Team DR-1 (DEFEND)

Overall impact score: 2
Overall impact summary: The investigators thoughtfully proposed a well-validated approach to tackle an important problem. My main concern is that the primary problem has already been tackled by the Osoktsky preprint, as cited by the investigators. I would recommend that the investigators either add further contributions to set themselves apart, such as by using single-cell RNA-seq endometriosis data (see Ma et. al., Cell & Bioscience), or focus on contributions that differ from the original preprint, such as by minimizing side effects.

Significance score: 1
Significance summary: The need for identifying drugs to tackle endometriosis was articulated well. The fact that 11% of American women are impacted by endometriosis but that no targeted therapies exist is a clear problem.

Innovation score: 3
Innovation summary: The investigators plan to use a well-validated approach: signature reversion strategy on differential expression signatures. This approach has been performed by Osoktsky using the microarray dataset GSE51981, as the authors noted. The authors noted some improvements that could be made to Osoktsky's approach, such as the use of additional public expression datasets, as well as validation in EHR systems outside of the University of California.

Although the investigators' claim that Osoktsky's paper failed to demonstrate systematic prioritization, I would not necessarily agree. Osoktsky's prioritization was systematic, but was based entirely on candidates' reversal scores. This proposal's primary innovation is in the incorporation of downstream steps within the prioritization scheme: clinical trials, usage within the EHR, or side effect information. In fact, the last data point is missing entirely within the Osoktsky paper.

Approach score: 2
Approach summary: The team's approach is well thought-out and focuses on well-established, accessible datasets.

I am curious about the choice of datasets, the RNA-seq dataset GSE134052 and the microarray dataset GSE51981. (1) How do the authors plan to use GSE51981 (77 cases diagnosed with endometriosis, 37 abnormal controls with uterine/pelvic pathology, and 34 normal controls without uterine/pelvic pathology)? Will the authors treat the abnormal controls differently from the normal controls? Will the authors use stage and phase? (2) Why did the authors choose an RNA-seq and a microarray dataset, instead of two RNA-seq datasets or two microarray datasets, in order to minimize variance? Is this to

maximize the sample size for power calculations? (3) I would also recommend looking at non-GSEA datasets, such as EndometDB.

For information on side effects, I would recommend that the investigators look into additional drug safety datasets. OffSIDES is built on the FDA Adverse Event Reporting System (FAERS), and is thus focused on post-marketing safety. Databases such as OnSIDES include side effects detected during clinical trials, and are also important for patient safety.

I would recommend that the authors clearly justify the prioritization scheme, which is a primary differentiator from the Osoktsky preprint. For example: (1) How will candidates be chosen, given that reversal scores are continuous? (2) Is concordance between datasets more important than previous use in treating non-endometriosis patients? If not, should some elements be weighted? (3) Should candidates that failed clinical trials be included (and rewarded), depending on the reason they failed? (4) Why do we reward candidates used to treat non-endometriosis patients?

Investigators score: 1
Investigators summary: The investigators are well equipped to tackle this problem. They are all PhD or Masters candidates with the Department of Biomedical Data Science with diverse skill sets in genomics, statistics, computer science, and biomedicine.

Environment score: 1
Environment summary: The investigators are at Stanford with good access to computational resources. I would recommend that the investigators communicate with a professor within the department with a record of working on RNA-seq data and women's health, such as Barbara Engelhardt.