Review of: "Improving an rRNA depletion protocol with statistical design of experiments"

Christos Papaneophytou[1]

1 University of Nicosia

Potential competing interests: The author(s) declared that no potential competing interests exist.

In this work, a protocol for rRNA depletion was improved using the Design of Experiments (DoE) approach. This is an interesting topic rRNA depletion from prokaryotic RNA-seq library reaction mixtures is an extensive and time-consuming process. Specifically, the authors employed a series of response surface (RSM) designs to optimize the rRNA depletion from the reaction mixtures. In the first design, a 2-factor/5- level central composite design (CCD) was employed to identify the best combination of 23S and 16S probes that minimize the presence of rRNA in the reaction mixtures. Subsequently, a 3-factor/5-level CCD was used to identify the optimum values of probes, RNA, and beads in the reaction mixture to minimize further the efficiency of RNA removal, while a third design (according to the authors a 2^2 factorial design) was also employed.

There are several concerns with this study which include the following:

1. It is not clear, at least to me, why three different DoE designs were required to identify the best combination of factors that minimize the responses. The authors should repeat the experiments and examine simultaneously the effect of the amount of the two probes (i.e., 23S- and 16S- probe), the amount of RNA, and the volume of beads in a single experiment using a 4-factor/5-level CCD because according to the equation in page 4, the interactions of the probe amount with both RNA amount and volume of beads do affect the response.
2. The authors did not provide sufficient information including the mathematical model for the first RSM. The main conclusion of this experiment is unclear.
3. The third design is unnecessary. The best compilation of factors that minimize (or maximize) the response can be easily identified from response surface plots and/or contour plots (graphical optimization) as well using numerical optimization. However, the response surface plots have not been included for both designs. Importantly, 2^2 design represents is a full factorial design with 4 runs and not a fractional factorial consisting of 6 runs.
4. Analysis of Variance (ANOVA), validation of the two models, and quality of fit for the RSM designs are missing.
5. The quality of a model in RSM is assessed using not only the coefficient of determination (R^2), adjusted R^2 and “adequate precision” and the “lack of fit” but also using the Normal plots of residuals and predicted vs actual values plot.
6. The authors say that “we used DoE and RSM” without explaining what DoE designs, in addition to RSM, they used.
7. The statement “A linear model is fit to the experimental data” is incorrect because in RSM a second-order equation (quadratic model) is fitted to the experimental data.
8. The authors didn’t explain in detail i) why they have chosen the central composite design over other designs, for

example, the Box–Behnken design, and ii) how they have selected the levels of all variables.

9. Figure 2, panel A is too confusing.
10. In several instances, the results in the main text do not align with the results presented in the supplementary tables, for example, the results discussed in lanes 103-108 do not align with the results presented in Supplementary Table 2.