Project 2 Group 4

von Anna, Ann-Sophie und Jana

Load data

```
NCI_TPW_gep_treated <- readRDS(url("https://ndownloader.figshare.com/files/14720180?private_link=db1411
NCI_TPW_gep_untreated <- readRDS(url("https://ndownloader.figshare.com/files/14720183?private_link=db14
NCI_TPW_metadata <- read.delim("https://ndownloader.figshare.com/files/14720186?private_link=db1411debc
NegLogGI50 <- readRDS(url("https://ndownloader.figshare.com/files/ 14720210?private_link=074e0120fe5e68
CCLE_basalexpression <- readRDS(url("https://ndownloader.figshare.com/files/14770127?private_link=fc0c7
CCLE_copynumber <- readRDS(url("https://ndownloader.figshare.com/files/14770130?private_link=fc0c71246d
CCLE_mutations <- readRDS(url("https://ndownloader.figshare.com/files/14770133?private_link=fc0c71246dc
cellline_annotation <-read.delim("https://ndownloader.figshare.com/files/14768981?private_link=efb6a529
drug_annotation <- read.delim("https://ndownloader.figshare.com/files/14768984?private_link=efb6a529eaf
```

1. Broad analysis

Data preparation and annotation

Calculate fold change due to drug treatment

```
fold_changes <- NCI_TPW_gep_treated - NCI_TPW_gep_untreated
fold_changes <- as.data.frame(fold_changes)
```

Renaming of cellline SK-MEL-2 Problem: name of cellline SK-MEL-2 is part of cellline SK-MEL_28 Solution: rename SK-MEL-2 to SK-MEL-2_ (first define it as new factor level)

```
levels(cellline_annotation$Cell_Line_Name) <- c(levels(cellline_annotation$Cell_Line_Name),
                                                "SK-MEL-2_")
cellline_annotation[33, 1] <- "SK-MEL-2_"
#delete level SK-MEL-2 (otherwise we would have 62, instead of 61 levels)
cellline_annotation$Cell_Line_Name <- factor(as.character(
  cellline_annotation$Cell_Line_Name))
```

Create annotation: A matrix is created, which contains for each sample name the drug, cellline and cancertype

1. Drug

```
sample_drug <- as.data.frame(sapply(levels(drug_annotation$Drug), grepl,
                                    colnames(fold_changes), ignore.case = TRUE))
  #creates table with TRUE and FALSE for each sample and drug
rownames(sample_drug) <- colnames(fold_changes)
drugs <- as.vector(apply(sample_drug, 1, function(x){
  colnames(sample_drug[which(x)])
}))
```

2. Cellline

```
sample_cellline <- as.data.frame(sapply(levels(cellline_annotation$Cell_Line_Name), grepl,
                                        colnames(fold_changes), ignore.case = TRUE))
  #creates table with TRUE and FALSE for each sample and cellline
rownames(sample_cellline) <- colnames(fold_changes)
cellline <- as.vector(unlist(apply(sample_cellline, 1, function(x){
  colnames(sample_cellline[which(x)])
})))

annotation <- cbind("Drug" = drugs, "Cellline" = cellline)
rownames(annotation) <- colnames(fold_changes)
```

3. Cancertype

```
cancertype <- sapply(annotation[, 2], function(x){
  #2nd column contains cellline annotation of samples
  cellline_annotation$Cancer_type[cellline_annotation$Cell_Line_Name == x]
})
cancertype <- as.vector(unlist(cancertype))

annotation <- cbind(annotation, "Cancertype" = cancertype)
rm(drugs, sample_drug, cellline, sample_cellline, cancertype)
```

Coloring:
Create a vector which assigns each drug or each cancertype a color

1. According to drug (color_vector_all_drugs)

```
#define a color palette with 15 chosen colors
color_palette_drug <- c("aquamarine", "brown", "forestgreen", "slategrey",
                        "chartreuse", "darkgoldenrod1", "cadetblue","purple",
                        "firebrick1", "deepskyblue", "gold", "violetred4",
                        "deeppink", "plum2", "blue" )
names(color_palette_drug) <- levels(drug_annotation$Drug)

#create vector containing a color name for each sample according to drug
color_vector_drug <- sapply(rownames(annotation), function(x){
  unname(color_palette_drug[annotation[x, 1]]) #first column of annotation contains drug
})
```

