Procesado de los datos del análisis de deconvolución:

Elena Eyre Sánchez, PhD

2024-10-26

Contents

1 Introducción y Objetivo

En este script proceso los datos obtenidos del análisis de deconvolución para después poder iniciar los análisis estadísticos.

2 Paquetes y datos

Inicialmente cargo los paquetes y los datos necesarios para este trabajo de post-análisis.

```r
knitr::opts_chunk$set(warning=FALSE)
package_to_load <- c("readr", "dplyr", "ggplot2", "tidyr", "dplyr",
                     "RColorBrewer", "gplots", "readxl")
for (package in package_to_load) {
  require(package, character.only = T); packageVersion(package)
}
extra_to_load <- c("knitr","stringr", "stringi", "ggrepel","ggpubr", "ggbreak",
                   "reshape2","ggfortify","cowplot","data.table")
for (package in extra_to_load) {
  require(package, character.only = T); packageVersion(package)
}
rm(package_to_load, extra_to_load)
```

Los datos a comparar son los resultados obtenidos de los análisis previos de deconvolución:

- GSE22155 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi) con dos plataformas: GPL6102 y GPL6947.

- GSE35640 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE35640)

- GSE50509 (<>)

- GSE61992 (<>)

- GSE54467 (<>)

- TCGA-SKCM (https://portal.gdc.cancer.gov/projects/TCGA-SKCM)

De manera excepcional, tenemos GSE120575 (<>), que consiste en un dataset scRNA-seq en el que han analizado muestras tumorales y controles.

```
setwd("~/Desktop/ELENA_UOC/TFM")
bisque_GSE22155_GPL6102 <- read_csv("Deconvolution_analysis/bisque_GSE22155_GPL6102.csv")
bisque_GSE22155_GPL6947 <- read_csv("Deconvolution_analysis/bisque_GSE22155_GPL6947.csv")
bisque_GSE35640 <- read_excel("Deconvolution_analysis/bisque_GSE35640.xlsx")
bisque_GSE50509 <- read_excel("Deconvolution_analysis/bisque_GSE50509.xlsx")
bisque_GSE61992 <- read_excel("Deconvolution_analysis/bisque_GSE61992.xlsx")
bisque_GSE54467 <- read_delim("Deconvolution_analysis/bisque_GSE54467.csv",
    delim = ";", escape_double = FALSE, trim_ws = TRUE)
#bisque_GSE65904 <- read_csv("Deconvolution_analysis/bisque_GSE65904.csv")
bisque_tcga_skcm <- read_csv("Deconvolution_analysis/bisque_tcga_skcm.csv")
TCGA_SKCM_clinical_data <- read_excel("Datasets/TCGA_SKCM_clinical_data.xlsx")
Proportions_per_patient_of_cell_type_GSE120575 <- read_excel("Proportions_per_patient_of_cell_type_GSE1
Final_metadata_GSE120575 <- read_excel("Final_metadata_GSE120575.xlsx")
```

