

- We're going to work on some data that exists in R already. This experiment was carried out using zebrafish as a model organism to study early development in vertebrates. Swirl is a point mutant in the BMP2 gene that affects the dorsal/ventral body axis. These data were provided by Katrin Wuennenberg-Stapleton from the Ngai Lab at UC Berkeley. The swirl embryos for this experiment were provided by David Kimelman and David Raible at the University of Washington.

- `library(limma)`

- To get the data to work on:

```
swirl.tg <- readTargets("SwirlSample.txt")  
swirl.rg <- read.maimages(swirl.tg$FileName, source="spot")
```

```
swirl.rg$genes <- readGAL("fish.gal")
```

```
swirl.rg$printer <- getLayout(swirl.rg$genes)
```

- To normalize the data:

```
swirl.ma <- normalizeWithinArrays(swirl.rg, method="none")  
swirl.med <- normalizeWithinArrays(swirl.rg, method="median")  
swirl.norm <- normalizeWithinArrays(swirl.rg, method="loess")
```

1. Use the command `?normalizeWithinArrays` to answer the following question: Did you correct for the background when you normalized above? If so, how did the correction work?
2. What are the different components in an `RGList` (note that `swirl.rg` is an `RGList` class)? Use the command `names`. Why do you think they call the class `RGList`?
3. What are the different components in an `MAList` (note that `swirl.ma`, `swirl.med`, and `swirl.norm` are all `MAList` class)? Use the command `names`. Why do you think they call the class `MAList`?
4. Why is it that `dim(swirl.rg$genes)` works and `dim(swirl.rg$printer)` doesn't work?
5. For the RG data, array 3, how many red foreground values are larger than the red background values? How many green foreground foreground values are larger than the green background values?
6. For the RG data, array 3, is the red background more intense than the green background or vice versa?

7. We had the following function with the wine data to bump up low intensities:

```
wine.filt2 <- function(x,threshold=100){  
  x$R <- ifelse(x$R < threshold, threshold, x$R)  
  x$G <- ifelse(x$G < threshold, threshold, x$G)  
  return(x)}  
wine.dat2 <- wine.filt2(wine.dat)
```

Adjust the function so that it works on the swirl data. How many of the original red values did you bump up?

8. Create 4 boxplots of the overall intensities (not of the ratios) of your 4 arrays for one of your above normalization methods. This should only be one very short line of code.
9. Find the mean and standard deviation of a random gene (what data will you use??) Note that the swirl data doesn't have any missing values, but your data surely does. What can you do to deal with missing data when computing the mean and standard deviation? (Note: try `?mean` or `?sd`.)
10. Give a scatterplot (`plot`) for two random genes. Change the axes to something meaningful (within the `plot` function, add `xlab="whatever I want to say"`, `ylab="something else I want to say"`. Note that `xlab` stands for x label).