HW3

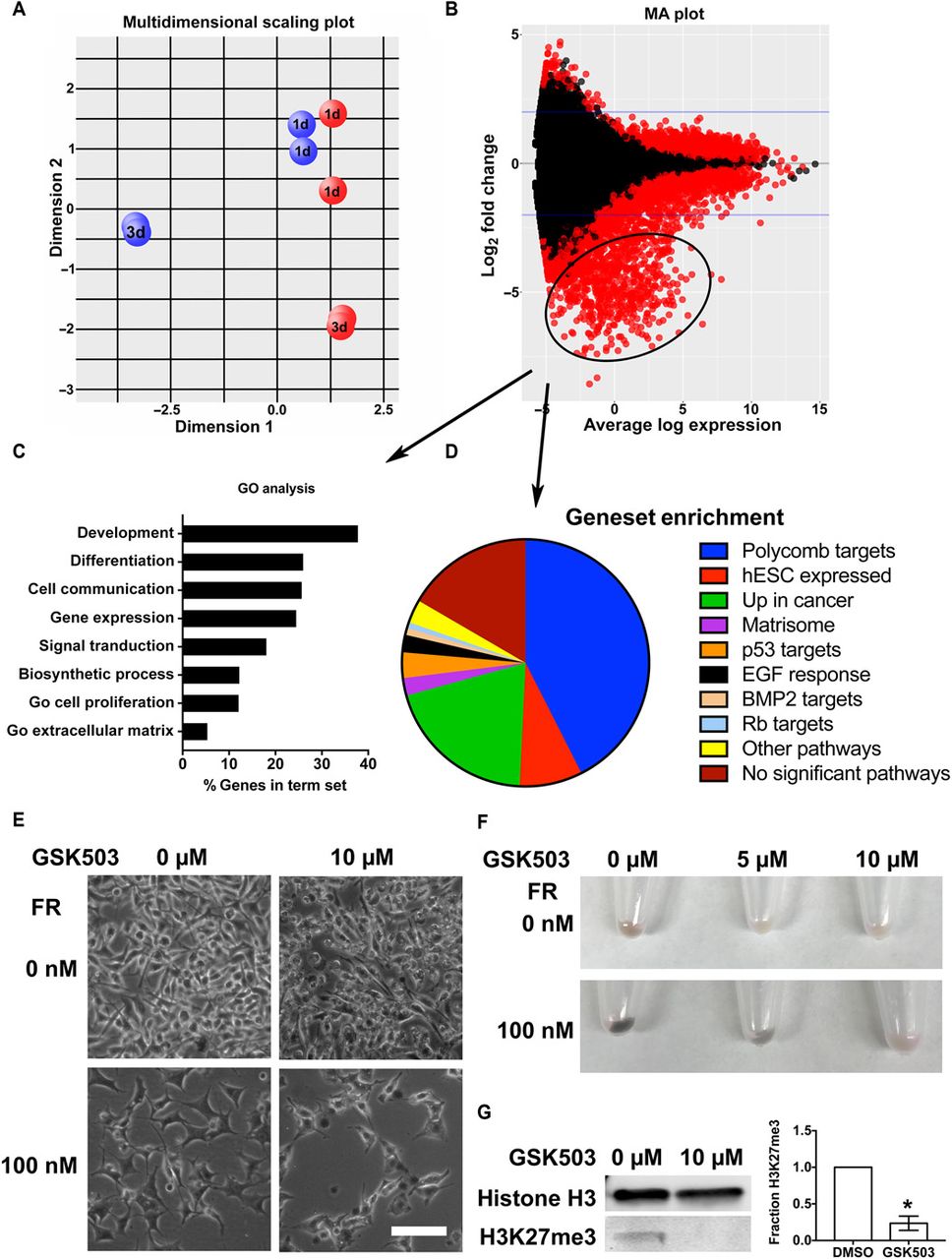
# Introduction

Uveal melanoma (UM) is the most common intraocular cancer afflicting 7 in 1,000,000 individuals in the U.S and UK. 90% of uveal melanoma cases have mutations in the GNAQ/11 genes. These genes encode for G-protein coupled receptor (GPCR) subunits found within the inner leaflet of the plasma membrane that act as an on/off switch for a wide range of cellular processes. Mutations in the GNAQ/GNA11 genes lead to constitutively active GPCRs, which causes increased cell growth and proliferation. In the paper referenced below the authors demonstrate that FR900359 (FR) prevents GDP/GTP exchange in GRPCs; a crucial step in signal transduction, via small molecule allosteric inhibition. Inactivation of the primary mutated pathway found in most UM cases leads to arrest of cell proliferation, inhibition of secondary signaling, and reinstated melanocyte differentiation leading to potential therapeutic uses.