2.1 Pre-procesar datos

Para poder realizar las comparaciones de los tratamientos, necesito harmonizar los datos: que .

```
df_GSE22155_GPL6102 <- bisque_GSE22155_GPL6102[,1:7]
df_GSE22155_GPL6102$samples <- bisque_GSE22155_GPL6102$geo_accession
#df_GSE22155_GPL6102$treatment <- "Standard treatment with dacarbazine (DTIC) monotherapy, 850 to 1,000
df_GSE22155_GPL6102$treatment <- "Standard (dacarbazine monotherapy, DTIC)"
df_GSE22155_GPL6102$sample_type[grep("metastasis", bisque_GSE22155_GPL6102$source_name_ch1)] <- "Metast
df_GSE22155_GPL6102$contact_country <- bisque_GSE22155_GPL6102$contact_country
df_GSE22155_GPL6102$age_diag <- bisque_GSE22155_GPL6102$`age at primary diagnosis:ch1`
df_GSE22155_GPL6102$age_metas <-bisque_GSE22155_GPL6102$`age at metastases:ch1`
df_GSE22155_GPL6102$sex <- bisque_GSE22155_GPL6102$`sex:ch1`
df_GSE22155_GPL6102$loc_melan <-bisque_GSE22155_GPL6102$`localization of primary melanoma:ch1`
df_GSE22155_GPL6102$os_event <- bisque_GSE22155_GPL6102$`event (0=alive, 1=dead):ch1`
df_GSE22155_GPL6102$os_days <- bisque_GSE22155_GPL6102$`os (days):ch1`
df_GSE22155_GPL6102$stage <- bisque_GSE22155_GPL6102$`stage:ch1`
df_GSE22155_GPL6102$response <- NA
df_GSE22155_GPL6102$study <- "GSE22155_GPL6102"

df_GSE22155_GPL6947 <- bisque_GSE22155_GPL6947[,1:7]
df_GSE22155_GPL6947$samples <- bisque_GSE22155_GPL6947$geo_accession
#df_GSE22155_GPL6947$treatment <- "Standard treatment with dacarbazine (DTIC) monotherapy, 850 to 1,000
df_GSE22155_GPL6947$treatment <- "Standard (dacarbazine monotherapy, DTIC)"
df_GSE22155_GPL6947$sample_type[grep("metastasis", bisque_GSE22155_GPL6947$source_name_ch1)] <- "Metast
df_GSE22155_GPL6947$contact_country <- bisque_GSE22155_GPL6947$contact_country
df_GSE22155_GPL6947$age_diag <- bisque_GSE22155_GPL6947$`age at primary diagnosis:ch1`
df_GSE22155_GPL6947$age_metas <-bisque_GSE22155_GPL6947$`age at metastases:ch1`
df_GSE22155_GPL6947$sex <-bisque_GSE22155_GPL6947$`sex:ch1`
df_GSE22155_GPL6947$loc_melan <-bisque_GSE22155_GPL6947$`localization of primary melanoma:ch1`
df_GSE22155_GPL6947$os_event <- bisque_GSE22155_GPL6947$`event (0=alive, 1=dead):ch1`
df_GSE22155_GPL6947$os_days <- bisque_GSE22155_GPL6947$`os (days):ch1`
df_GSE22155_GPL6947$stage <- bisque_GSE22155_GPL6947$`stage:ch1`
df_GSE22155_GPL6947$response <- NA
df_GSE22155_GPL6947$study <- "GSE22155_GPL6947"

df_GSE35640 <- bisque_GSE35640[,1:7]
df_GSE35640$samples <- bisque_GSE35640$geo_accession
```

```r
df_GSE35640$treatment[grep("prior",bisque_GSE35640$source_name_ch1)] <- "untreated"
df_GSE35640$sample_type[grep("Melanoma", bisque_GSE35640$source_name_ch1)] <- "Primary Tumor"
df_GSE35640$contact_country <- bisque_GSE35640$contact_country
df_GSE35640$age_diag <- NA
df_GSE35640$age_metas <- NA
df_GSE35640$sex <- NA
df_GSE35640$loc_melan <-NA
df_GSE35640$os_event <- NA
df_GSE35640$os_days <- NA
df_GSE35640$stage <- NA
df_GSE35640$response <- bisque_GSE35640$`response:ch1`
df_GSE35640$study <- "GSE35640"

df_GSE50509 <- bisque_GSE50509[,43:49]
df_GSE50509$samples <- bisque_GSE50509$geo_accession
df_GSE50509$treatment <- bisque_GSE50509$`treatment:ch1`
df_GSE50509$treatment[df_GSE50509$treatment == "none"] <- "untreated"
df_GSE50509$sample_type<- "Primary Tumor"
df_GSE50509$contact_country <- bisque_GSE50509$contact_country
df_GSE50509$age_diag <- NA
df_GSE50509$age_metas <- NA
df_GSE50509$sex <- NA
df_GSE50509$loc_melan <-bisque_GSE50509$`location:ch1`
df_GSE50509$os_event <- NA
df_GSE50509$os_days <- NA
df_GSE50509$stage <- NA
df_GSE50509$response <- NA
df_GSE50509$study <- "GSE50509"

df_GSE54467 <- bisque_GSE54467[,1:7]
df_GSE54467$samples <- bisque_GSE54467$geo_accession
df_GSE54467$treatment <- "standard?"
df_GSE54467$sample_type[grep("melanoma", bisque_GSE54467$title)] <- "Primary Tumor"
df_GSE54467$contact_country <- bisque_GSE54467$contact_country
df_GSE54467$age_diag <- bisque_GSE54467$`patient age at primary diagnosis (years):ch1`
df_GSE54467$age_metas <- bisque_GSE54467$`patient age at stage iii sample banked (years):ch1`
df_GSE54467$sex <- bisque_GSE54467$`patient sex:ch1`
df_GSE54467$loc_melan <-NA
df_GSE54467$os_event <- 1
df_GSE54467$os_event[grep("Alive",bisque_GSE54467$`patient last status:ch1`)] <- 0
df_GSE54467$os_days <- (bisque_GSE54467$`survival from primary melanoma (months):ch1`)*30
df_GSE54467$stage <- bisque_GSE54467$`stage at primary diagnosis 5th edition:ch1`
df_GSE54467$response <- NA
df_GSE54467$response[grep("NSR",bisque_GSE54467$`patient last status:ch1`)] <- "not sustained response"
df_GSE54467$response[grep("Not Melanoma",bisque_GSE54467$`patient last status:ch1`)] <- "responder"
