Meyer-Arendt et al., Supplementary Materials

Suppl. Figures
Suppl. Figure S1.  IsoformResolver input file format
Suppl. Figure S2.  IsoformResolver protein profile output, Sections 1-3
Suppl. Figure S3.  Methods of assigning bridge peptides used by different protein inference programs
Suppl. Figure S4.  Challenges to protein inference include volatility in choosing primary proteins.
Suppl. Figure S5.  Protein inference is affected by joining multiple datasets in different ways

Suppl. Tables
Suppl. Table S1.  Datasets used in these studies
Suppl. Table S2.  Comparison of protein inference programs
Suppl. Table S3.  Examples of ISD groups found in common by six protein inference programs

Suppl. Worksheets
Worksheet_1.xlsx:  IsoformResolver compare profile from Dataset 1A. Referred to by Suppl. Fig. S2.
Worksheet_2.xlsx:  Annotated IsoformResolver compare profile from Dataset 2. Referred to by Table 2 and Fig. 5.
Worksheet_3.xlsx:  Annotated IsoformResolver compare profile from Dataset 1D, analyzed by six different protein assembly programs. Referred to by Fig. 9 and Suppl. Table S3.
Suppl. Figure S1. IsoformResolver Input File Format

The IsoformResolver software requires an input directory with files in a comma-separated values (CSV) file format with one header line, and one row per MS/MS record.

Columns in each input .csv file must include the following:

- **File**: an identifier for each MS/MS file. This filename must be an identifier that is unique across all data files which are profiled at the same time.
- **Sample**: the LC-MS/MS (aka sample or rawfile) from which the MS/MS or MS^n file was extracted. This value must be a subset of the value in the File column.
- **Charge**: 1, 2, 3, or higher.
- **Golden**: A validated peptide sequence which the user believes to be the correct identification for the MS/MS named in the File field. This field may be empty (for an ambiguous identification) or contain an “X” (for a clearly rejected identification), in which case the entire line is ignored during IsoformResolver protein assembly.
- **Confidence**: A string indicating the basis for validation of the peptide. For example, values we often use include S for an identification by Sequest, M1 for an identification by Mascot, S-M1 for a peptide identified by both Sequest and Mascot, and MAE confidences. Other confidences may be provided by the user.
- **MrExp**: The observed mass of the peptide sequenced by MS/MS, derived from the parent ion mass.
- **CalcMass1**: The calculated mass of the peptide, including modifications.
- **Mowse1**: The Mascot score for the peptide identification. This value is optional, but if present, is carried through to the protein profile output. Other columns optionally supported include XCORR (Sequest score) and MISIM (MAE score).
- **VarMods1**: The variable modifications proposed by the search engine.
- **Intensity**: Peptide ion intensity, measured from the parent ion. Optional, but often helpful to view in the output. Stddev (standard deviation value per MS/MS) is also accepted when intensity information isn’t available.

Any row with ‘#’ in the first column is interpreted by the script as a comment and ignored unless followed by specific keywords. For instance, the use of the keyword “Peptide Tolerance (Units)” shown in the figure above directs IsoformResolver to parse the values in this line as part of peptide isoform detection.
Suppl. Figure S2. IsoformResolver Protein Profile Output

IsoformResolver creates an output file which concisely displays information about the peptides, proteins, MSD and ISD protein groups. In Section 1, peptides and all possible proteins from which the peptides could be derived are displayed within MSD groups, clearly indicating the inferred primary and secondary proteins. A simple example is shown in Panel a below in MSD group 1976 ([1]) which lists three peptides, each matching a single primary protein (IPI00023664.3), labeled by an identifier (2316) which tracks proteins in the minimal list. In contrast, MSD group 917 ([2]) has nine peptides which match four proteins, two of which are primary (labeled 1207, 1208) and two (labeled a, b) which lack independent peptide evidence and are therefore classified as secondary. Each peptide is also labeled with an identifier which reports the proteins to which they map ([3]). Identifiers for shared peptides combine two or more protein identifiers, for example, TFNLPLLMLGGG… ([4]) maps both to primary protein ‘1208’ (IPI00289601.10) and secondary protein ‘b’ (IPI00555868.1). MSD group 917.01 ([5]) shows two proteins that are indistinguishable because they are each accounted for by the same set of peptides. Identifiers of indistinguishable primary proteins are labeled with asterisks.