In this report I plan to reproduce the MA plot in Fig.6 B from Onken et al. 2018 with RNA expression data published with the paper. This figure depicts the fold change vs. average log expression of RNA between FR and non FR treated cells. Identification of gene clusters suppressed by FR, or not, helps us determine whether FR will be a useful treatment in UM. Once I’ve recreated the MA plot depicted below I plan on using my own RNAseq data collected from cells treated with a B-Raf inhibitor. B-Raf is a proto-oncogenic serine/threonine kinase that is commonly expressed in FR resistant cells. Expression profiles of FR treated cells, and B-Raf inhibitor treated cells, allows us to compare their individual efficacy while also illuminating a potential multidrug treatment.

# Reference

Onken MD, Makepeace CM, Kaltenbronn KM, et al. Targeting nucleotide exchange to inhibit constitutively active G protein α subunits in cancer cells. Science Signaling. 2018;11(546):eaao6852. <doi:10.1126/scisignal.aao6852>



MA Plot

Fig. 6. FR represses expression of differentiation genes by restoring function of the PRC2. (A)Gaq-mutant 92.1 UM cells were treated with FR or vehicle, and RNA was collected 1 and 3 days (1d and 3d, respectively) after treatment for RNAseq analysis. Results of a multidimensional gene expression analysis that compares the relative patterns of expression of all genes across all samples and groups genes with similar patterns. The graph shows samples positioned by their relative gene expression values within each pattern. Dimension 1 (x axis; the most represented pattern) shows separation based on vehicle treatment (red balls) versus FR treatment (blue balls), whereas dimension 2 (y axis; the second most represented pattern) shows separation based on time in culture (indicated by 1d or 3d on balls). (B)MA plot (M, log ratio; A, mean average) comparing gene expression between FR- and vehicle-treated 92.1 samples identifies a group of significantly reduced genes (circled; fold change, >2; FDR, q <0.01) associated with FR treatment. (C) GO analysis of the FR-repressed gene set [circled in (B) with arrow]. (D) FR-repressed genes [circled in (B) with arrow] identified as targets of the polycomb repressive complex2(PRC2)byGSEA.EGF,epidermalgrowth factor; BMP2, bone morphogenetic protein; hESC, human embryonic stem cells. (E) Effect of the EZH1/2 inhibitor GSK503 on morphological differentiation elicited by FR. Representative fields are shown from one of three experiments of 92.1 UM cells treated for 7 days with GSK503 and for 3 days with FR and then imaged by phase-contrast microscopy. Scale bar, 100 μm. (F) Effect of GSK503 on pigmentation of FR-treated cells, visualized by macroscopic inspection. 92.1 cells were treated for 7 days with GSK503 and for 3 days with FR and pelleted; representative images from one of three experiments. (G) PRC2 inhibition by GSK503. Immunoblots of 92.1 cells treated for 7 days with GSK503 show reduced histone H3K27 trimethylation. Plot shows relative fraction of trimethyl-histone H3K27 compared to dimethyl sulfoxide (DMSO) control and normalized to total histone H3 from densitometry data from three independent experiments. \*P < 0.01 by t test; significance was confirmed using q < 0.01 by the FDR method of Benjamini and Hochberg.

# Materials and Methods

To collect the data seen in Figure 6 B the authors grew 92.1 uveal melanoma cells in 100 nM FR or DMSO in RPMI growth medium. After 1 and 3 days of treatment RNA was collected from each treatment. HiSeq2500 was used to generate FastQ raw data. This data was aligned to the whole genome using Bioconductor in EdgeR. Once aligned, a 2.0 fold change in expression was considered significant comparing FR treated to non-FR treated cells. This data was then plotted as an MA plot, M = log ratio and A = mean average.

Since I have a CSV file containing the log ratio and mean averages, I’ll make the volcano plot using Rstudio. The next step I need to take now that I’ve generated the MA plot is to trim some of the data. I’ve set the perimeters to only graph fold change >2.0 but I need to filter out results with a significance less than 10%. The authors set their ‘*q*’<0.01 meaning only genes with a false discovery rate (FDR) greater than 10% are graphed. I however, do not have the statistical analysis available to me. Instead I’ll have to derive significance from the change between days.

