Supplemental Information

EndoGenius: Optimized neuropeptide identification from mass spectrometry datasets

Lauren Fields,[1] Nhu Vu,[1] Tina Dang,[3] Hsu-Ching Yen,[2] Min Ma,[3] Wenxin Wu,[1] Lingjun Li[1,3,4,5]*

[1]Department of Chemistry, University of Wisconsin-Madison, 1101 University Avenue, Madison, WI 53706, USA

[2]Department of Biochemistry, University of Wisconsin-Madison, 433 Babcock Drive, Madison, WI 53706, USA

[3]School of Pharmacy, University of Wisconsin-Madison, 777 Highland Avenue, Madison, WI 53705, USA

[4]Lachman Institute for Pharmaceutical Development, School of Pharmacy, University of Wisconsin-Madison, Madison, WI 53705, USA

[5]Wisconsin Center for NanoBioSystems, School of Pharmacy, University of Wisconsin-Madison, Madison, WI 53705, USA

*Corresponding author

Tel.: +1 (608) 265-8491

Fax: +1 (608) 262-5345

E-mail: lingun.li@wisc.edu

Table of Contents

Supplemental Figures (located within this document)

- **Figure S1:** PSM assignment decision tree
- **Figure S2:** Evaluation of motif database perturbations
- **Figure S3:** Target-decoy entrapment search results
- **Figure S4:** DSD results
- **Figure S5:** FDR versus EndoGenius score evaluation

Supplemental files

- **Supplemental File 1:** Definitive screening design
- **Supplemental File 2:** Full-factorial design
- **Supplemental File 3:** Crustacean neuropeptide motif database
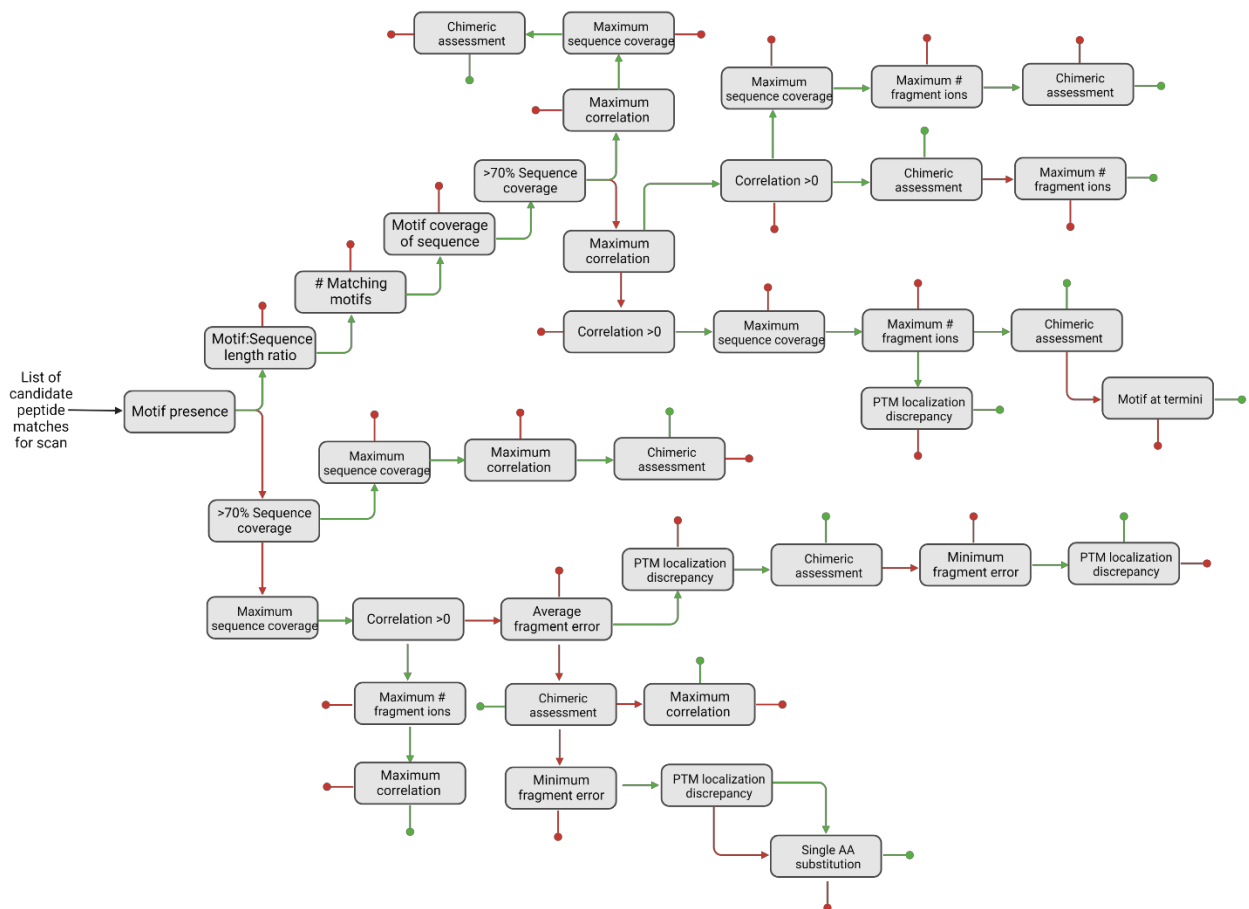- **Supplemental File 4:** EndoGenius user manual