Although the peptide evidence may separate proteins into multiple MSD groups (e.g., 917 [2] and 917.01 [5]), such proteins are linked by gene family and belong to the same ISD group ([6]), even if the shared peptide sequences were not captured by MS/MS. (The .01 suffix in MSD group identifier 917.01 is used to associate it with MSD identifier 917.) IsoformResolver facilitates detection of these cases, by placing MSD groups belonging to the same ISD group adjacent to each other. The program also displays the number of peptides which report each protein and MSD group ([7]), the number of proteins per ISD group ([8]), the mass of each protein ([9]), and the percentage protein sequence coverage by observed peptides ([10]).

Information from the input files are displayed, including calculated peptide mass ([11]), closest observed mass ([12]), criteria used for validation (“Confidence”, [13]), highest search program scores ([14]), and standard deviation of fragment ion spectral quality ([15]). Total spectral counts ([16]) as well as spectral counts for each charge state are reported (not shown). Annotations are provided for each protein and ISD group, including the ISD group descriptor ([17]), protein descriptor ([18]), gene symbol ([19]), and GO and other cross references (not shown). Multiple experiments, each containing one or more LC/MS runs, can be compared (“compare profiles”), reporting results from each experiment in separate sets of columns ([20]). When a compare profile is requested, two additional files are generated – one which reports the numbers and names of proteins in the minimal list, and the other which reports the number of identified peptides in each experiment – which can be used as input for a Venn diagram display.

The detailed report of Section 1 is followed by a summary of results in Sections 2 and 3. Section 2 (Panel b) consists of a brief paragraph with the number of spectra, peptides, proteins, and protein groups, as well as the number of proteins supported by different numbers of peptides. Section 3 (Panel c) summarizes proteins inferred to be in the minimal list in the same order as Section 1, showing spectral counts and protein annotations (e.g., gene, ISD group, protein molecular weight, descriptor). When IsoformResolver is used to compare multiple experiments (“compare profile”), spectral counts for each experiment (e.g., Expt_1 and Expt_2) are listed side by side, allowing easy assessment of quantitative changes. Indistinguishable proteins are marked with an asterisk, and accession numbers of representative and other proteins are separated out. The number of peptides for each protein are indicated (# Peps), as are the number of peptides shared between two or more primary proteins (“bridge peptides”) which are reported on a separate line. Bridge peptides and their spectral counts are apportioned between proteins based
on spectral counts of non-bridge peptides, and the sum of non-redundant peptides and bridge peptides apportioned for each protein are reported (# App Peps). The sum of non-bridge and apportioned bridge peptide spectral counts equal total spectral counts (# App SC) (28). In the example shown here, the two spectral counts corresponding to bridge peptides 1207_1208 are apportioned to protein 1207 (0.8 SC) and 1208 (1.2 SC), while the single bridge peptide is apportioned to 1207 (0.4 peptides) and 1208 (0.6 peptides). The format of Section 3 is useful for further automation in spectral count analyses. The full profile output can be viewed in Suppl. Worksheet_1.xlsx.
### a. Isoform Resolver output, Section 1: MSD and ISD protein groups