2. According to cancertype (color_vector_cancertype)

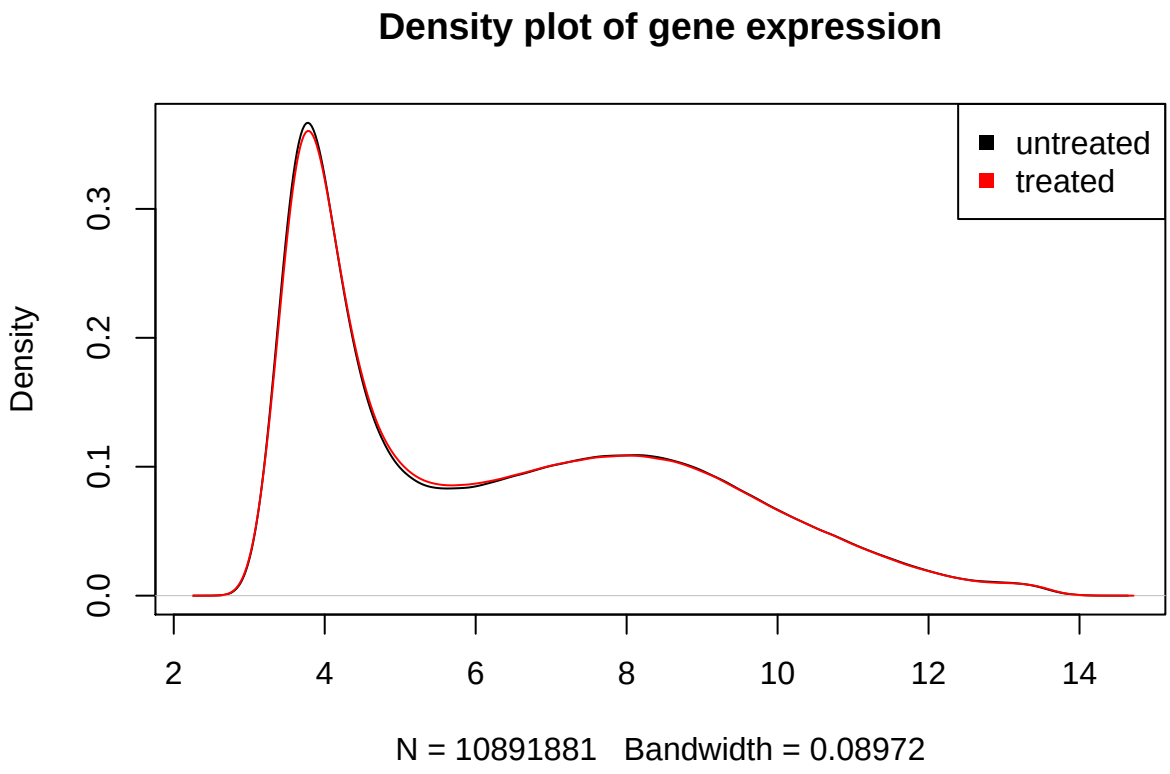
```
#define a color palette with 9 chosen colors
color_palette_cancertype <- c("aquamarine", "brown", "forestgreen", "chartreuse",
                              "darkgoldenrod1", "cadetblue","purple",
                              "firebrick1", "deepskyblue")
names(color_palette_cancertype) <- levels(cellline_annotation$Cancer_type)

#create vector containing a color name for each sample according to cancertype
color_vector_cancertype <- sapply(rownames(annotation), function(x){
  unname(color_palette_cancertype[annotation[x, 3]]) #3rd columns of annotation contains cancertype
})
```

Density plot

To show the distribution of all gene expression values of all samples, a density plot was drawn. The black line contains all values measured for control samples (untreated). In red the distribution of the gene expressiion of all samples treated with 15 drugs is shown.

