



BIOINFORMATICS INTERDEPARTMENTAL
GRADUATE PROGRAM
611 CHARLES E. YOUNG DRIVE EAST
LOS ANGELES, CALIFORNIA 90095-7246

September 12, 2022

RE: Written Qualifying Examination Results for Sandy Kim [REDACTED]

Dear Sandy,

The Bioinformatics IDP Faculty is happy to inform you that you have successfully **PASSED** your WQE. Please see evaluator notes. If you have questions, please let [REDACTED] know as soon as possible. Once again, congratulations on passing your WQE.

Sincerely,

A handwritten signature in black ink, appearing to read "Hee Ky", is written over a black rectangular redaction box.

[REDACTED]
Student Affairs Officer
Bioinformatics IDP
[REDACTED]

Evaluators #1

Overall, this is a fantastic WQE! The research question is very interesting, and Sandy brings up some extremely good points, in particular that normalization methods suppress variation, which could interfere with outlier detection in RNA-seq experiments. The proposed idea here, to use Bayesian methods on raw counts for outlier detection, is very promising and well-motivated. The idea was implemented well, and the design for testing the approach was extremely convincing.

My only comments here have to do with the writing and perhaps a bit the motivation for the choice of various prior distributions from which the elements of the Bayesian model were drawn. In the Methods, it is very unclear why certain choices were made--why have outliers with an $e^I * \mu$ difference in the mean from the “non-outliers?” I know why the counts were sampled from a negative binomial (though maybe it would be useful to have zero inflation?), but the dispersion of 100 there seems arbitrary. Also, a number of other choices felt arbitrary, especially the choice of prior distributions in the inference model (though perhaps that does not matter much). Figures 1 and 2, while technically correct, did not really make it easier to understand how either model worked, or what the motivation for various technical choices was. None of these issues, however, are very major.

The results were presented in a clear manner, and the analysis was convincing. Given the constraints of a WQE, even promising and interesting ideas can't be developed beyond a certain point. Overall, this was a fantastic project.

Evaluators #2

This proposal describes a method to identify outliers in RNA seq data. Existing methods rely on PCA and are somewhat arbitrary so it makes sense to attempt to develop a more rigorous probabilistic framework. While the paper accomplished this goal, I have two critiques that would like to see addressed:

- 1) it would be good to apply the existing PCA methods to the simulated data to see if they in fact perform better than the proposed method
- 2) I found the description of the inference model difficult to follow and was unable to evaluate whether the approach is robust or not