

Using the entrapment sequences method as a standard to evaluate key steps of proteomics data analysis process

Additional file 1

Figure S1. Plot figures of the numbers of PSM, peptide and protein identified by five search engines under the estimated FDRs on *Pfu* dataset (A-C) and *LM3* dataset (D-F). The original scores of all five search engines are used.

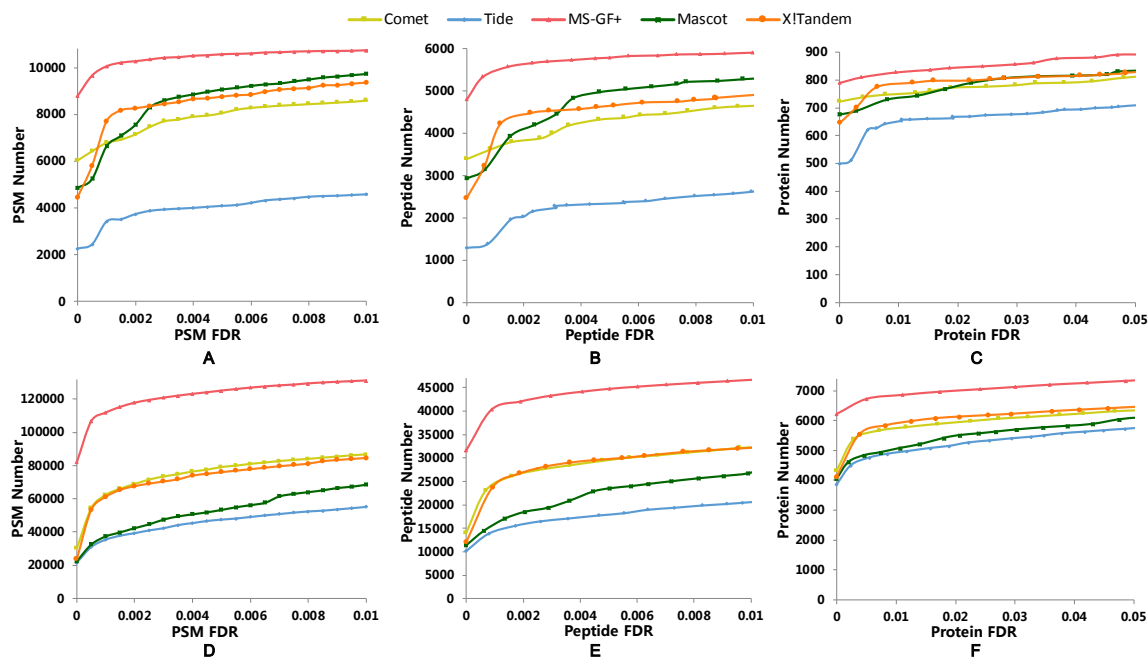


Figure S2. Distribution figures of the identification numbers and FMRs under 0.01 FDR of spectrum, peptide and protein level for five search engines on *Pfu* dataset (A-C) and *LM3* dataset (D-F). The original scores of all five search engines are used.