df_GSE54467$response[grep("with Melanoma",bisque_GSE54467$`patient last status:ch1`)] <- "non-responder`
df_GSE54467$study <- "GSE54467"

df_GSE61992 <- bisque_GSE61992[,36:42]
df_GSE61992$samples <- bisque_GSE61992$geo_accession
df_GSE61992$treatment <- "untreated"
df_GSE61992$treatment[grep("trametinib", bisque_GSE61992$description)] <- "dabrafenib + trametinib"
#df_GSE61992$treatment[grep("EDT", bisque_GSE61992$description)] <- "EDT(early during treatment) BRAFi
```

```r
df_GSE61992$treatment[grep("EDT", bisque_GSE61992$description)] <- "dabrafenib"
df_GSE61992$sample_type[grep("Melanoma", bisque_GSE61992$source_name_ch1)] <- "melanoma"
df_GSE61992$contact_country <- bisque_GSE61992$contact_country
df_GSE61992$age_diag <- NA
df_GSE61992$age_metas <- NA
df_GSE61992$sex <- NA
df_GSE61992$loc_melan <- ifelse(bisque_GSE61992$`tissue:ch1` == "Melanoma- subcutaneous", "subcutaneous"
                                ifelse(bisque_GSE61992$`tissue:ch1` == "Melanoma- lymph node", "lymph no
                                       ifelse(bisque_GSE61992$`tissue:ch1` == "Melanoma- bowel", "bowel"
                                              ifelse(bisque_GSE61992$`tissue:ch1` == "Melanoma- peritone
df_GSE61992$os_event <- NA
df_GSE61992$os_days <- NA
df_GSE61992$stage <- NA
df_GSE61992$response <- NA
df_GSE61992$study <- "GSE61992"

TCGA_SKCM_clinical_data$submitter_id <- TCGA_SKCM_clinical_data$case_submitter_id
bisque_tcga_skcm <- as.data.frame(inner_join(bisque_tcga_skcm, TCGA_SKCM_clinical_data, by = 'submitter_
df_tcga_skcm <- bisque_tcga_skcm[,1:7]
df_tcga_skcm$samples <- bisque_tcga_skcm$submitter_id.samples
df_tcga_skcm$treatment <- bisque_tcga_skcm$treatment_type
df_tcga_skcm$treatment[grep("NOS",bisque_tcga_skcm$treatment_type)] <- "NOS, pharmaceutical and radiati
df_tcga_skcm$sample_type <- bisque_tcga_skcm$sample_type.samples
df_tcga_skcm$contact_country <- bisque_tcga_skcm$name.tissue_source_site
df_tcga_skcm$age_diag <- bisque_tcga_skcm$age_at_initial_pathologic_diagnosis
df_tcga_skcm$age_metas <- NA
df_tcga_skcm$sex <- ifelse(bisque_tcga_skcm$gender == "female", "Female", "Male")
df_tcga_skcm$loc_melan <- bisque_tcga_skcm$submitted_tumor_location
df_tcga_skcm$loc_melan[df_tcga_skcm$loc_melan == "Regional Lymph Node"] <- "lymph node"
df_tcga_skcm$loc_melan[df_tcga_skcm$loc_melan == "Primary Tumor"] <- "primary detected"
df_tcga_skcm$loc_melan[df_tcga_skcm$loc_melan == "Regional Cutaneous or Subcutaneous Tissue (includes sa
df_tcga_skcm$os_event <- ifelse(bisque_tcga_skcm$vital_status.demographic == "Dead", 1,0)
df_tcga_skcm$os_days <- bisque_tcga_skcm$days_to_death
df_tcga_skcm$stage <- bisque_tcga_skcm$tumor_stage.diagnoses
df_tcga_skcm$stage <- ifelse(bisque_tcga_skcm$tumor_stage.diagnoses == "stage i", "I",
                             ifelse(bisque_tcga_skcm$tumor_stage.diagnoses == "stage ia", "I",
                                    ifelse(bisque_tcga_skcm$tumor_stage.diagnoses == "stage ib", "I",
                                           ifelse(bisque_tcga_skcm$tumor_stage.diagnoses == "stage ii",
                                                  ifelse(bisque_tcga_skcm$tumor_stage.diagnoses == "stag
                                                         ifelse(bisque_tcga_skcm$tumor_stage.diagnoses
                                                                ifelse(bisque_tcga_skcm$tumor_stage.dia
                                                                       ifelse(bisque_tcga_skcm$tumor_sta
                                                                       ifelse(bisque_tcga_skcm$tumor_sta
                                                                              ifelse(bisque_tcga_skcm$tu
                                                                                     ifelse(bisque_tcga_
                                                                                            ifelse(bisqu
df_tcga_skcm$response <- NA
df_tcga_skcm$study <- "tcga_skcm"

rownames(Proportions_per_patient_of_cell_type_GSE120575) <- Proportions_per_patient_of_cell_type_GSE1205
proportions_GSE120575 <- as.data.frame(t(Proportions_per_patient_of_cell_type_GSE120575[,-1]))
colnames(proportions_GSE120575) <- rownames(Proportions_per_patient_of_cell_type_GSE120575)
proportions_GSE120575$`Patient ID`<- rownames(proportions_GSE120575)
```

```r
proportions_GSE120575 <- inner_join(proportions_GSE120575, Final_metadata_GSE120575, by = "Patient ID")
proportions_GSE120575$treatment <- proportions_GSE120575$Therapy
proportions_GSE120575$study <- "GSE120575"
```