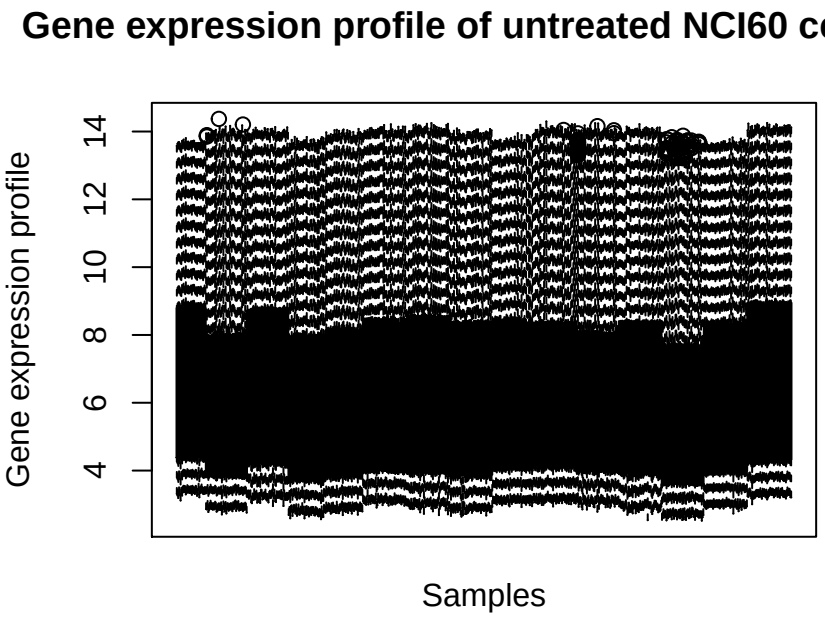
```
plot(density(NCI_TPW_gep_untreated), "Density plot of gene expression")
lines(density(NCI_TPW_gep_treated), col = "red")
legend("topright", legend = c("untreated", "treated"), col = c("black", "red"), pch = 15)
```



Boxplot

Create a boxplot to show the distribution of the foldchanges of all genes in one box per sample

```
#par makes spaces outside the plot larger, xaxt: removes labels on x-axis
#title() used to move xlab nearer to the axis
par(oma = c(1, 1, 1, 8), xpd = "TRUE")
boxplot(NCI_TPW_gep_untreated,
        xaxt = "n",
        ylab = "Gene expression profile",
        vertical =  T,
        main = "Gene expression profile of untreated NCI60 celllines")
title(xlab = "Samples", line = 1.0)
```



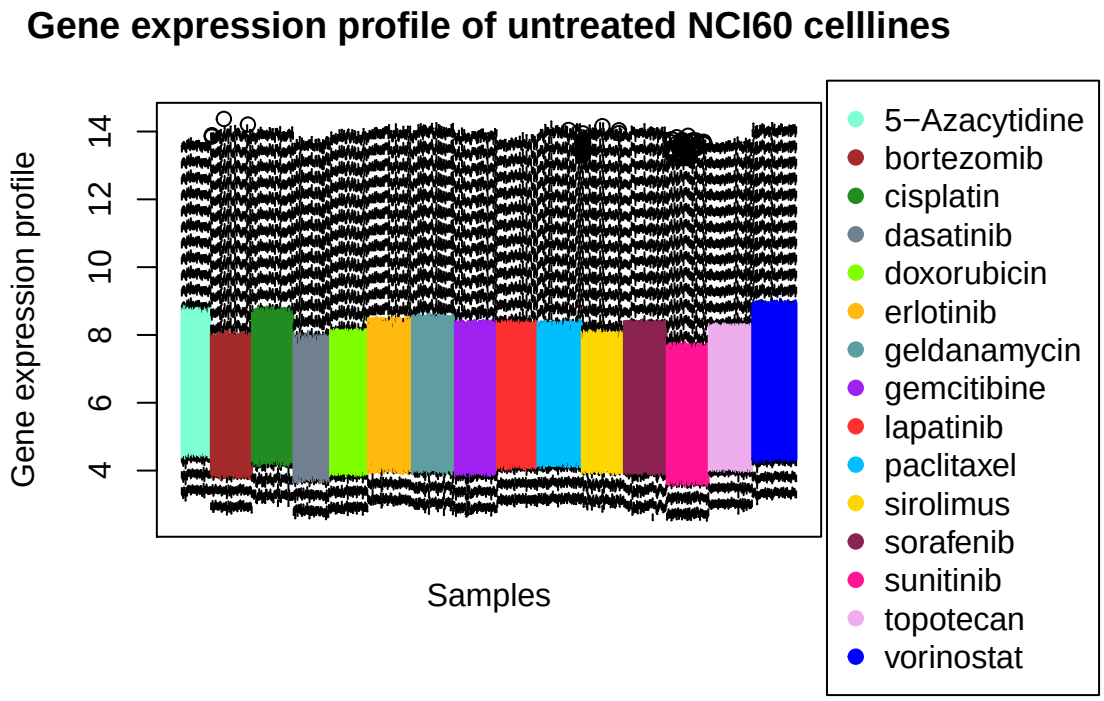
Batch effect was seen –> corresponding to drugs?