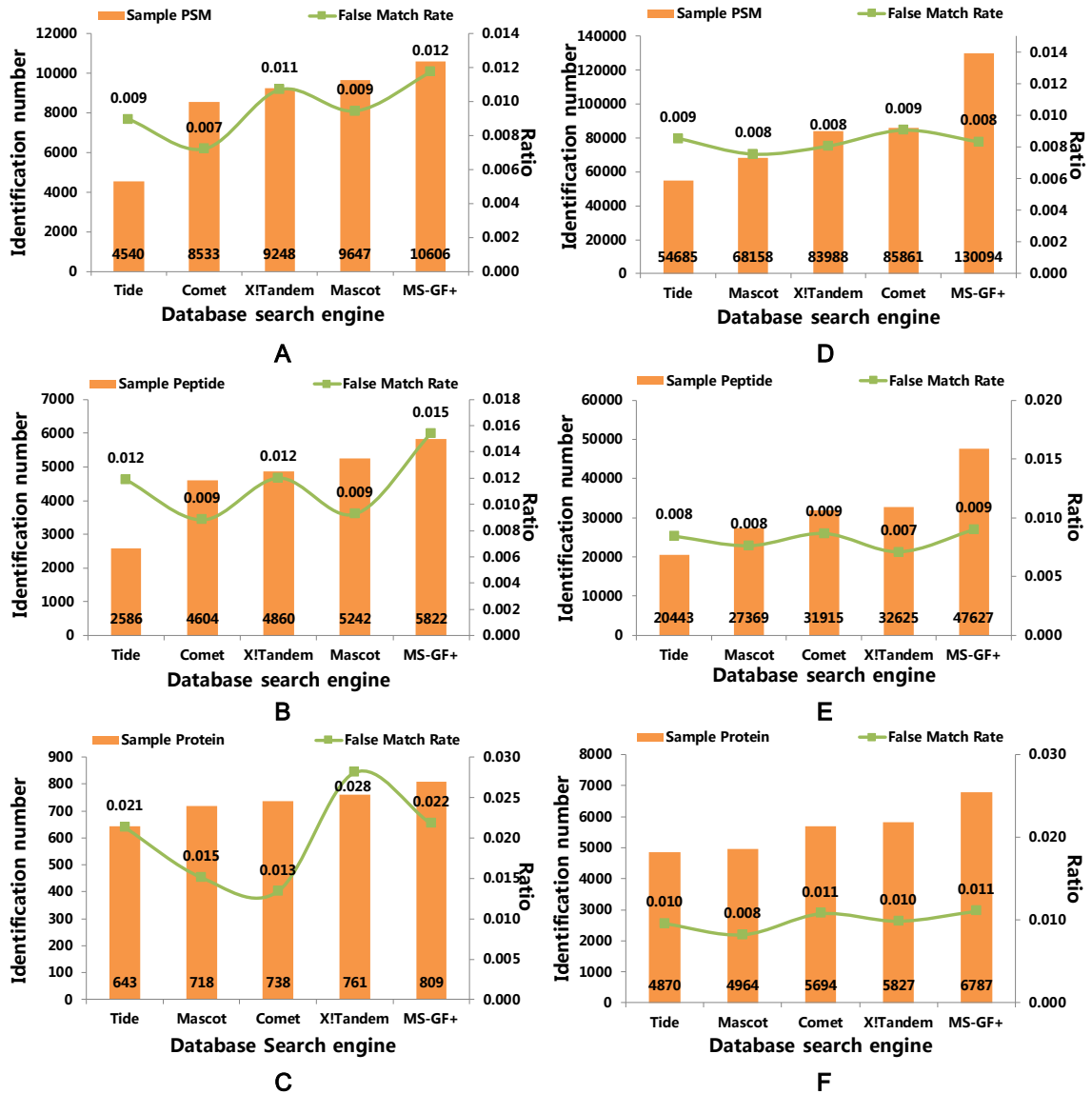


Figure S3. Overlap of PSMs identified by five search engines' reprocessed scores (A, B) and four quality control methods (C, D) on *Pfu* dataset (A, C) and *LM3* dataset (B, D) under 0.01 PSM FDR. The venn diagram was drew using the web tool on <http://bioinformatics.psb.ugent.be/webtools/Venn/>.

