**Student Handout B: Understanding Log Transformations**

*Learning objectives:*

1) Describe log transformation and the effect of log transformation on data analysis.

Case study using: Tomasetti, C. & Vogelstein, B. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* **347**, 78–81 (2015).

The data in this study compared lifetime cancer risk to total stem cell divisions in a variety of tissue types. The data spanned over five orders of magnitude.

* + Lifetime cancer risk ranged from 0.0000302 (Osteosarcoma of the head) to 1 (Colorectal adenocarcinoma with FAP)
	+ Total number of stem cell divisions ranged from 3.150 x 106 (osteosarcoma of the pelvis) to 3.550 x 1012 (basal cell carcinoma)

To analyze this data, logarithmic transformation of the data was employed. This exercise is designed to help you review log functions and to better understand logarithmic transformation.

*In 2015, C. Tomasetti & B. Vogelstein published a study asking if the overall risk of cancer over a person’s lifetime for a given tissue increase for tissues with a greater number of stem cell divisions (and therefore greater amount of DNA replication). The data in figure 1 is from this paper. The following exercises will help you to better understand the data presented in this case study.*



*Figure 1*

The raw data collected by researchers in this paper are large and these values are broad in range**.** Log transformation was used to change these large numbers into something more easily comparable. Log scale transformation consists of taking the log of an original value and plotting the result.

For example:



Additionally, for a scatterplot, converting values to the log scale can change weak relationships into a potentially linear relationship. By only plotting exponents we are reducing the variability of the values. To understand log transformation, we must first practice expressing numbers in scientific notation, percentage, and log form. To practice the concept of logs, remember that:

* + The log of x will be the exponent on 10 that gives the value of x.
	+ y = log(x), so 10Y = x. For example: 2= log(100) because 102 = 100
	+ *Remember: The answer to a log is an exponent!*

*Without a calculator, practice calculating the approximate log scale value of 54,321:*

1. Express this number in scientific notation
2. What is the approximate log scale value?

*Refer to Figure 1. How was melanoma plotted in this figure?*

1. Consider the x-axis. For melanoma, the total stem cell divisions equals 7.638 x 1011. Calculate the value that will be used on the x-axis on the scatterplot.
2. Consider the y-axis. For melanoma, the lifetime cancer risk is 2.3%. Calculate the value that will be used on the y-axis on the scatterplot.

**Discussion Question**

Why did researchers transform the raw data of this study into log form?

**References**

1. Tomasetti, C. & Vogelstein, B. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* **347**, 78–81 (2015); <http://science.sciencemag.org/content/347/6217/78.full>
2. OpenStax CNX; https://cnx.org/contents/esgfrPlv@3/Accessory-Organs-in-Digestion-
3. HHMI Biointeractive; <https://www.hhmi.org/biointeractive/data-points>
4. Khan Academy; https://www.khanacademy.org
5. Math Bench Biology Modules; <http://mathbench.umd.edu/modules/misc_scaling/page07.htm>