1. What is the hypothesis of the HiHost2 project? Are there any assumptions readers, users, and authors should keep in mind?

2. Describe the data in sheet 1. What metadata are included? What are column names? What do the values represent? How were these values obtained? What is the reference sequence?

3. Refer to the histogram of variant frequencies. Are there more rare than common variants in this region?

Explore the data with your group to answer the following questions:

4. How many variants are in this region?

5. Are there more cell lines from males or female donors? How many? How does this look at the population level?

6. Which population has, on average, the largest number of variants? Describe the data transformations that occur to come to the answer. What's the standard deviation for each population? *(bonus: are these differences statistically significant?)*

7. How many unique variants occur in each population? How many variants are shared between only 2 populations?

Normalization & Log2-transformation:

8. What is the purpose of taking the log2() of the data?

9. How might you interpret the normality of the transformed data if you were only looking at the statistical test versus the QQ plots?

10. What does it mean that the data may be "confounded" by population and/or sex?

11. Are T-test and ANOVA p-values more or less significant when run on data that has *not* been log-transformed?

Use the Results data file to answer the following questions with your group:

12. Describe the data in sheet 2. What is Beta? P-value?

13. For the 2 variants with the most significant p-values, is the major or minor (most or least prevalent) allele associated with higher IP10 levels during infection?

14. Use UCSC human genome browser to locate these variants. What genes are they in or nearby? Are they in coding regions? If not, how might these variants affect CXCL10 expression?

15. What future projects or questions might you propose based on these results? How might you caution against using these results to directly inform treatment of Chlamydia infections?