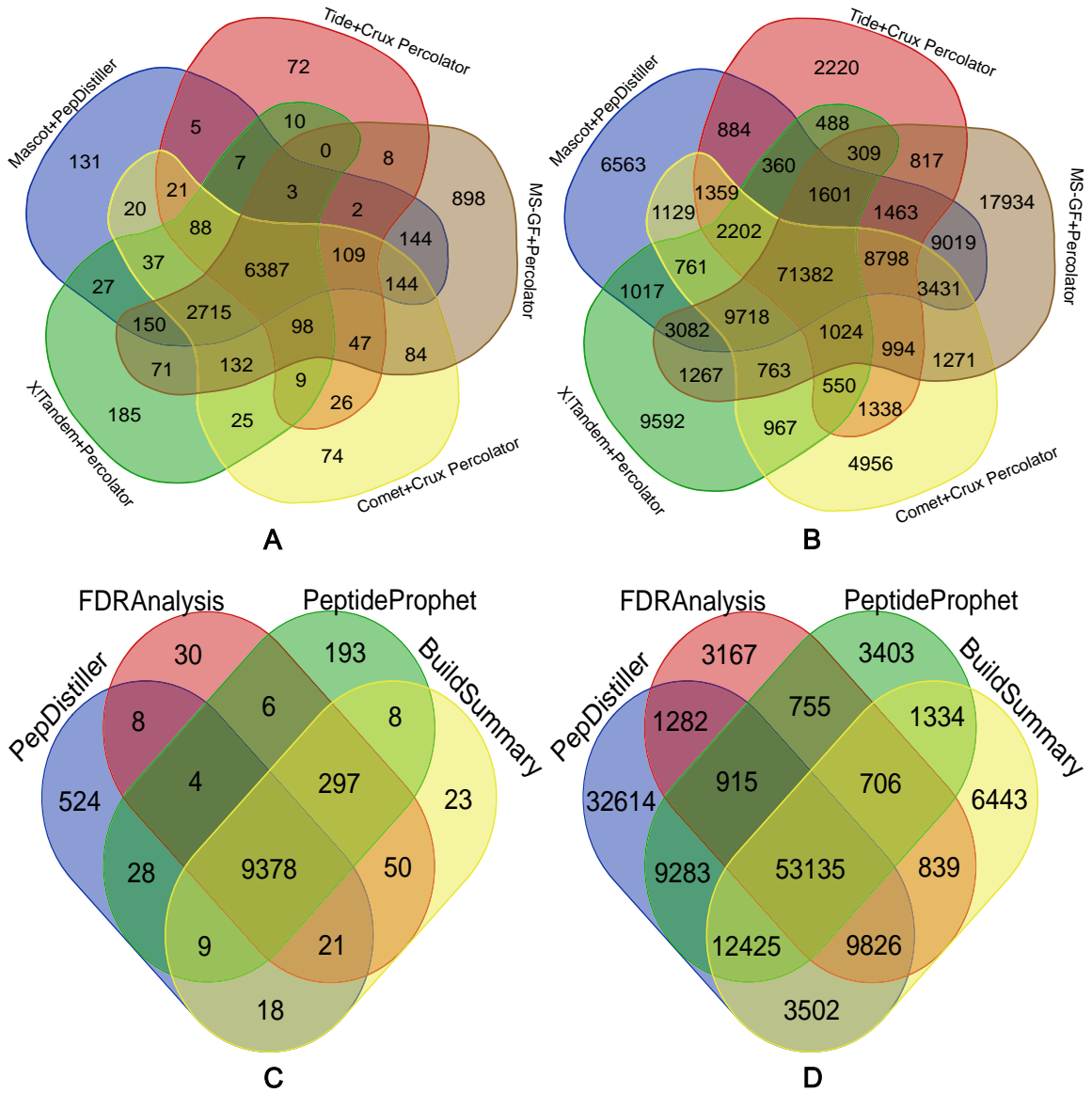


Figure S4. Distribution figures of the PSMs, FDRs and FMRs identified by different number of quality control methods on *LM3* dataset. (A) Distributions of original FDRs and FMRs under 0.01 PSM FDR. (B) Distributions of refined FDRs and FMRs under 0.01 PSM FDR, as the PSMs in each subgroup are further filtered to keep their sub-FDRs lower than 0.01 PSM FDR.