Color plot according to drugs

```
par(oma = c(1, 1, 1, 8), xpd = "TRUE")
```

```
## Warning in par(oma = c(1, 1, 1, 8), xpd = "TRUE"): NAs durch Umwandlung
## erzeugt
```

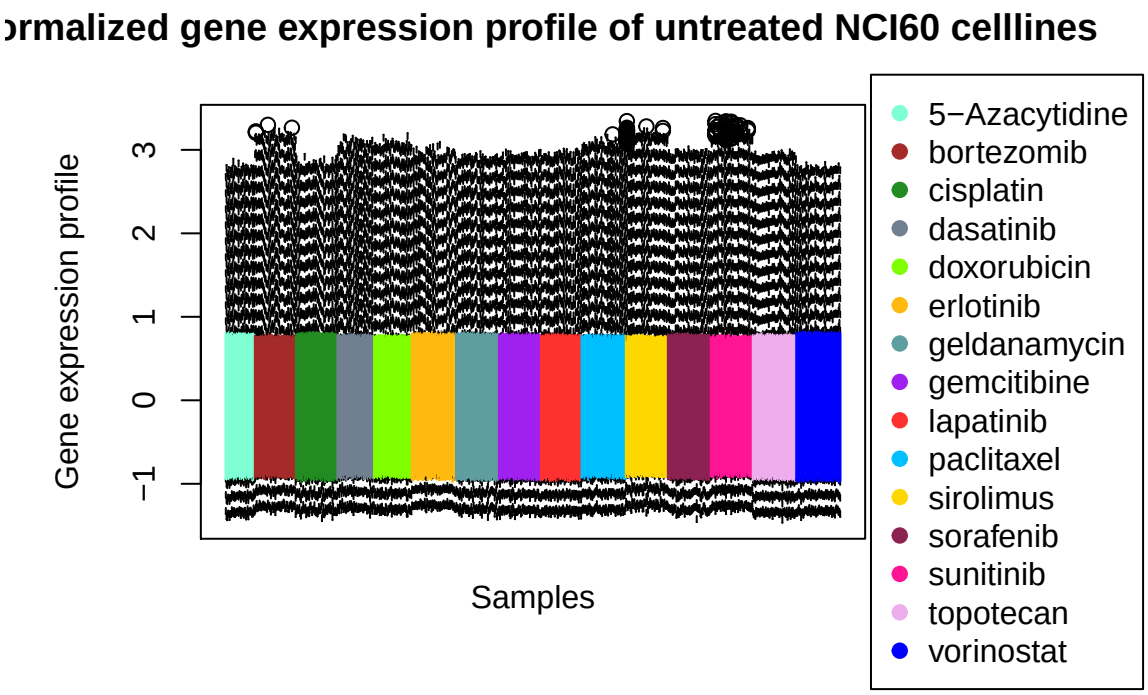
```
boxplot(NCI_TPW_gep_untreated,
        xaxt = "n",
        ylab = "Gene expression profile",
        vertical =  T,
        main = "Gene expression profile of untreated NCI60 celllines",
        boxcol = color_vector_drug)
title(xlab = "Samples", line = 1.0)
legend(x = 860,
       y = 15.5,
       legend = names(color_palette_drug),
       col = color_palette_drug,
       pch = 19)
```



Normalization of data is necessary

```
#each sample should have mean 0 and sd 1
untreated_normalized <- apply(NCI_TPW_gep_untreated, 2, function(x){
  (x - mean(x)) / sd(x)
})
FC_normalized <- apply(fold_changes, 2, function(x){
  (x - mean(x)) / sd(x)
})

#boxplot of normalized untreated values
par(oma = c(1, 1, 1, 8), xpd = "TRUE")
boxplot(untreated_normalized,
        xaxt = "n",
        ylab = "Gene expression profile",
        vertical =  T,
        main = "Normalized gene expression profile of untreated NCI60 celllines",
        boxcol = color_vector_drug)
title(xlab = "Samples", line = 1.0)
legend(x = 860,
       y = 3.9,
       legend = names(color_palette_drug),
       col = color_palette_drug,
       pch = 19)
```