| Protein ID | Peptide sequence | Pep MW | Confidence | XC | Mow | Obs MW | Stddev | # | XC | Mow | Obs MW | Stddev | # |
|------------|-----------------|--------|------------|----|-----|--------|--------|---|----|-----|--------|--------|---|
| MSD 1976   | ISD8-12137      |        |            |    |     |        |        | 3 |    |     |        |        |   |
| 2316       | IPI00023664.3   | 3      | 54420      | 3.5| 40.9| 3893.922| 3.7    | 6 | 34.3| 3893.917| 68     | 2  |
|            | LVQGISFSQPTC..  | 2302.095| Seq,Mas    | 5.2| 80.8| 3933.922| 46     | 2 |    |     |        |        |   |
| 2316       | SCMNQVVTSTTR    | 1512.678| Seq,Mas    | 3.6| 51   | 1512.679| 11     | 3 |    |     |        |        |   |
| 917        | ISD8-1383       | 9      | 7          |    |     |        |        | 3 |    |     |        |        |   |
| 1207       | IPI00013774.1   | 3      | 55103      | 16.8|     |        |        |   |    |     |        |        |   |
| 1208       | IPI000289601.10 | 5      | 65538      | 12  |     |        |        |   |    |     |        |        |   |
| a          | IPI00514649.1   |        | 24545      | 10  |     |        |        |   |    |     |        |        |   |
| 1208_b     | IPI00555886.1   |        | 27008      | 15.1|     |        |        |   |    |     |        |        |   |
| 1207       | DQIDDESYEAIFK...|        | 2043.958   | MAE | 12.9| 2043.959| 26.6   | 2 | 127.8| 2233.055| 373    | 16 |
| 1207       | LHIPPSNMTNQNTP...|        | 2233.055   | Seq, Mas | 5.6| 113.6| 2233.055| 15     | 2 | 127.8| 2233.055| 373    | 16 |
| 1207_1208_a| YGEYFPQTDGLR    | 1374.632| Seq, Mas   | 3.2| 60.3| 1374.631| 618    | 2 |    |     |        |        |   |
| 1207_a     | YYAVNPLR        | 1158.594| Seq, Mas   | 1.7| 16.3| 1158.594| 34     | 1 |    |     |        |        |   |
| 1208       | DGIDDESYGQIPK...|        | 2025.017   | Seq, Mas | 4.5| 47.8| 2025.017| 9      | 9 | 41   | 2025.018| 2768   | 5  |
| 1208       | VMEMYQPSAVVL... | 2513.147| Seq, Mas   | 4.5| 95.3| 2513.147| 45     | 2 | 17.7| 1160.56| 887    | 2  |
| 1208_b     | LHIPPSNMTNQNTN...|        | 2234.022   | Seq, Mas | 5.4| 86.2| 2234.021| 45     | 16| 85.6| 2234.022| 78     | 10 |
| 1208_b     | TFNLPLLMLGGG...  | 1822.988| Seq, Mas   | 4.6| 97.8| 1822.988| 129    | 5 | 4.9 | 107.9 | 1822.99 | 149  |
| MSD 917.01 | ISD8-1383       | 2      | 7          |    |     |        |        | 3 |    |     |        |        |   |
| 2745*      | IPI00006187.1   | 2      | 48848      | 12.4|     |        |        | 9 |    |     |        |        |   |
| 2745*      | IPI000212965.1  |        | 49111      | 12.4|     |        |        |   |    |     |        |        |   |
| 2745*      | HLFQPVINQVVD...  | 3749.787| Seq, Mas   | 4.9| 59.4| 3749.814| 15     | 1 |    |     |        |        |   |
| 2745*      | YGNYFPPTGD...    | 2316.987| Seq, Mas   | 2.6| 16.3| 2316.997| 25     | 1 |    |     |        |        |   |
b. Isoform Resolver output, Section 2: Summary of peptide and protein counts

Input files referenced 929171 dta files; 928490 not counting charge decoys.
We found a total of 372642 golden peptides and 26225 unique golden peptides and 40199 unique golden peptide ions.
After peptide isoform removal we had 26225 unique golden peptides left.
We qualified 3668 proteins and 3328 MSD protein groups and 3050 ISD protein groups.
This comes to 25.55% single peptide proteins.