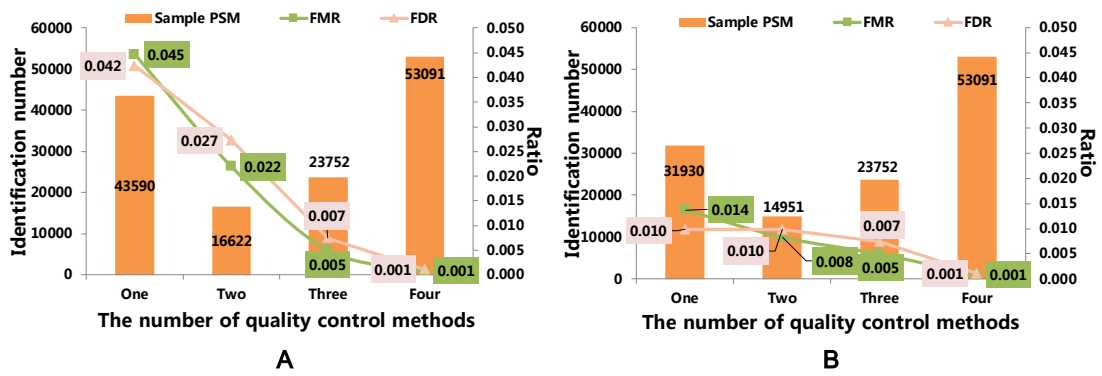
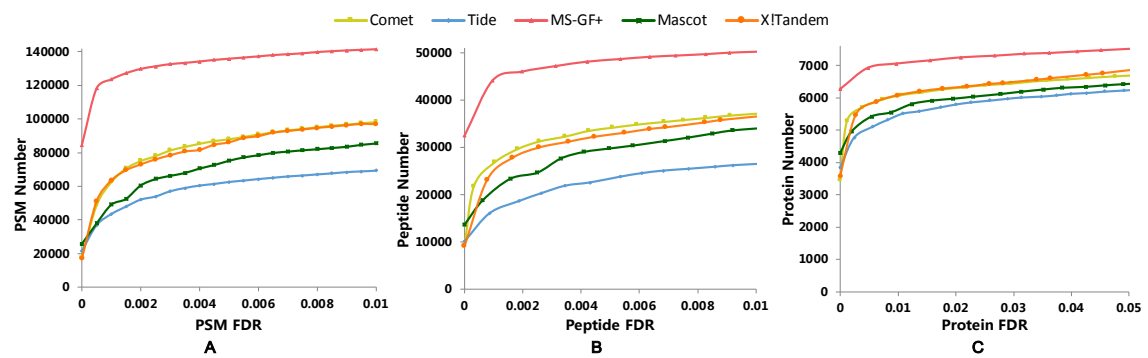
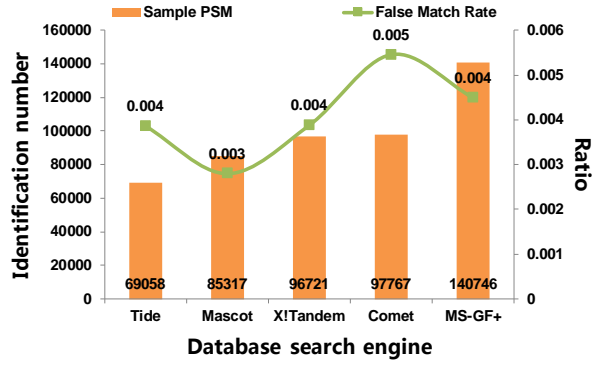
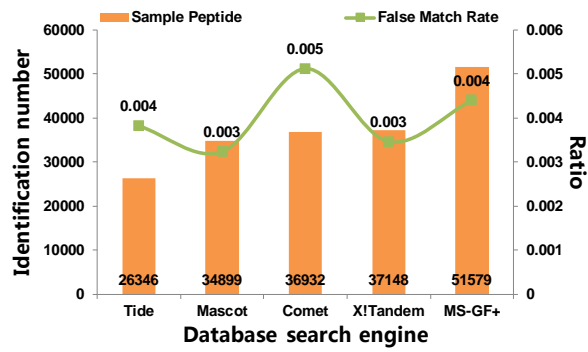


Figure S5. Using a small size of entrapment sequences to evaluate the search engines' original scores on LM3 dataset. The original scores of all five search engines are used. (A-C) Plot figures of the numbers of PSM, peptide and protein identified by five search engines under the estimated FDRs. (D-F) Distribution figures of identification numbers and FMRs for five database search engines under 0.01 PSM FDR, peptide FDR and protein FDR.

