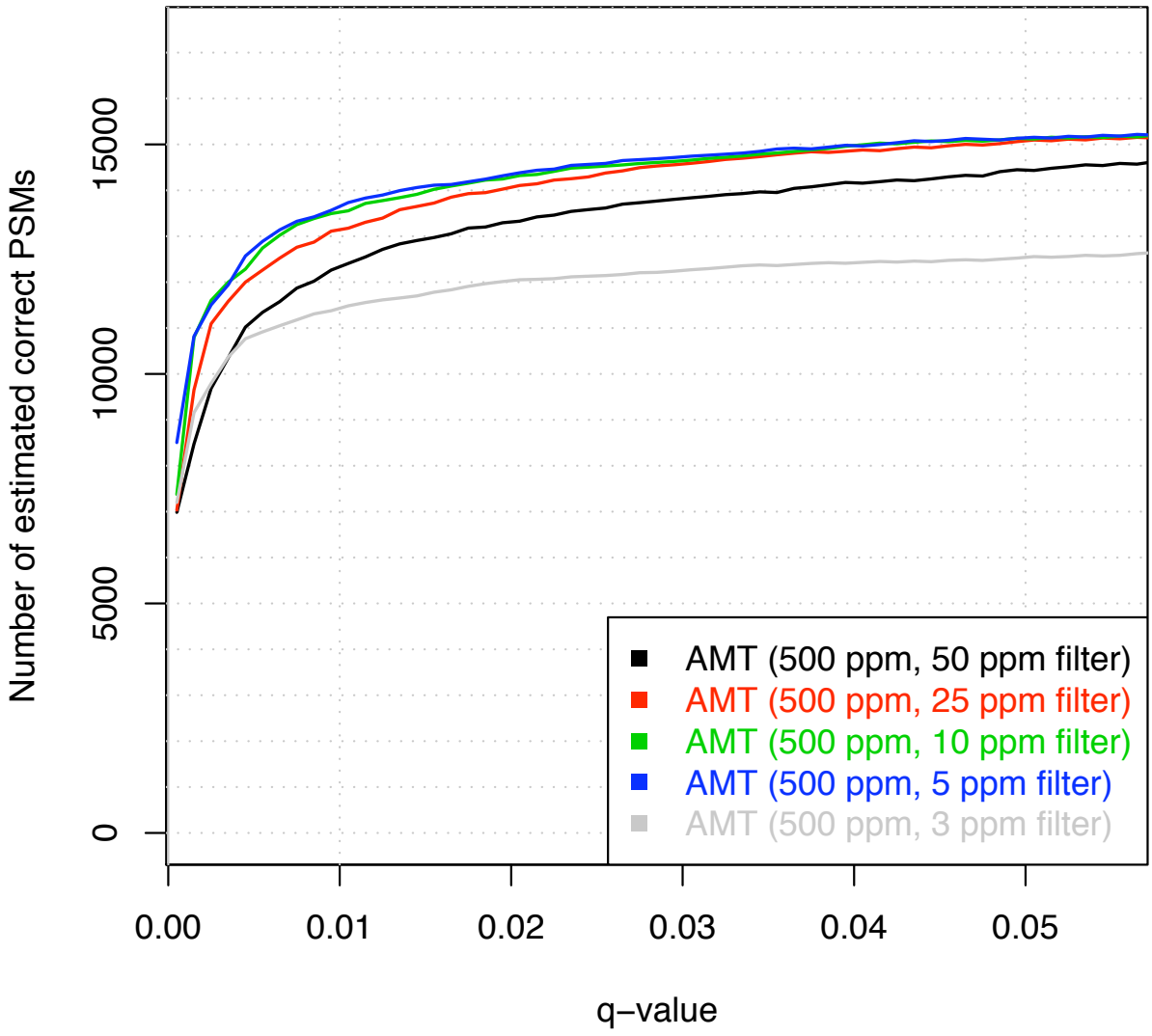