#Install needed packages   
options(repos = c(CRAN = "https://cran.r-project.org"))  
install.packages('R.utils')

## package 'R.utils' successfully unpacked and MD5 sums checked  
##   
## The downloaded binary packages are in  
## C:\Users\eanderson34\AppData\Local\Temp\RtmpWUpBdg\downloaded\_packages

library(data.table)  
install.packages("tidyverse")

## package 'tidyverse' successfully unpacked and MD5 sums checked  
##   
## The downloaded binary packages are in  
## C:\Users\eanderson34\AppData\Local\Temp\RtmpWUpBdg\downloaded\_packages

library(tidyverse)

#Import data files   
GSM2781365\_sample.92.1\_d1\_1.txt <- fread("Data/GSM2781365\_sample.92.1\_d1\_1.txt.gz")  
GSM2781367\_sample.92.1\_d3\_1.txt <- fread("Data/GSM2781367\_sample.92.1\_d3\_1.txt.gz")  
GSM2781369\_sample.92.1\_fr1\_1.txt <- fread("Data/GSM2781369\_sample.92.1\_fr1\_1.txt.gz")  
GSM2781371\_sample.92.1\_fr3\_1.txt <- fread("Data/GSM2781371\_sample.92.1\_fr3\_1.txt.gz")

#Renames files   
d1 <- GSM2781365\_sample.92.1\_d1\_1.txt[-1,]  
d3 <- GSM2781367\_sample.92.1\_d3\_1.txt[-1,]  
d1\_FR <- GSM2781369\_sample.92.1\_fr1\_1.txt[-1,]  
d3\_FR <- GSM2781371\_sample.92.1\_fr3\_1.txt[-1,]

# Create a new data table from columns in other tables  
expression\_data\_d1 <- data.table("Gene\_Name" = d1$"external\_gene\_name",  
 "D1\_Expression" = d1$"sample.92.1\_d1\_1",  
 "D1\_FR\_Expression" = d1\_FR$"sample.92.1\_fr1\_1")  
  
expression\_data\_d3 <- data.table("Gene Name" = d1$"external\_gene\_name",  
 "D3\_Expression" = d3$"sample.92.1\_d3\_1",  
 "D3\_FR\_Expression" = d3\_FR$"sample.92.1\_fr3\_1")  
data <- data.table("D1\_Expression" = d1$"sample.92.1\_d1\_1",  
 "D1\_FR\_Expression" = d1\_FR$"sample.92.1\_fr1\_1",  
 "D3\_Expression" = d3$"sample.92.1\_d3\_1",  
 "D3\_FR\_Expression" = d3\_FR$"sample.92.1\_fr3\_1")

#Convert Data table characters into Numbers   
expression\_data\_d1$D1\_Expression <- as.numeric(expression\_data\_d1$D1\_Expression)  
expression\_data\_d1$D1\_FR\_Expression <- as.numeric(expression\_data\_d1$D1\_FR\_Expression)  
expression\_data\_d3$D3\_Expression <- as.numeric(expression\_data\_d3$D3\_Expression)  
expression\_data\_d3$D3\_FR\_Expression <- as.numeric(expression\_data\_d3$D3\_FR\_Expression)  
data$D1\_Expression <- as.numeric(expression\_data\_d1$D1\_Expression)  
data$D1\_FR\_Expression <- as.numeric(expression\_data\_d1$D1\_FR\_Expression)  
data$D3\_Expression <- as.numeric(expression\_data\_d3$D3\_Expression)  
data$D3\_FR\_Expression <- as.numeric(expression\_data\_d3$D3\_FR\_Expression)

# Calculate the mean expression level (M) and the log fold change (A) for each gene  
logFC\_D1 <- log2(data$D1\_FR\_Expression / data$D1\_Expression)

## Warning: NaNs produced

aveExpr\_D1 <- log2(rowMeans(data[,c("D1\_FR\_Expression","D1\_Expression")], na.rm = TRUE))

## Warning: NaNs produced

logFC\_D3 <- log2(data$D3\_FR\_Expression / data$D3\_Expression)

## Warning: NaNs produced

aveExpr\_D3 <- log2(rowMeans(data[,c("D3\_FR\_Expression","D3\_Expression")], na.rm = TRUE))

## Warning: NaNs produced

# Combine the M and A values with the Day column in a data frame  
MA\_data <- data.frame(Day = rep(c("D1", "D3"), each = length(logFC\_D1)),  
 logFC = c(logFC\_D1, logFC\_D3),  
 aveExpr = c(aveExpr\_D1, aveExpr\_D3))  
# Set threshold for outlier removal  
logFC\_threshold <- 2  
# Remove outliers  
MA\_plot\_filtered <- MA\_data %+% subset(data, abs(logFC\_D1) < logFC\_threshold & logFC\_D1 > 2)  
  
# Create the MA plot  
ggplot(MA\_data, aes(x = aveExpr, y = logFC, color = Day)) +  
 geom\_point(alpha = 0.5) +  
 ggtitle("MA Plot") +  
 xlab("Average Log Expression") +  
 ylab("Log Fold Change") +  
 theme\_bw()+  
 xlim(-5, 5) +  
 ylim(-5, 5)

## Warning: Removed 47577 rows containing missing values (`geom\_point()`).

