**Supplementary Materials for:**
**Scribe: next generation library searching for DDA experiments**

Brian C. Searle[1, 2, 3, 4 *] , Ariana E. Shannon[1, 2, 3], and Damien Beau Wilburn[1, 2, 3]

[1]Department of Biomedical Informatics, The Ohio State University Medical Center, Columbus, Ohio 43210, United States.
[2]Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210, United States.
[3]Pelotonia Institute for Immuno-Oncology, The Ohio State University Comprehensive Cancer Center, Columbus, Ohio 43210, United States.
[4]Proteome Software Inc., Portland, OR

*Corresponding Address: Brian C. Searle, Ph.D., brian.searle@osumc.edu

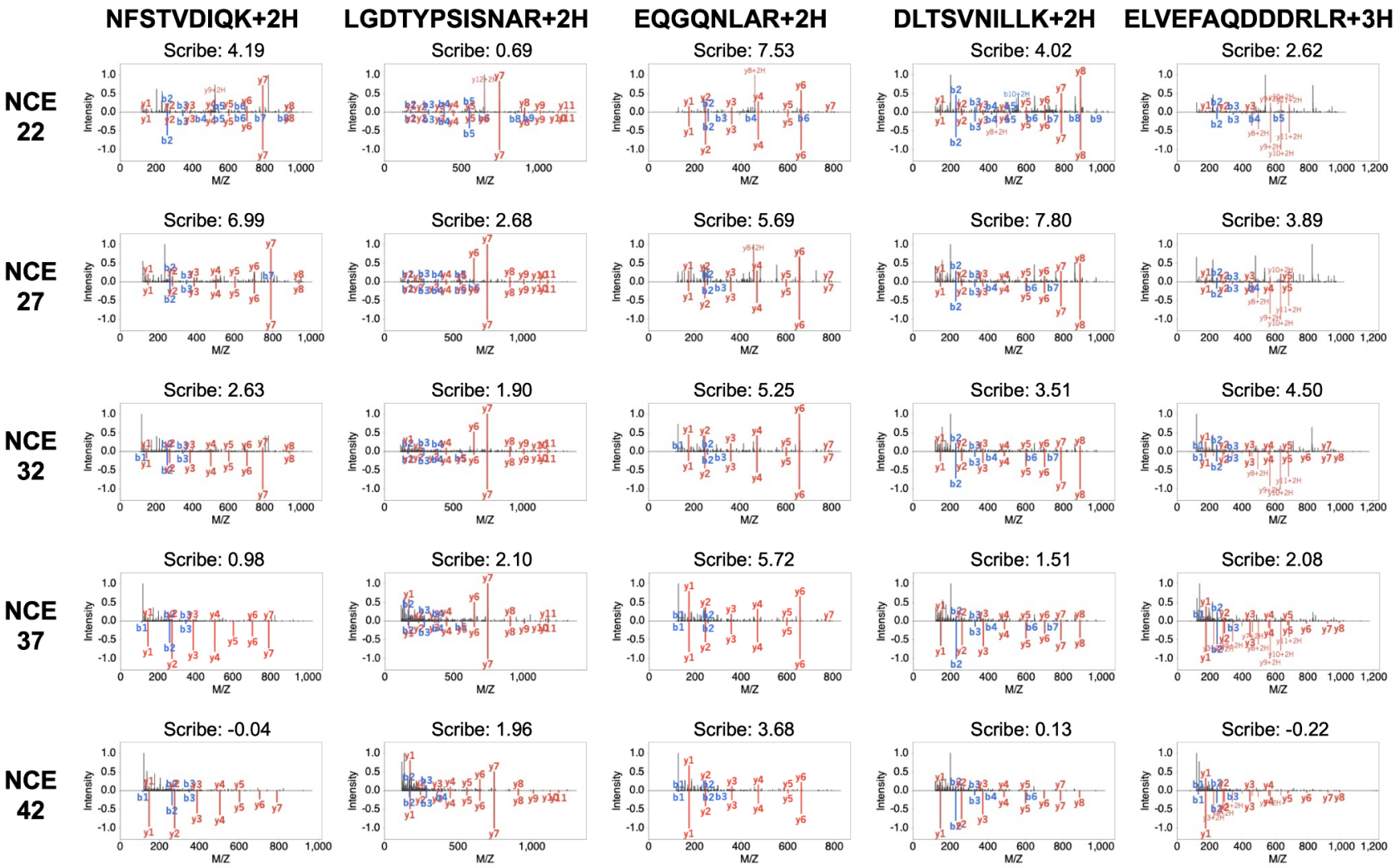
NFSTVDIQK+2H
LGDTYPSISNAR+2H
EQGQNLAR+2H
DLTSVNILLK+2H
ELVEFAQDDDRLR+3H
NCE 22
NCE 27
NCE 32
NCE 37
NCE 42
Scribe: 4.19
Scribe: 0.69
Scribe: 7.53
Scribe: 4.02
Scribe: 2.62
Scribe: 6.99
Scribe: 2.68
Scribe: 5.69
Scribe: 7.80
Scribe: 3.89
Scribe: 2.63
Scribe: 1.90
Scribe: 5.25
Scribe: 3.51
Scribe: 4.50
Scribe: 0.98
Scribe: 2.10
Scribe: 5.72
Scribe: 1.51
Scribe: 2.08
Scribe: -0.04
Scribe: 1.96
Scribe: 3.68
Scribe: 0.13
Scribe: -0.22
Intensity
M/Z