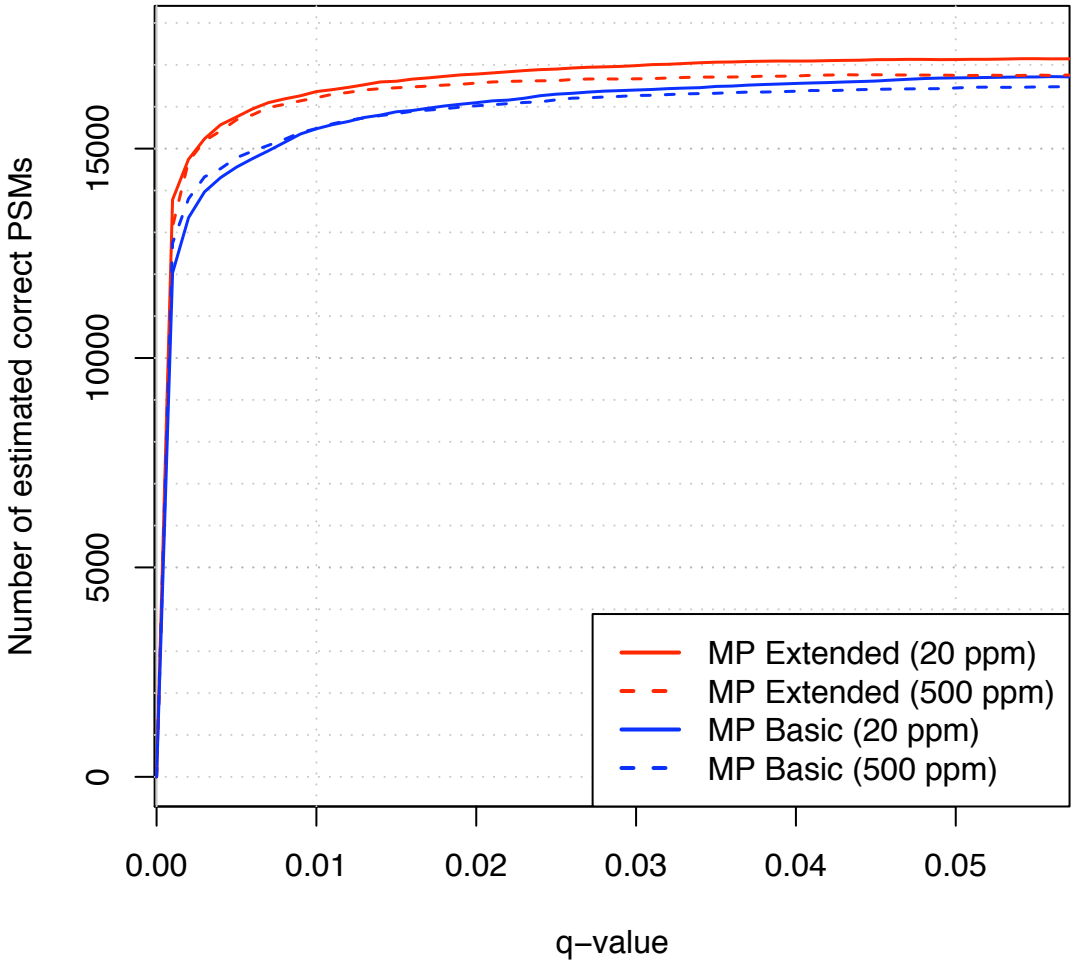
Supplementary figure 1



