MacroSEQUEST: Efficient candidate-centric searching and high resolution correlation analysis for large-scale proteomics datasets

Brendan K. Faherty[1] and Scott A. Gerber[1,2,*]

Supplementary Information

1. Department of Genetics, Dartmouth Medical School, Lebanon, NH 03756
2. Norris Cotton Cancer Center, Lebanon, NH 03756

* to whom correspondence should be addressed: scott.a.gerber@dartmouth.edu

a

```
./Macro8m -h
USAGE Macro <-i exp list file> <-d database>
ARGUMENTS
     -i  FILE    exp list file is the path to the list of
         experimental spectra file (a list of dtas, one per line)
     -d  FILE    database is a Fasta file to look through
     -s  STRING OPTIONAL static modifictions string consisting of
                residues and masses, ie: -s C:57.021461
     -m  STRING OPTIONAL dynamic modifications string consisting of
                residues and masses, ie: -m M:15.9949146221,STY:
                79.9663304084
     -e  STRING enzyme, currently 'trypsin-full' (DEFAULT),
                'trypsin-partial' and 'none'
     -c  INT    number missed cleavage events per peptide, default 3
     -t  FLOAT  searching match tolerance in Daltons, default 1.1Da
     -p  FLOAT  searching match tolerance in PPM.
                * USE OVERRIDES -t and used 0.1Da. bin size
     -b  FLOAT  binning MS2 peak tolerance in Daltons, default
                1.0Da. Used in scoring
     -o  FILE   file name to use for outputting csv file.
                Default is Macro-date-time.csv
     -l  INT    print level, default 0 - quiet. 1:info,
                2:progress, 3: debug, 4: verbose debug
     -h         print out a usage message
     -v         print out versioning information
```

b

```
HeLa90/HeLa90.12459.12459.2.dta
HeLa90/HeLa90.1178.1178.3.dta
HeLa90/HeLa90.0884.0884.3.dta
HeLa90/HeLa90.14463.14463.3.dta
HeLa90/HeLa90.1686.1686.2.dta
HeLa90/HeLa90.10834.10834.2.dta
HeLa90/HeLa90.14363.14363.2.dta
HeLa90/HeLa90.13850.13850.2.dta
HeLa90/HeLa90.10405.10405.2.dta
HeLa90/HeLa90.4545.4545.3.dta
```

Figure S-1: Usage and command line arguments for Macro. A) “Help” information regarding usage for the current version of Macro. **B)** Formatting of the input file list.

