

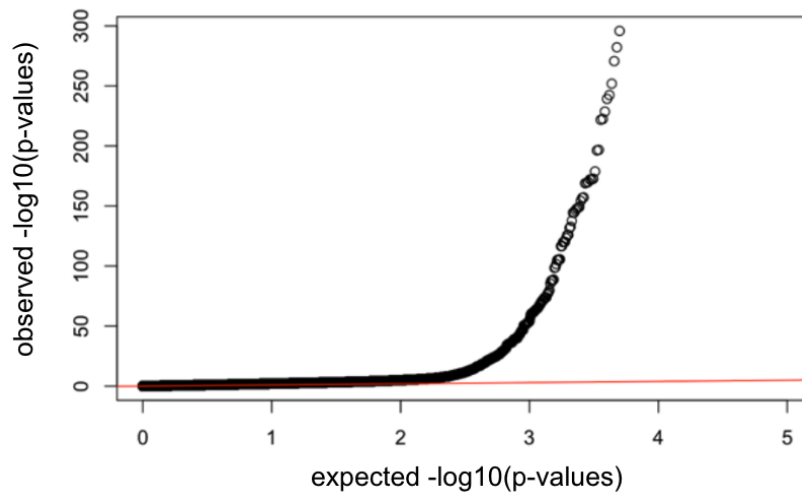
CMM 262 2023

Homework 2: GWAS and single cell RNA-seq (Graham McVicker and Robert Morey)

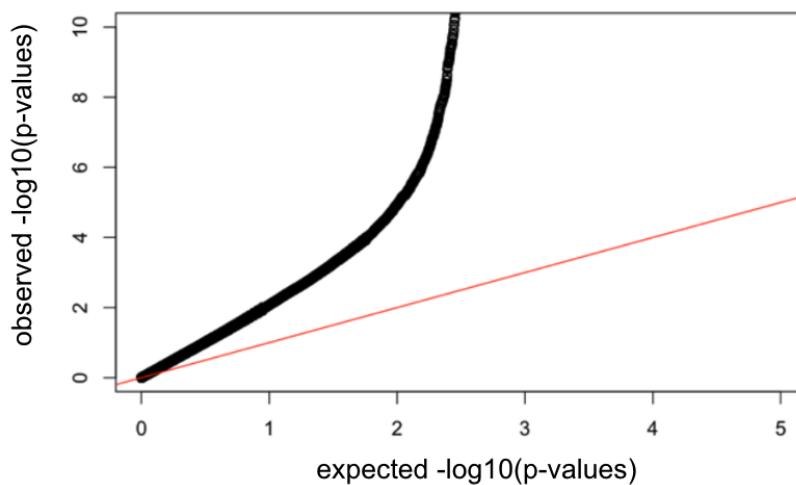
Instructions

Answer the [following questions](#) in your own words and upload a PDF of your answers to Gradescope. Make sure to write your name and PID at the start of your answers. This assignment is due **2/16/23 at 9:00AM**.

Part 1: GWAS



Upon zoom in:



Shown above is a qq-plot for a list of p-values. Answer the following questions by referring to this plot.

1. Under the null hypothesis, how should the p-values be distributed (i.e. What distribution should they come from)? (1 point)
2. Do you notice a problem with the p-values from the zoomed-in plot? What word best

describes what is occurring with the p-values? (1 point)

3. Are we more likely to have false negatives or false positives? (1 point)
4. What could cause this problem with the p-values? Discuss two different explanations for what could cause this problem for full credit. (2 points)
5. How could you correct this problem experimentally? Provide one suggested experimental correction for each explanation you listed above. (2 points)

Part 2: Single cell RNA-seq

6. Describe the experimental differences between droplet based methods and physical separation methods. Start by explaining each. Then discuss when it might be better to use one or the other. (3 point)
7. When analyzing scRNA-seq data, how do you know which cell a read comes from? (1 point)
8. Describe what a doublet is and how it is generated. Why do you not want doublets in your dataset when you perform analysis? (2 points)
9. When preprocessing scRNA-seq data, there are multiple metrics to perform QC on. Please list three QC metrics and explain what each controls for? (3 points)
10. Come up with an experiment in which you'd prefer to use scRNA-seq rather than RNA-seq. Justify why you believe this assay would be preferable for this experiment. (2 points)