**Figure S1: Fragmentation examples of peptides with Scribe scores.** Butterfly fragmentation plots show acquired spectra on top and predicted spectra on bottom for five peptides at five different NCE settings. Each peptide was detected at a 1% peptide-level FDR in every NCE acquisition, and each form of that peptide was found within tight retention time windows of each other. This additional level of stringency ensures that the peptide detections are high quality matches even though some have very low scores.

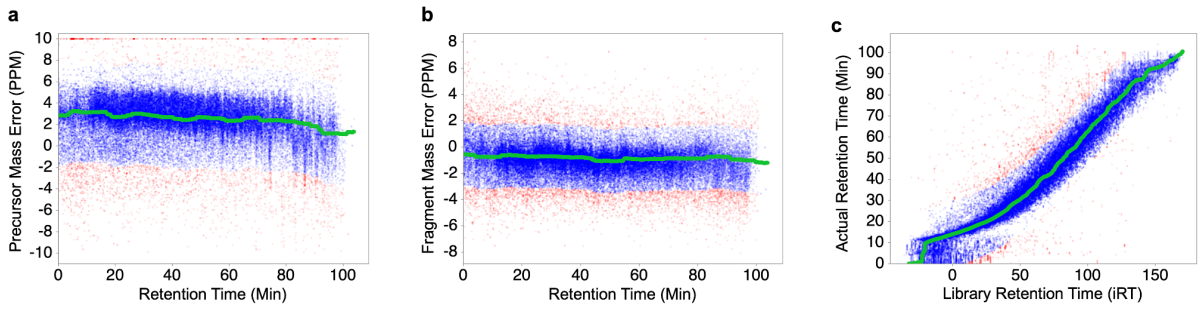


**Figure S2: Model refinement after first pass processing with Percolator.** Global delta mass errors at the precursor (**e**) and fragment (**f**) level, as well as delta retention time (**g**) are calculated and fit using the KDE method (green line). For these error scores, the middle 90% of the target distribution are labeled as blue dots, while potential PSM outliers are labeled as red dots.

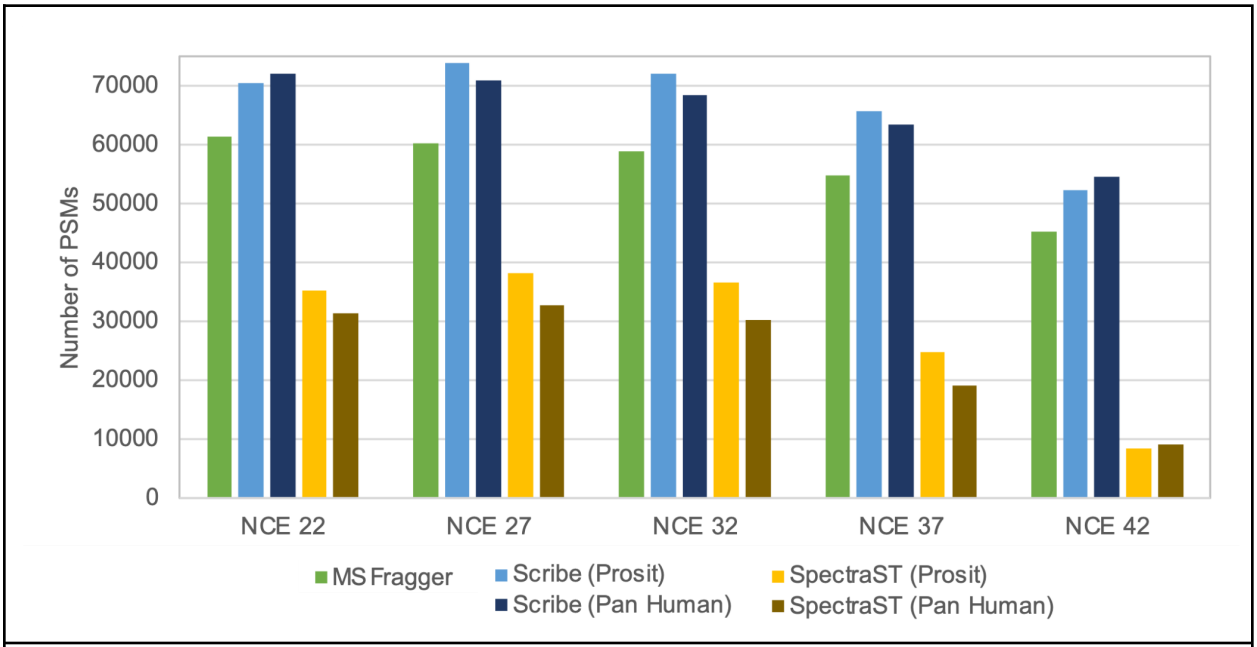


**Figure S3: Performance with tryptic PSMs.** The number of peptide-spectrum matches (PSMs) detected by MSFragger, Scribe, or SpectraST in HeLa experiments acquired on an Exporis 480 at different NCE settings. Scribe and SpectraST were provided with either the Pan Human Library or Prosit predicted spectrum libraries tuned specifically for the acquired NCE setting on this instrument.