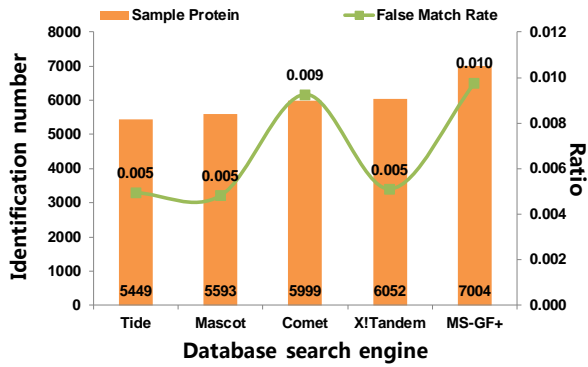




D

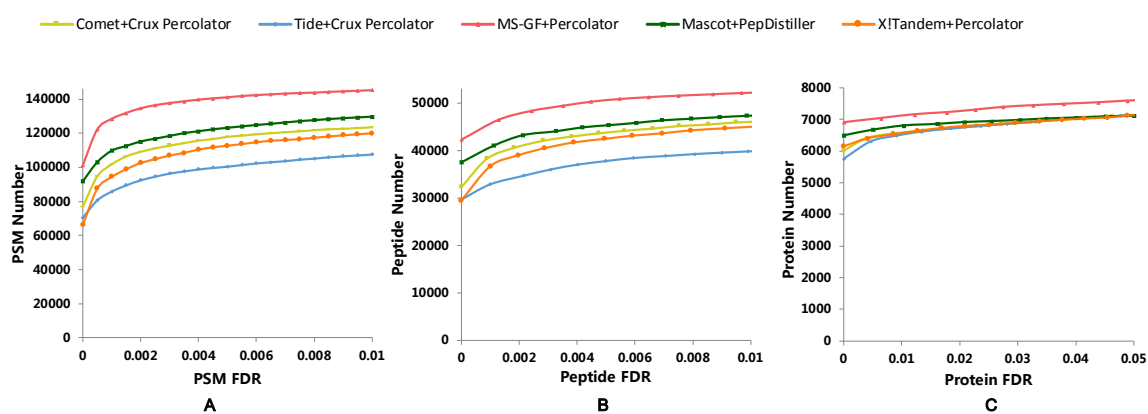


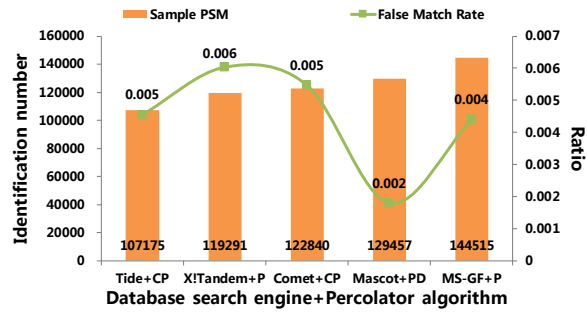
E



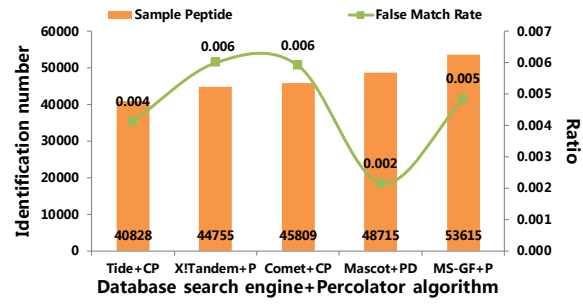
F

Figure S6. Using a small size of entrapment sequences to evaluate the search engines' reprocessed scores on *LM3* dataset. (A-C) Plot figures of the number of spectrum, peptide and protein identified by five search engines under the estimated FDRs. (D-F) Distribution figures of the identification numbers and FMRs for five search engines under 0.01 FDR of spectrum, peptide and protein level. The reprocessed scores of all five search engines are used. And CP, P and PD in figure stand for Crux Percolator, Percolator and PepDistiller respectively.

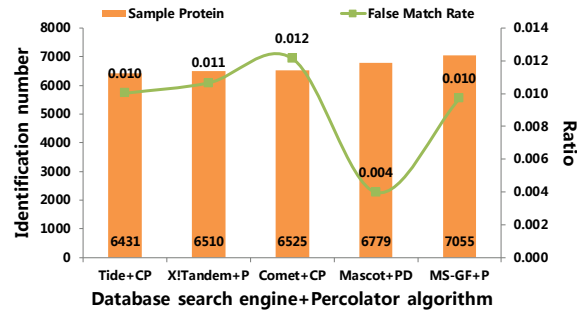




D



E



F