**WORKSHOP 1 2nd October 2017**

**Worksheet – welcome to the ggunle.**

**Tutor: Brendan Palmer**

**Type out the code you used for each part below the task**

**Part A: Some basics**

1. Save all the files you were emailed to a new folder on your computer and create an R project file to that folder
2. Load the tidyverse package using the library() function
3. Create a new object called x using the assign command (<-) and give it a numerical value
4. Create a new object called y and assign a sequence of numbers to it (e.g 1:10)
5. Create a string character (“I am a one of these”) and call it z. You need to include the quotation marks
6. Use the “c()” command to create an object with more than one character string, each of which is separated by a comma (e.g. “I”, “am”, “one”, “of”, “these”)
7. Check the type of objects you have created using the function typeof(). What is returned?
8. Write a line of text preceded by the hash (“#”) symbol and hit return. What happens?

PART B: Examination of the Brauer2008 data set

1. Open the data set using Notepad or another basic text editor. What makes this data set untidy?
2. Use the import button in RStudio to import the data. How does the import view change if you select Comma instead of Tab under the options for “Delimiter”? What happens when you untick the “First Row as Names” box?
3. Looking at the “NAME” column, can you suggest a strategy to separate the information it contains?

PART C: What’s going on in the script1 code file?

1. View separated\_gene\_df and mutated\_gene\_df in the Editor window of RStudio. Can you see any differences?
2. The function head() gives a view of the top of your data set and the function tail() gives a view of the bottom. The symbol $ allows you to define a specific column.

What changes were applied to the data when we used mutate\_at() function in the Script 2 code?

Clue: Type head(mutated\_gene\_df$BP) and compare it to head(separated\_gene\_df). Can you see any differences now?

1. How many columns are in the selected\_gene\_df?

Tip: The function ncol() might help here

1. How many rows are in the selected\_gene\_df data set?

The function nrow() could come in handy

1. Can the number of rows and number of columns be found by examining the RStudio workspace window?
2. How many rows and columns does the gathered\_gene\_df data set have?
3. Look at your workspace and note all the individual data sets we have created. Now, clear your work space and re-run script 1. Can you see any advantages to using the %<% operator and the tidyverse?

PART D: Plotting the data

1. What are the three main elements needed to generate a ggplot2 graph?
2. By examining the ggplot2 geoms, which will allow you to draw a histogram?

Hint: type ggplot2:: into the RStudio console and scroll through the options

1. Re-produce a plot by replacing “LEU1” in the last code chunk of script1 with any other gene name of your choice.