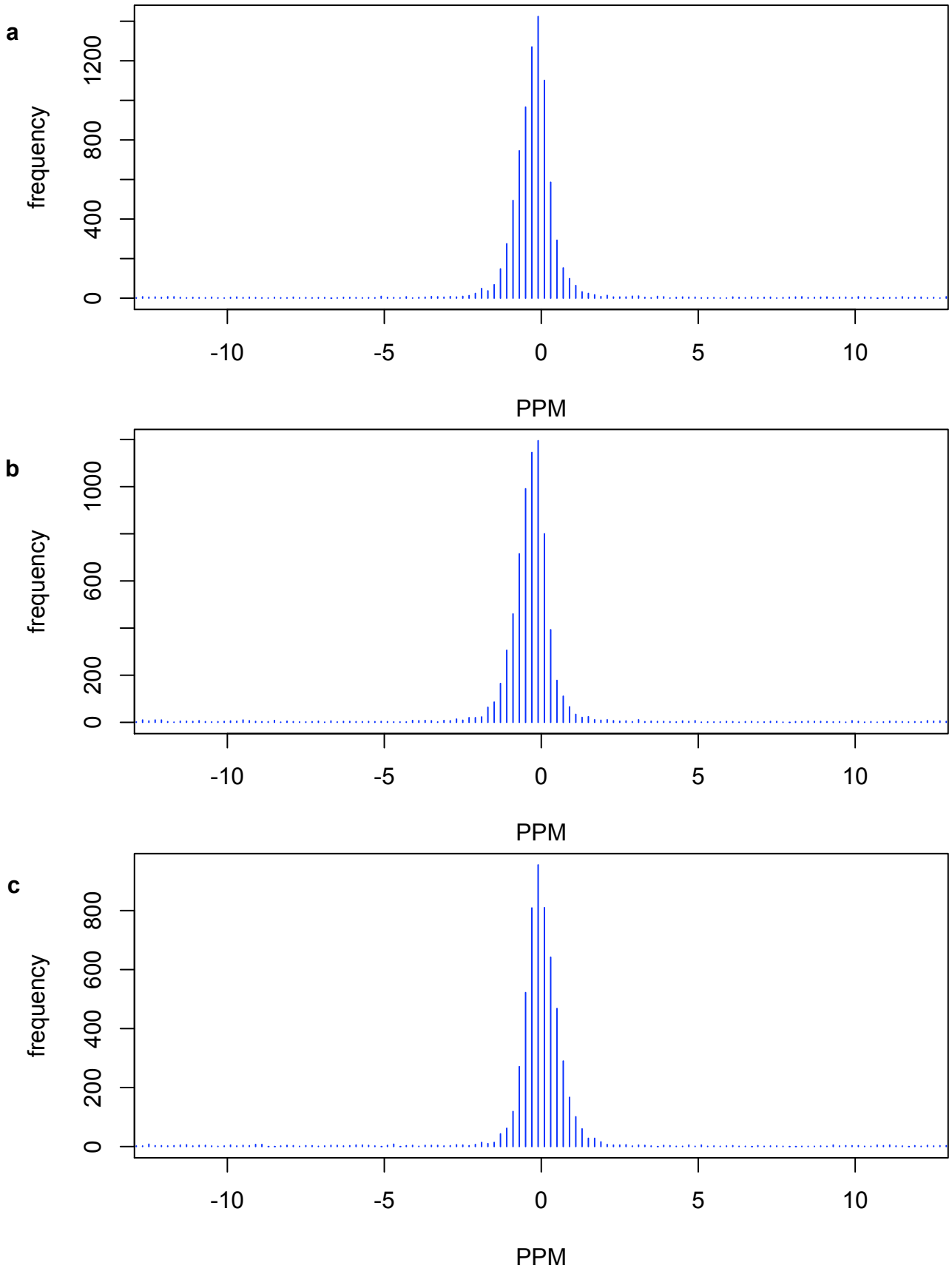

Figure S-2: Precursor ion mass measurement accuracy distributions. Precursor ion mass measurement accuracy distributions for the **A)** yeast, **B)** human, and **C)** human phosphopeptide datasets used in this work.

a

	Yeast 90	HeLa 90	HeLa Phos 90
number spectra	12,979	13,129	10,761
number tryptic peptides	3,042,150	27,492,945	331,035,378
number matched peptides	1,769,687	16,851,426	321,399,594
number scoring operations	29,053,497	277,789,229	1,430,087,343
number modifications	461,154	4,700,878	307,323,410

b

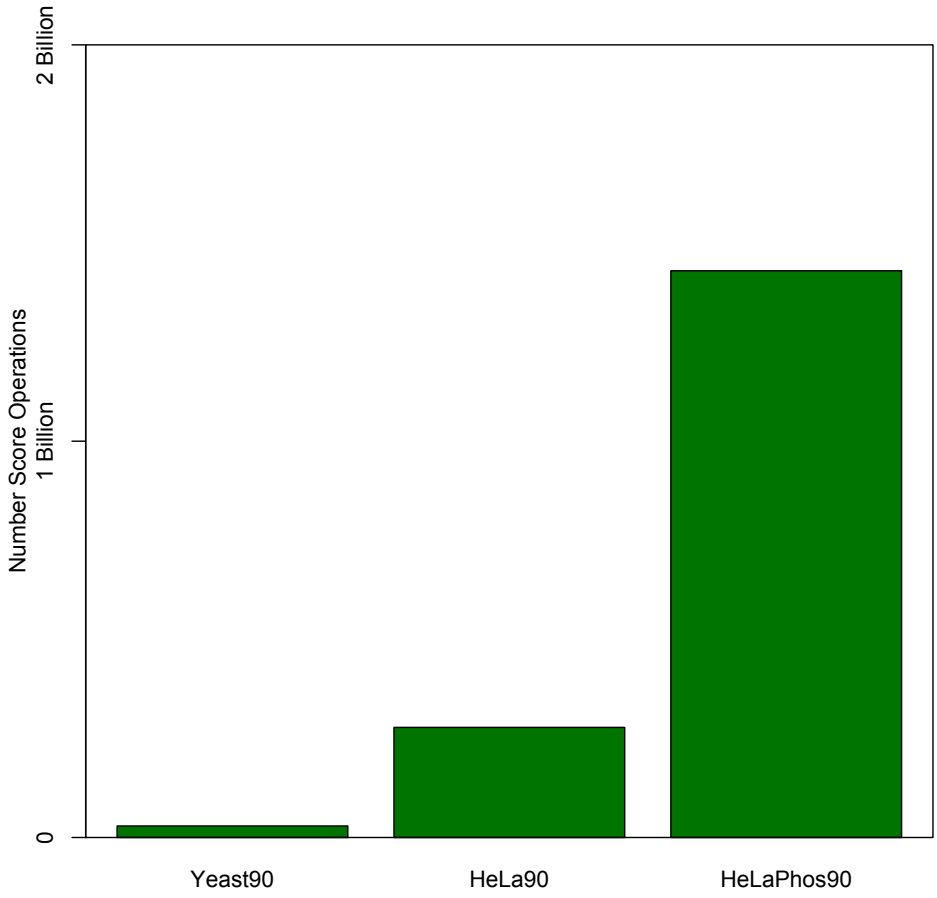

Figure S-3: Scoring summary statistics. Relevant scoring statistics for the datasets used in this work.

a

MS/MS bin width	cutoffs deltaCorr	2+	3+	4+	TP Number	FP Rate
1.000 Da.	0.08	0.7142	0.8786	n/a	3408	0.38%
0.750 Da.	0.08	0.7142	0.8786	n/a	3408	0.38%
0.500 Da.	0.08	0.5922	0.7759	0.9996	3690	0.46%
0.250 Da.	0.08	0.4324	0.6255	0.9109	3755	0.45%
0.100 Da.	0.08	1.4979	0.4738	0.6743	3889	0.75%
0.075 Da.	0.08	1.3506	0.4285	0.6910	4009	0.82%
0.050 Da.	0.10	1.3736	0.4682	0.5915	4029	0.77%
0.025 Da.	0.1	0.9593	0.3594	0.5058	4244	0.97%
0.010 Da.	0.10	1.2926	1.0167	1.4701	3968	0.93%

b

Search	cutoffs deltaCorr	2+	3+	4+	TP Number	FP Rate
1.1 Da. / 1.0 Da.	0.08	0.7142	0.8786	n/a	3408	0.38%
10 PPM / 1.0 Da.	0.08	1.4430	1.5871	1.7966	3596	0.95%
1.1 Da / 0.025 Da.	0.10	0.9593	0.3594	0.5058	4244	0.97%
10 PPM / 0.025 Da.	0.10	1.1230	1.1600	1.2632	4109	0.92%

Table S-1: True positive determination statistics.