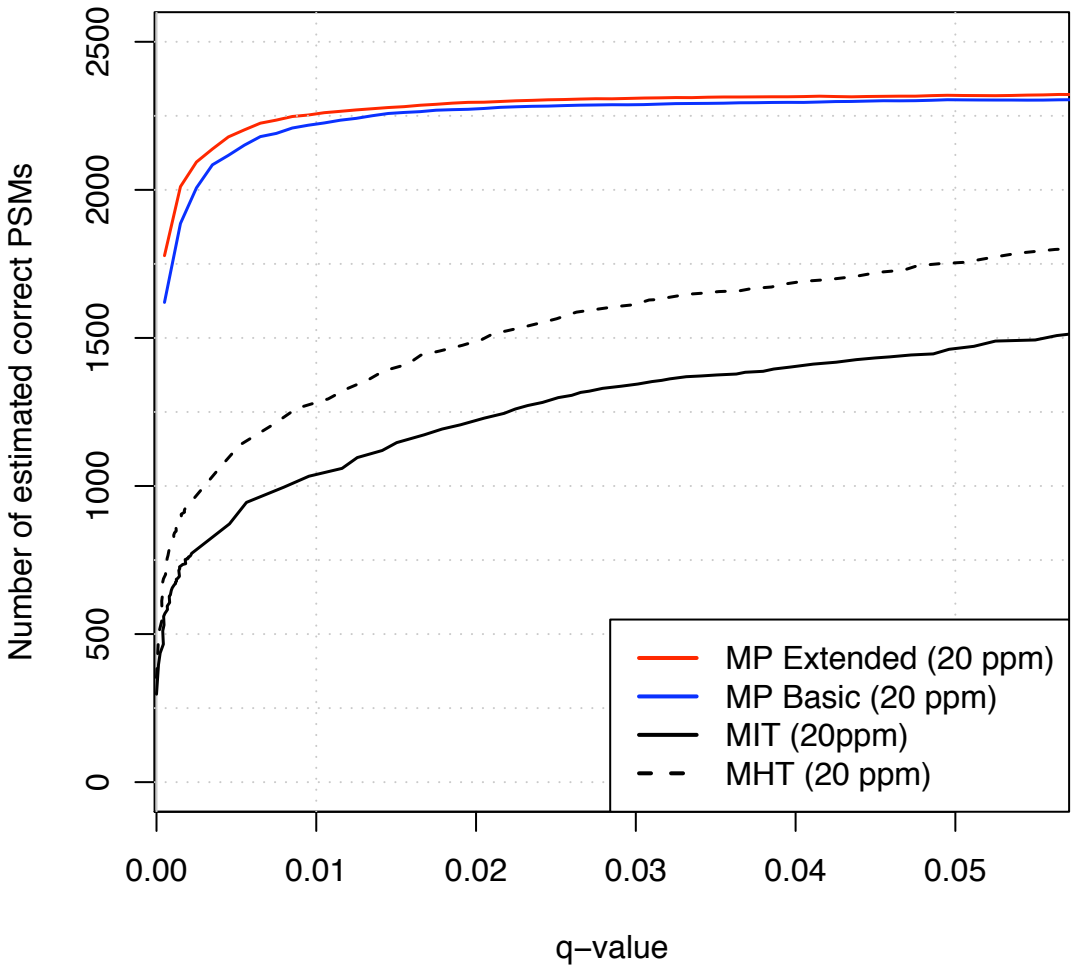
The performance of the Adjusted Mascot Threshold (AMT) was evaluated using mass deviation filter settings of 50, 25, 10, 5 and 3ppm: for each, the number of estimated correct PSMs was determined across a range of q-values. These results show the trade-off between improving specificity with more stringent mass tolerance filters and conversely excluding potentially correct PSMs when the filters become too stringent. For this dataset the best mass filter was found to be 5 ppm.

