df %>% dplyr::rename(ADRN_1298 = List_RNA_No
GSVA_GeneExpression_Matrix_df <- GSVA_GeneExpression_Matrix_df %>% dplyr::rename(MES_1469 = List_RNA_Nor
GSVA_GeneExpression_Matrix_df <- GSVA_GeneExpression_Matrix_df %>% dplyr::rename(Hypo_Up_554 = ADRN_Norm
GSVA_GeneExpression_Matrix_df <- GSVA_GeneExpression_Matrix_df %>% dplyr::rename(Hypo_Down_635 = ADRN_No

GSVA_GeneExpression_Matrix_df <- GSVA_GeneExpression_Matrix_df %>% dplyr::rename(Hypo_UP_287 = MES_Norm_
GSVA_GeneExpression_Matrix_df <- GSVA_GeneExpression_Matrix_df %>% dplyr::rename(Hypo_DN_171 = MES_Norm_
GSVA_GeneExpression_Matrix_df <- GSVA_GeneExpression_Matrix_df %>% dplyr::rename(Hypo_UP_121 = MES_Norm_
GSVA_GeneExpression_Matrix_df <- GSVA_GeneExpression_Matrix_df %>% dplyr::rename(Hypo_DN_103 = MES_Norm_
```

Make character, numeric

```r
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.1.2
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:expss':
##
##     compute, contains, na_if, recode, vars, where
```

```
## The following objects are masked from 'package:maditr':
##
##     between, coalesce, first, last
```

```
## The following objects are masked from 'package:data.table':
##
##     between, first, last
```

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

```
## The following object is masked from 'package:graph':
##
##     union
```

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

```
## The following objects are masked from 'package:IRanges':
##
##     collapse, desc, intersect, setdiff, slice, union
```

```
## The following objects are masked from 'package:S4Vectors':
##
##     first, intersect, rename, setdiff, setequal, union


## The following object is masked from 'package:Biobase':
##
##     combine


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


## The following objects are masked from 'package:stats':
##
##     filter, lag


## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
GSVA_GeneExpression_Matrix_df <- GSVA_GeneExpression_Matrix_df %>% mutate_if(is.character, as.numeric)
```

Calculate correlation

```r
GSVA_GeneExpression_Matrix_cor_df <- cor(GSVA_GeneExpression_Matrix_df)
```
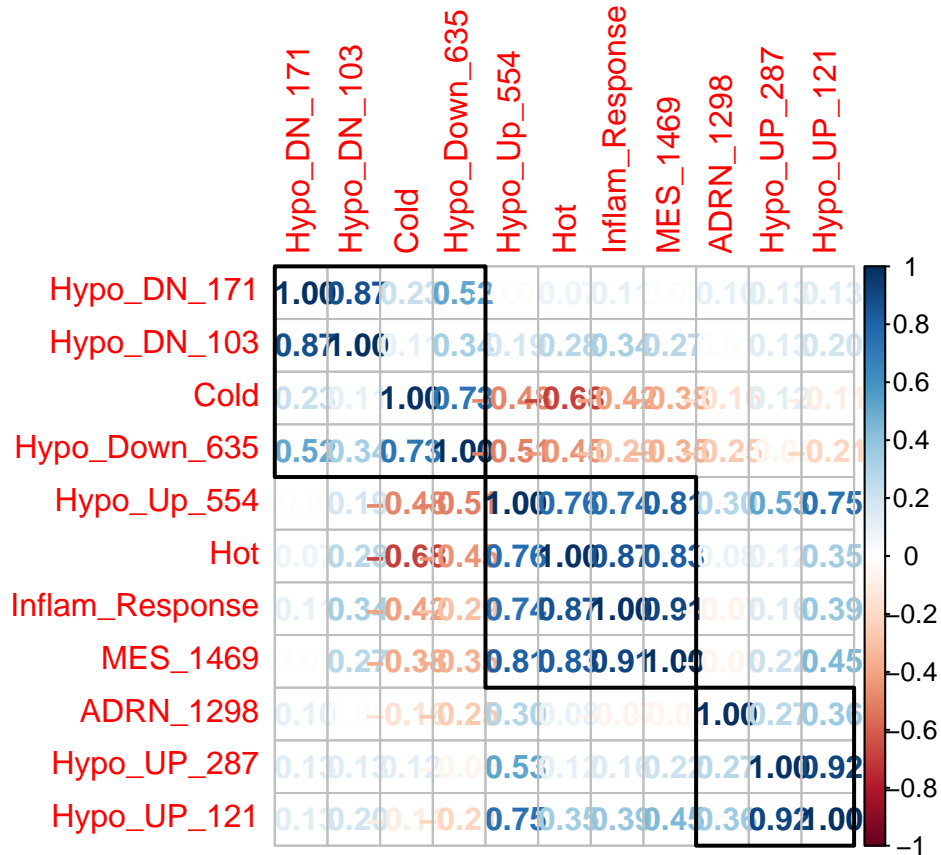
Plot correlation matrix

Plot with colorful number

```r
library(corrplot )
```

```
## corrplot 0.92 loaded
```

```r
corrplot(GSVA_GeneExpression_Matrix_cor_df, method = 'number', order="hclust", addrect = 3)
```

Plot circles

```r
corrplot(GSVA_GeneExpression_Matrix_cor_df,
         type="lower",
         order="hclust",
         #method = c(pie),
         title = "Correlation Matrix Hot and Cold Tumors",
         tl.col = "black",
         #tl.offset=1,
         tl.srt = 45,
         addrect = 4,
         col=colorRampPalette(c("#053061",
                                "#2166AC",
                                "#4393C3",
                                "#92C5DE",
                                "#D1E5F0",
                                "#FFFFFF",
                                "#FDDBC7",
                                "#F4A582",
                                "#D6604D",
                                "#B2182B",
                                "#67001F"
                                ))(200)
)
```

Correlation Matrix Hot and Cold Tumors