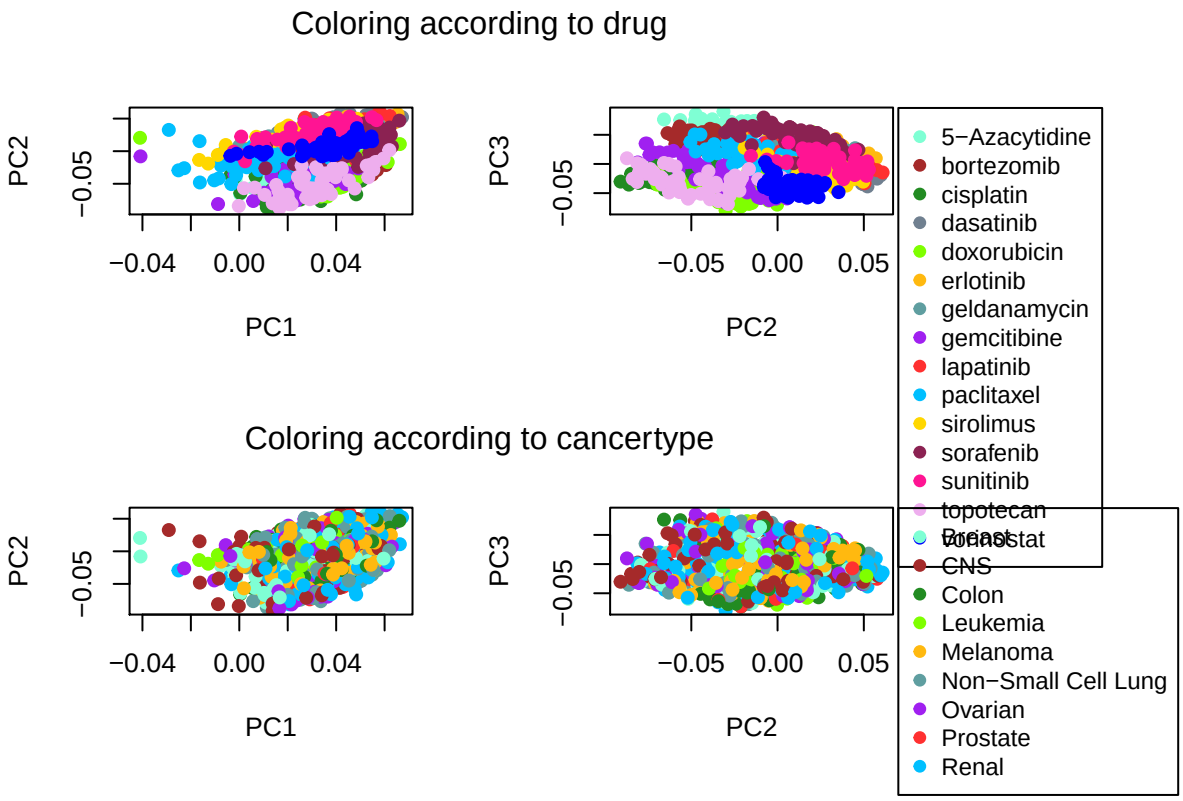
PCA

```
pca <- prcomp(FC_normalized)

#color PCA according to drug
par(oma = c(1, 1, 1, 8), mfrow = c(2, 2)) #mfrow to create multiple plots
#PC1 and PC2
plot(pca$rotation[,1],
     pca$rotation[,2],
     col = color_vector_drug,
     pch = 19,
     xlab = "PC1",
     ylab = "PC2")
#PC2 and PC3
plot(pca$rotation[,2],
     pca$rotation[,3],
     col = color_vector_drug,
     pch = 19, xlab = "PC2",
     ylab = "PC3")
#create legend on the right side
legend(x = 0.07,
       y = 0.096,
       legend = names(color_palette_drug),
       col = color_palette_drug,
```

```
       pch = 19,
       xpd = "TRUE",
       cex = 0.9)
#Title: mtext = margin text, side = 3 (upside)
mtext("Coloring according to drug", side = 3, line = -2, outer = TRUE)

#Color PCA according to cancertype
#PC1 and PC2
plot(pca$rotation[,1],
     pca$rotation[,2],
     col = color_vector_cancertype,
     pch = 19, xlab = "PC1",
     ylab = "PC2")
#PC2 and PC3
plot(pca$rotation[,2],
     pca$rotation[,3],
     col = color_vector_cancertype,
     pch = 19, xlab = "PC2",
     ylab = "PC3")
legend(x = 0.07,
       y = 0.096,
       legend = names(color_palette_cancertype),
       col = color_palette_cancertype,
       pch = 19,
       xpd = "TRUE",
       cex = 0.9)
mtext("Coloring according to cancertype", side = 3, line = -15, outer = TRUE)
```




```
rm(pca)
```

Biomarkers

Barplot and Boxplot

2. Specific analysis: Erlotinib