Supplementary figure 2



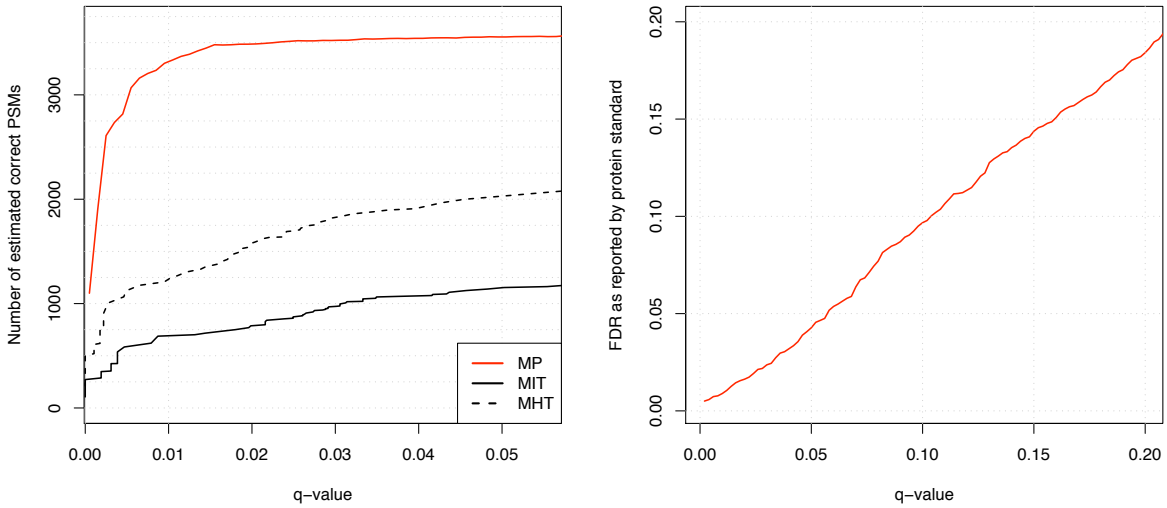
The Mascot Percolator performance was evaluated for the relaxed (500 ppm) and stringent (20 ppm) Mascot search over a range of q-values (0.00 - 0.06) using the basic and extended feature set. The close tracking between the 500 ppm and 20 ppm results indicate that the peptide mass deviation features do not further boost performance for the more relaxed mass tolerance setting.

Supplementary figure 3



The Mascot Percolator, MIT and MHT performance was evaluated for the protein standard dataset using the basic and extended feature set over a range of q-values (0.00 - 0.06).

Supplementary figure 4



The performance of Mascot Percolator, MIT and MHT was evaluated for the protein standard dataset that was searched in Mascot without any enzyme constraints. Mascot Percolator identified 260% and 470% more peptides than using the MHT or MHT, respectively, at a q-value of 1% (figure left). The estimated q-values were also compared against the expected sequences from the protein standard dataset to validate of these results (figure right).

Overall this demonstrates that Mascot Percolator can also be applied to more challenging conditions than standard tryptic searches, where the search space increases by several orders of magnitude (http://www.matrixscience.com/help/search_field_help.html), such as searches without any enzyme constrains or excessive PTM settings.

Supplementary information 1

Feature information: (feature 3) the native Mascot score correlates well with the quality of a PSM; (feature 4) the difference between the Mascot score and next best non-isobaric peptide hit indicates the level of ambiguity between two competing matches; (feature 5-7) peptide mass accuracies can serve as a discriminator; (feature 8) multiple distinct peptide sequences matching to the same protein are more likely to be correct (alternatively modified peptides are collapsed and not discriminated). Also note, that this feature is associated with some controversy since it introduces protein level evidence and does not not fully compensate for protein size differences, e.g. the feature is limited to two distinct sequences. We have not detected any unexpected behaviour in our subsequent evaluations and advocate the use of protein level information where maximal identification sensitivity is required.

The main idea of the extended feature set is to include fragment ion matching statistics: (feature 10) higher total ion intensity can result in better spectrum quality; (feature 11) the sum of total matching ions, (feature 12) the fraction of matched ion intensity over the total intensity and (feature 16-18) the fraction of matching ions, the derived sequence coverage and matched ion intensity of a given ion series can correlate with the PSM quality; (feature 13) alternative PSM score; (feature 14-15) fragment mass error often contributes to discrimination of an correct and incorrect PSM.

Feature 1-2 and 9 are not expected to provide discrimination power by themselves, but they may correlate with other features and thereby improve discrimination.