Salvo para el caso de GSE120575, unifico los resultados del resto de estudios en lo que se refiere a las células resultantes de la deconvolución y el tratamiento.

```r
#df <-
#  df_GSE35640[, c("B_cell","CAF","Endo_cell","Macrophage","NK","Other_cells","T_cell","treatment")] %>
#  bind_rows(df_GSE22155_GPL6102[, c("B_cell","CAF","Endo_cell","Macrophage","NK","Other_cells","T_cell
#df <- df %>% bind_rows(df_GSE22155_GPL6947[, c("B_cell","CAF","Endo_cell","Macrophage","NK","Other_cel
#df <- df %>% bind_rows(df_GSE50509[, c("B_cell","CAF","Endo_cell","Macrophage","NK","Other_cells","T_c
#df <- df %>% bind_rows(df_GSE61992[, c("B_cell","CAF","Endo_cell","Macrophage","NK","Other_cells","T_c
#df <- df %>% bind_rows(df_tcga_skcm[, c("B_cell","CAF","Endo_cell","Macrophage","NK","Other_cells","T_
#df <- rbind(df ,df_GSE54467[, c("B_cell","CAF","Endo_cell","Macrophage","NK","Other_cells","T_cell","t

df <- rbind(as.data.frame(df_GSE35640), as.data.frame(df_GSE22155_GPL6102))
df <- rbind(df,as.data.frame(df_GSE22155_GPL6947))
df <- rbind(df,as.data.frame(df_GSE50509))
df <- rbind(df, as.data.frame(df_GSE61992))
df <- rbind(df,as.data.frame(df_tcga_skcm))
df <- rbind(df,as.data.frame(df_GSE54467))


df2 <- reshape2::melt(df, "treatment")
df2 <- reshape2::melt(df, id.vars = c("treatment","samples", "sample_type","contact_country", "age_diag
  variable.name = "cell_type",
  value.name = "deconv_value")
df2$deconv_value <- sapply(df2$deconv_value, as.numeric)

pdf(file = "treatment_comparison.pdf", width = 15, height = 25)
ggplot(df2, aes(treatment, deconv_value)) + geom_boxplot() + facet_grid(cell_type ~ sex) +
  #theme_classic() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1))
ggplot(df2, aes(treatment, deconv_value)) + geom_boxplot(aes(fill = sample_type)) + facet_grid(loc_mela
  #theme_classic() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1))
ggplot(df2, aes(treatment, deconv_value)) + geom_boxplot(aes(fill = sample_type)) + facet_grid(cell_type
  #theme_classic() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1))
dev.off()
```

```
## pdf
##   2
```