Figure S1: A specialized peptide-spectrum match (PSM) filtering decision-tree was developed to reward neuropeptide-typic attributes of a given PSM. At each step in the decision tree, if only one peptide match was remaining for a given spectrum, that peptide was assigned as the PSM. Otherwise, the next round of criteria was assessed. In this tree, green color denotes a positive response, wherein a candidate PSM satisfied the criterion, whereas a red color denotes a negative response, where the criterion was not satisfied. Nodes represent where no additional criteria were presented, and a peptide was selected at random from the remaining candidate PSMs, in line with more traditional PSM assignment approaches.
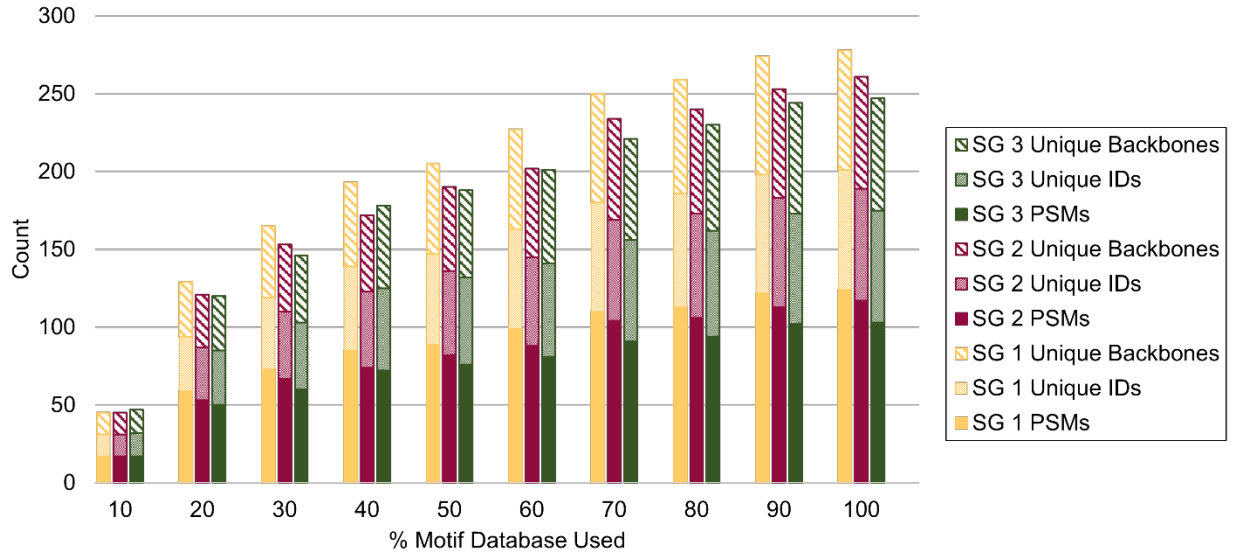
Figure S2: Evaluation of number of PSMs, unique IDs, and unique backbones across three technical replicates when a fraction of the motif database, selected at