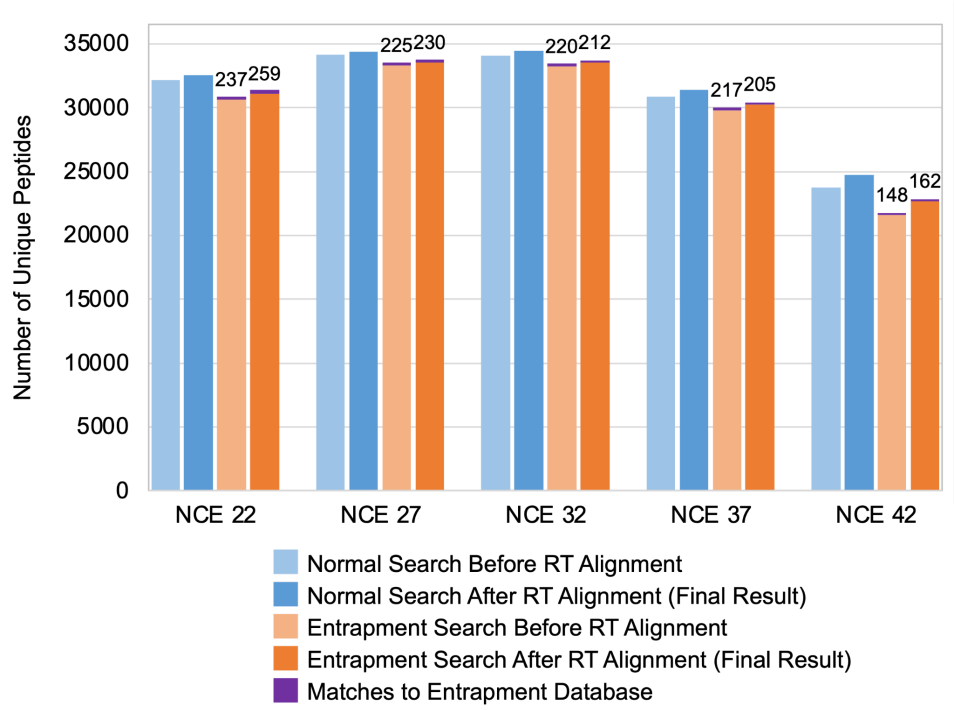


**Figure S4: Scribe performance with entrapment searches.** The number of unique peptides detected by Scribe in the first round (before retention time fitting, light color) or final round (after retention time fitting, dark color) using either a normal (blue) or a shuffled entrapment (orange) database search. The shuffled entrapment database doubles the size of the normal target/decoy search, and consequently Scribe detects slightly fewer peptides. Matches to the entrapment database are colored purple and the number of entrapment peptide hits is labeled above the bars. In all cases, the number of entrapment peptides is less than 1% of the total assigned peptides, matching the 1% peptide-level FDR.

**Table S1:**

| # | Score Name | Score Description |
|---|---|---|
| 1 | averageParentDeltaMass | Average mass error of matched precursors (including isotopes) in PPM |
| 2 | averageFragmentDeltaMasses | Average mass error of matched fragments in PPM |
| 3 | DotProduct | $\sum_{i=1}^{n} \left(A_i * L_i\right)$ |
| 4 | contrastAngle | $1 - (2 * acos(DotProduct))/\pi$ |
| 5 | logit | $log(DotProduct)/(1 - DotProduct)$ |
| 6 | primary | $log(DotProduct) * !(numberOfMatchingPeaks)$ |
| 7 | xCorrLib | XCorr comparing against Sequest normalized library spectrum |
| 8 | xCorrModel | XCorr comparing against Sequest model spectrum |
| 9 | scribeScore | $-ln\left(\sum_{i=1}^{n} \left(A_i - L_i\right)^2\right)$ |
| 10 | numberOfMatchingPeaks | Total number of matching peaks between library and acquired spectra |
| 11 | isotopeDotProduct | Normalized DotProduct of acquired and predicted precursor isotope intensities |
| 12 | percentBlankOverMono | Precursor isotope-1 intensity divided by the monoisotopic intensity |
| 13 | numberPrecursorMatch | Number of precursor isotopes with >0 intensity (up to 3) |
| 14 | logSp | $Log_{10}$ of the Sp score from Sequest |
| 15 | maxLadderLength | Maximum connected fragment ions in a ladder |
| 16 | eValue | E-value of scribeScore as calculated by the Fenyo et al 2003 method |
| 17 | deltaScore | (bestScribeScore - secondBestScribeScore) / (bestScribeScore + 10) |
| 18 | numConsidered | Total number of candidate library entries considered |
| 19 | chargeMatch | 1 if acquired charge matches library entry, 0 if mismatch |