**Proteoform identification using multiplexed top-down mass spectra**

Zhige Wang, Xingzhao Xiong, and Xiaowen Liu

1Deming Department of Medicine, Tulane University, New Orleans, Louisiana, 70112, United States

**Table S1**: Parameter settings for TopFD

|  |  |
| --- | --- |
| Input Parameter | Value |
| Maximum charge | 30 |
| Maximum mass | 70,000 Da |
| m/z error tolerance of spectral peaks | 0.02 m/z |
| MS1 signal-to-noise ratio | 3 |
| MS/MS signal-to-noise ratio | 1 |
| Use MSDeconv score | False |
| ECScore cutoff | 0.5 |
| Number of MS1 scans to detect a feature | 3 |
| Use noise levels in single scans to filter | False |
| Disable final filtering of envelopes | False |
| Disable additional feature search | False |

**Table S2**: Parameter settings of TopPIC for computing the intensity ratio between the first and second most abundant precursors in the spectra with two proteoform identifications in the E. coli data set

|  |  |
| --- | --- |
| Input Parameter | Value |
| Fixed modification | Carbamidomethylation on cysteine |
| Allowed N-terminal forms | None, NME, NME\_ACETYLATION, M\_ACETYLATION |
| Maximum number of mass shift | 1 |
| Minimum value of a mass shift | -500 Da |
| Maximum value of a mass shift | 500 Da |
| Use a shuffled decoy database | True |
| Mass error tolerance | 10 ppm |
| Proteoform error tolerance | 1.2 Da |
| Spectrum level cutoff type | FDR |
| Spectrum level cutoff value | 0.01 |
| Proteoform level cutoff type | FDR |
| Proteoform level cutoff value | 0.01 |
| Use TopFD Features | True |

**Table S3**: Parameter settings of TopPIC for generating the MATCH and SHIFT dataset

|  |  |
| --- | --- |
| Input Parameter | Value |
| Fixed modification | Carbamidomethylation on cysteine |
| Allowed N-terminal forms | None, NME, NME\_ACETYLATION, M\_ACETYLATION |
| Maximum number of mass shift | 1 |
| Minimum value of a mass shift | -500 Da |
| Maximum value of a mass shift | 500 Da |
| Use a shuffled decoy database | False |
| Mass error tolerance | 10 ppm |
| Proteoform error tolerance | 1.2 Da |
| Spectrum level cutoff type | E-value |
| Spectrum level cutoff value | 10-6 |
| Proteoform level cutoff type | E-value |
| Proteoform level cutoff value | 0.01 |
| Use TopFD Features | True |

**Table S4**: Parameter settings of TopPIC for searching the MATCH-ERROR and SHIFT-ERROR dataset

|  |  |
| --- | --- |
| Input Parameter | Value |
| Fixed modification | Carbamidomethylation on cysteine |
| Allowed N-terminal forms | None, NME, NME\_ACETYLATION, M\_ACETYLATION |
| Maximum number of mass shift | 1 |
| Minimum value of a mass shift | -500 Da |
| Maximum value of a mass shift | 500 Da |
| Use a shuffled decoy database | False |
| Mass error tolerance | 10 ppm |
| Proteoform error tolerance | 1.2 Da |
| Spectrum level cutoff type | E-value |
| Spectrum level cutoff value | 0.01 |
| Proteoform level cutoff type | E-value |
| Proteoform level cutoff value | 0.01 |
| Use TopFD Features | False |

**Table S5**: Parameter settings of TopPIC for generating the pseudo-multiplexed dataset

|  |  |
| --- | --- |
| Input Parameter | Value |
| Fixed modification | Carbamidomethylation on cysteine |
| Allowed N-terminal forms | None, NME, NME\_ACETYLATION, M\_ACETYLATION |
| Maximum number of mass shift | 1 |
| Minimum value of a mass shift | -500 Da |
| Maximum value of a mass shift | 500 Da |
| Use a shuffled decoy database | False |
| Mass error tolerance | 10 ppm |
| Proteoform error tolerance | 1.2 Da |
| Spectrum level cutoff type | E-value |
| Spectrum level cutoff value | 0.01 |
| Proteoform level cutoff type | E-value |
| Proteoform level cutoff value | 0.01 |
| Use TopFD Features | True |

Supplementary Table S3: Parameter settings for TopPIC for precursor selection test data generation and pseudo-multiplexed data generation.

|  |  |
| --- | --- |
| Input Parameter | Value |
| Fixed modification | Carbamidomethylation on cysteine |
| Allowed N-terminal forms | None, NME, NME\_ACETYLATION, M\_ACETYLATION |
| Maximum number of mass shift | 1 |
| Minimum value of a mass shift | -500 Da |
| Maximum value of a mass shift | 500 Da |
| Use a shuffled decoy database | False |
| Mass error tolerance | 10 ppm |
| Proteoform error tolerance | 1.2 Da |
| Spectrum level cutoff type | E-value |
| Spectrum level cutoff value | 0.01 |
| Proteoform level cutoff type | E-value |
| Proteoform level cutoff value | 0.01 |
| Score threshold for filtering modification | 0.15 |

Supplementary Table S5: Parameter settings for TopPIC for searching noised pseudo-multiplexed dataset

|  |  |
| --- | --- |
| Input Parameter | Value |
| Fixed modification | Carbamidomethylation on cysteine |
| Allowed N-terminal forms | None, NME, NME\_ACETYLATION, M\_ACETYLATION |
| Maximum number of mass shift | 1 |
| Minimum value of a mass shift | -20075 Da |
| Maximum value of a mass shift | 20075 Da |
| Use a shuffled decoy database | False |
| Mass error tolerance | 10 ppm |
| Proteoform error tolerance | 1.2 Da |
| Spectrum level cutoff type | E-value |
| Spectrum level cutoff value | 0.01 |
| Proteoform level cutoff type | E-value |
| Proteoform level cutoff value | 0.01 |
| Score threshold for filtering modification | 0.15 |

Supplementary Table S6: Parameter settings for TopPIC for SCPF correction analysis

|  |  |
| --- | --- |
| Input Parameter | Value |
| Fixed modification | None |
| Allowed N-terminal forms | None, NME, NME\_ACETYLATION, M\_ACETYLATION |
| Maximum number of mass shift | 1 |
| Minimum value of a mass shift | -500 Da |
| Maximum value of a mass shift | 500 Da |
| Use a shuffled decoy database | True |
| Mass error tolerance | 10 ppm |
| Proteoform error tolerance | 1.2 Da |
| Spectrum level cutoff type | E-value |
| Spectrum level cutoff value | 100,000 |
| Proteoform level cutoff type | E-value |
| Proteoform level cutoff value | 100,000 |
| Score threshold for filtering modification | 0.15 |