random, was included in the analysis. Unique IDs refers to a neuropeptide including any post-translational modifications (PTMs), while unique backbones refer to the amino acid sequence of the neuropeptide only.
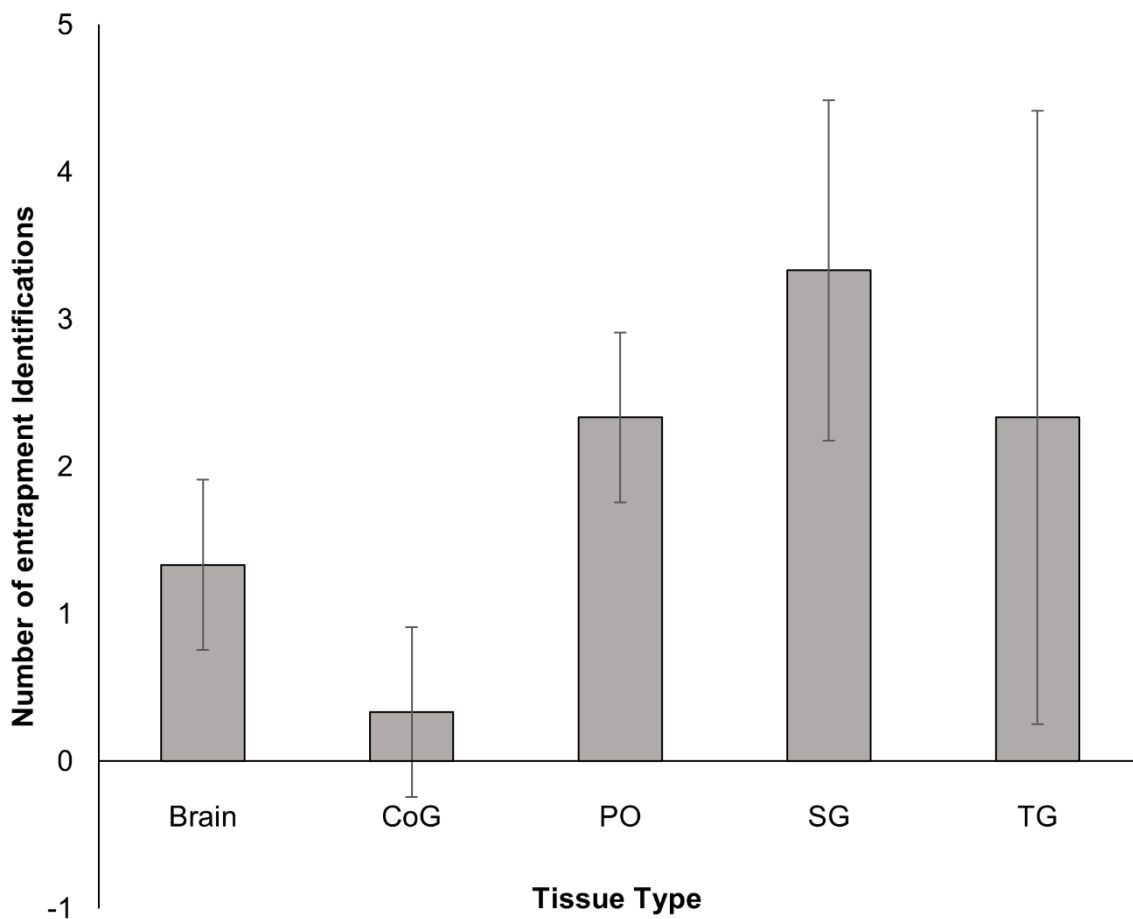
Figure S3: Average number of unique backbones yielded by EndoGenius searches of five tissue types across three technical replicates when searching against an entrapment database. Error bars representing mean ± standard deviation.