En este punto también necesito harmonizar más. Sería especialmente en la localización del melanoma y el tipo de muestra.

```r
#unique(df2$loc_melan)
#[1] NA                        "Trunk"                  "Lower extremity"          "Head"
# [6] "Upper extremity, acral"  "Upper extremity"        "Head and eye"             "Lower extrem
#[11] "Genital"                 "GI tractus"             "Eye"                      "Mucosa, head
#[16] "brain"                   "subcutaneous"           "bowel"                    "lymph node"
#[21] "peritoneal"              "omentum"                "Cutaneous or Subcutaneous" "Distant Meta

df2$loc_melan2 <- ifelse(df2$loc_melan == "subcutaneous", "Cutaneous or Subcutaneous",
```

```r
                              ifelse(df2$loc_melan == "primary detected", "primary detected", "Other_regions_
unique(df2$sample_type)
```

```
## [1] "Primary Tumor"        "Metastatic"           "melanoma"
## [4] "Additional Metastatic" "Solid Tissue Normal"
```

```r
df2$sample_type2 <- ifelse(df2$sample_type == "Additional Metastatic", "Metastatic",
                           ifelse(df2$sample_type == "melanoma", "Primary Tumor", df2$sample_type))
unique(df2$sample_type2)
```

```
## [1] "Primary Tumor"        "Metastatic"           "Solid Tissue Normal"
```

```r
pdf(file = "treatment_comparison_harmonyzed.pdf", width = 15, height = 10)
ggplot(df2, aes(treatment, deconv_value)) + geom_boxplot(aes(fill = sample_type2)) + facet_grid(sample_
  #theme_classic() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1))
ggplot(df2, aes(treatment, deconv_value)) + geom_boxplot(aes(fill = sample_type2)) + facet_grid(loc_mela
  #theme_classic() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1))
ggplot(df2, aes(treatment, deconv_value)) + geom_boxplot(aes(fill = sample_type2)) + facet_grid(loc_mela
  #theme_classic() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1))
ggplot(df2, aes(treatment, deconv_value)) + geom_boxplot(aes(fill = sample_type2)) + facet_grid(cell_typ
  #theme_classic() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1))
dev.off()
```

```
## pdf
##   2
```