We would like to thank the reviewer for his/her comments. The resulting changes will provide a more robust evaluation and description of the Meta-SPS approach.

# RESPONSES TO THE REVIEWER

**Reviewer comment #1**: “The first issue relates to the documentation and support of the approach. In order to make this manuscript more approachable by the proteomics community, the authors need to expand their supplemental section in the following manner. The authors should assemble in an orderly manner the MS/MS, for example, that allowed them to determine the aBLTA protein, for example. Each MS/MS would have the meta-SPS pipeline interpretation of that individual MS/MS spectra. Next, the authors need to show how the individual MS/MS spectra were assembled to provide additional sequence assembly information. Essentially, what would greatly strengthen this manuscript is if the authors expanded the supplement and provided one detailed example with mass spectrometry data that walks the reader through Figure 1a, which is the overview.”

*Although the last paragraph of the results section contained a link to a Tranche upload with the requested reports, it was only briefly mentioned and no description of how to read the reports was given in the manuscript. Thus, we have added Figure S-11 to supplemental materials that illustrates an example contig and describes how the contig was obtained from unidentified MS/MS spectra . The following text was also updated in the second-to-last paragraph of the Results section:*

“All SPS contigs, meta-contigs, input MS/MS spectra, identified spectra, and annotated de novo sequences associated with this paper may be downloaded from Tranche/ProteomeCommons.org at the following hash:

s+8iy5TbHHydsOPmTf9yqotRGvkxeJPF8BXJxMxxZOnCRXqbje8wbn+Orpxr51YR3L0S2sZBTYljUdHUF35LjfTqeukAAAAAAv6xWQ==. *This link also contains de novo sequencing reports that visualize how MS/MS spectra from each data set were used to generate de novo protein sequences. A subset of these reports detailing all 6-prot meta-contigs can also be found directly at <http://proteomics.ucsd.edu/Software/MetaSPS/6-prot_meta-contigs/index.html>. Figure S-11 in Supplemental Materials provides a description of how to interpret these reports in relation to algorithmic steps outlined Figure 1a.”*

**Reviewer comment #2**: “A second issue relates to how much of a challenge the authors gave the meta-SPS approach. The authors describe two scenarios, one a 6 protein mixture and one a single antibody. It is possible to imagine a biotech company purifying an antibody to complete homogeneity prior to analysis. However, it is more likely that someone will want to carry out an analysis of a more complex mixture with a few critical proteins of interest in them. Given the state of modern proteomics, this comes across as a relatively easy test. It would be beneficial if the authors actually mixed these two samples and analyzed them as one with the meta-SPS approach. A comparison of what is achieved would be valuable. Are they able to obtain the same for this 8 protein mixture as they do for the two mixtures separately? This would go a long way towards convincing the reader that this approach could be useful with much more challenging samples.”

*We agree with the reviewer that results from mixing the two samples would provide readers with a useful understanding of performance expectations from a more complex sample. Since we lack remaining sample to collect additional data, we have addressed this computationally by combining the data instead. Given that the aBTLA sample was digested with similar enzymes and analyzed with a similar instrument compared to the 6-prot sample, the principle difference between this experiment and that suggested by the reviewer is that combining the samples and analyzing with the same LC gradient would yield reduced MS/MS coverage of the target proteins (due to incomplete peptide sampling by the instrument). This would likely translate to lower sequencing coverage and length, but accuracy should be unaffected because we have clearly demonstrated that Meta-SPS can differentiate between spectra from unrelated peptide sequences in a larger mixture and still assemble de novo sequences with high accuracy In practice, one would simply extend the LC gradient time or collect the data on a faster scanning instrument in order to maintain adequate peptide sampling. Consequently, we added the following paragraph was added to the end of the Results section to address this concern:*

“Although the 6-prot sample contained a mixture of proteins, applications of de novo protein sequencing are often targeted towards specific proteins within a larger mixture. To test how Meta-SPS performance might be impacted by such samples, we combined the 6-prot CID MS/MS spectra with the high resolution CID MS/MS spectra from the aBTLA sample and executed the algorithmic steps outlined in Figure 1a on the combined set of MS/MS spectra. Here, the proteins of interest were the heavy and light chain of aBTLA antibody and the background mixture was represented by the 6-prot data. Since the high resolution CID spectra from the aBTLA and 6prot samples were acquired on a similar model of instrument (LTQ Orbitrap XL and LTQ Orbitrap, respectively), the low-resolution aBTLA spectra were excluded from this experiment to better simulate high resolution data acquisition of an aBTLA/6-prot mixture sample. Although this does not rigorously simulate the expected loss in MS/MS coverage one might expect from such a mixture (due to incomplete peptide sampling by the instrument), it is still a fair approximation of the algorithmic challenges associated with sequencing a small subset of proteins within the background of higher complexity. In practice, one would simply extend the LC gradient time or collect the data on a faster scanning instrument in order to maintain adequate peptide sampling. Compared to sequencing results on the aBTLA high resolution spectra, Meta-SPS produced the same sequencing accuracy (98.1% compared to 98.6%) and average length (18 AA compared to 17 AA) of the aBTLA antibody from the combined aBTLA/6-prot set of MS/MS spectra at the cost of reduced sequencing coverage (58% compared to 71%) and shorter maximum sequence length (35 AA compared to 45 AA). Compared to SPS, Meta-SPS generated de novo sequences 100% longer on average from the combined set with ~2x as many correct sequence calls per incorrect sequence call. We note that in a real MS experiment mixing the 6-prot and aBTLA samples, the absence of a faster spectral acquisition rate and/or extended peptide separation time could diminish protein sequence coverage by MS/MS spectra and thus further limit the overall sequencing length and coverage.”

*We also mention in the discussion section that Meta-SPS provides a substantial boost in performance over SPS for this more complex mixture:*

“Nonetheless, even in the background of the 6prot MS/MS spectra, Meta-SPS still improved upon SPS sequencing accuracy (from 97% to 98%), average sequence length (from 11 AA to 20 AA), and maximum sequence length (from 25 AA to 35 AA) for the aBTLA antibody. Analyzing more complex mixtures with greater effectiveness may require faster spectral acquisition rates or extended peptide separations to generate enough spectra to cover all proteins in a sample.”

*Towards the end of the discussion section we describe foreseeable improvements that could also possibly improve the performance of Meta-SPS on complex mixtures.*