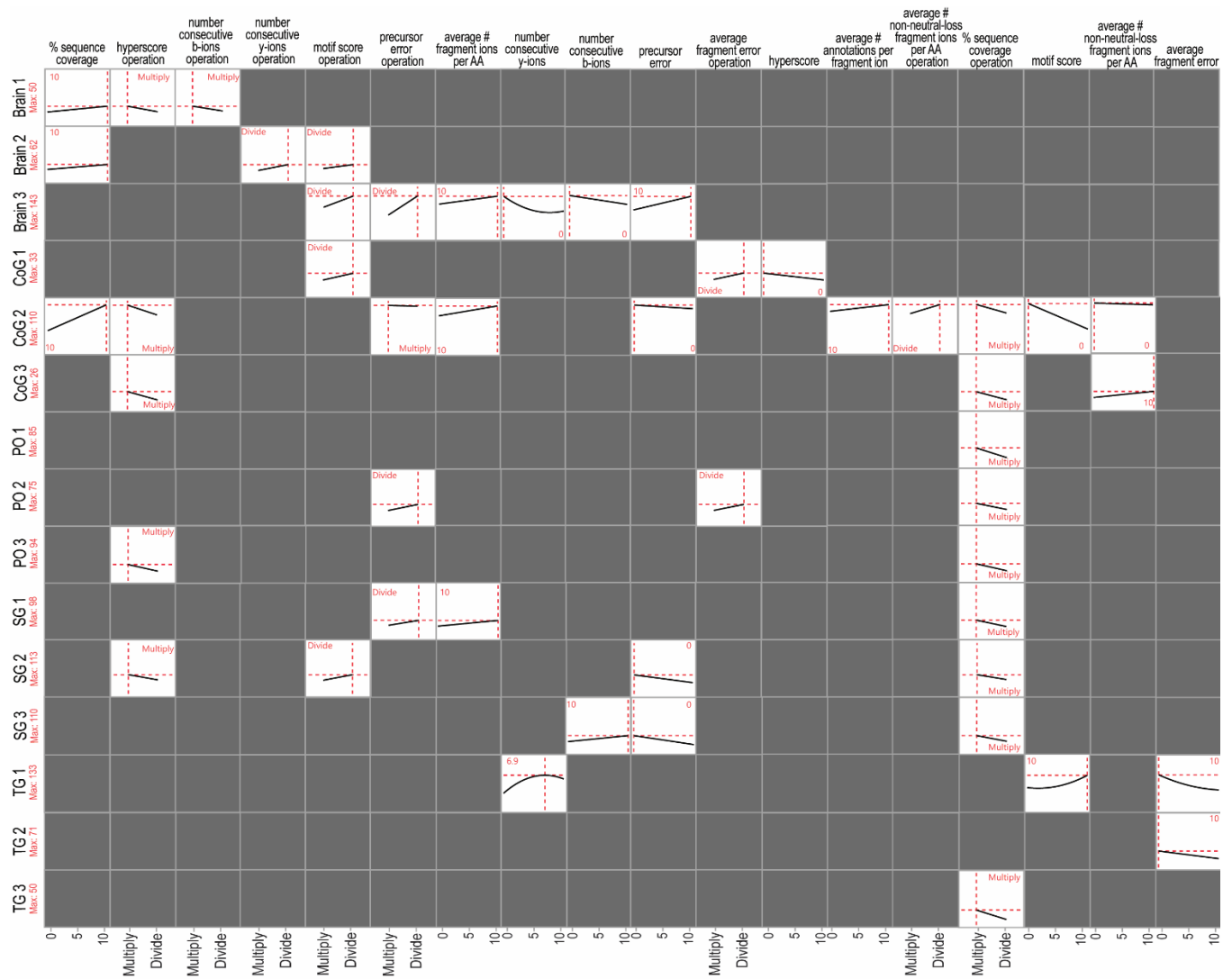
Figure S4: Results of definitive screening design when optimized for desirability, marked by the maximum number of unique identifications at 1% FDR, with responses from 15 samples. Results shown here are reported as significant factors.
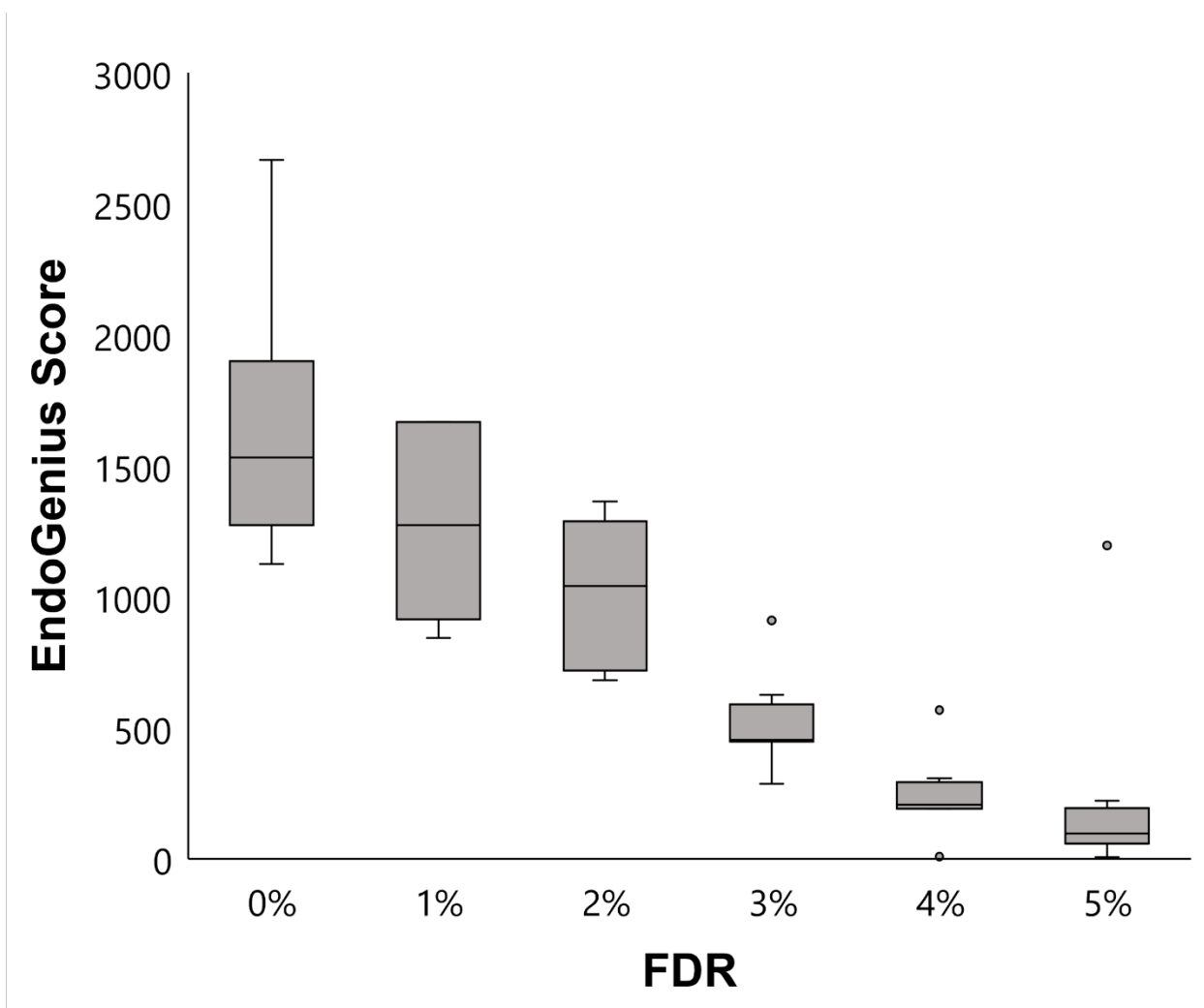
Figure S5: Correlation of false-discovery rate to the EndoGenius score, with results summarized from 15 spectral datasets.