Breakdown of protein per number of peptides:

| # peptides in protein | count |
|-----------------------|-------|
| 1                     | 937   |
| 2                     | 438   |
| 3                     | 355   |
| 4                     | 241   |
| 5                     | 218   |
| 6                     | 161   |

c. Isoform Resolver output, Section 3: Summary of spectral counts

| Distribution of minimum set proteins across sample sets: |
|---------------------------------------------------------|
| Expt_1 | Expt_2 | Prot ID | Prot accession | Gene | ISD group | # Peps | # App Peps | # App SC | MW     | Protein descriptor                                                                 | Indistinguishables                                      |
|--------|--------|---------|----------------|------|-----------|--------|------------|---------|--------|-----------------------------------------------------------------------------------|--------------------------------------------------------|
| 26     | 34     | 1206    | IPI00031768.1  | HOOK3| ISD8-5356 | 9      | 9          | 60      | 83126  | HOOK HOMOLOG 3                                                                    |                                                        |
| 21     | 18     | 1207    | IPI00013774.1  | HDAC1| ISD8-1383 | 3      | 3.4        | 39.8    | 55103  | HISTONE DEACETYLASE 1                                                             |                                                        |
| 2      | 0      | 1207_1208| IPI00013774.1_IPL | HDAC2| ISD8-1383 | 1      |            |         |        |                                                                                   |                                                        |
| 30     | 30     | 1208    | IPI00289601.10 | HDAC2| ISD8-1383 | 5      | 5.6        | 61.2    | 65538  | HISTONE DEACETYLASE 2                                                             |                                                        |
| 97     | 68     | 1209*   | IPI000018804.3 | TRIP10| ISD8-2599 | 9      | 9          | 165     | 62592  | ISOFORM 2 OF CDC42-INTERACTING PROTEIN 4                                         | IPI00168849.3-IPI                                      |
| 63     | 78     | 1210    | IPI00443909.1  | TMEM4| ISD8-2294 | 9      | 9          | 141     | 20652  | ISOFORM 1 OF MIR-INTERACTING Saposin-Like Protein Precursor                      |                                                        |
| 12     | 15     | 1211*   | IPI00302688.7  | ECHDC1| ISD8-896  | 9      | 9          | 27      | 33698  | ISOFORM 1 OF SENOYL-COA HYDRATASE DOMAIN-CONTAINING PROTEIN                       |                                                        |
| 40     | 40     | 1212    | IPI00217949.12 | UBE25| ISD8-9004 | 9      | 9          | 80      | 23845  | INTERFERON REGULATORY FACTOR 3                                                   |                                                        |
Suppl. Figure S3. Methods of assigning bridge peptides used by different protein inference programs. Protein-centric profiling methods replicate or arbitrarily assign bridge peptides to proteins, whereas IsoformResolver reports bridge peptides in the context of their MSD protein groups and ISD protein groups, thus displaying related proteins adjacently in the output.

a. TPP ProteinProphet. The output replicates bridge peptides between different proteins, and assigns weights according to distribution. Bridge peptides are those with weight < 1.0, assigned to proteins which are often displayed in different areas of the profile. Replication of peptides in the output may suggest higher confidence in inferred proteins than warranted.

In this example, peptides listed with weight = 0.50 were bridge peptides, assigned to high confidence proteins IPI00030783.1 and IPI00103415.1 (STAT5A and STAT5B). Each of the two protein identifications were supported by only 2 or 3 unique peptides, although 9 and 10 peptides were listed in each case.
In this example, although they are related to STAT5, the proteins IPI0030782.1 and IPI00743541.1 (STAT6) are displayed in a different location in the protein profile output. The peptide in IPI00743541.1 is also contained within IPI0030782.1, so it is considered a secondary (subset) protein and assigned a weight of 0.0.

b. IsoformResolver. Peptides, proteins, and MSD protein groups which belong to the same ISD protein group are displayed adjacently. Here STAT5 and STAT6 are listed together in the output, despite the fact that none of the observed peptides were common to both.
Suppl. Figure S4. Challenges to protein inference include volatility in choosing primary proteins. An illustration of how protein inference introduces ambiguity in assigning primary proteins. The MSD group includes seven proteins, from which IsoformResolver infers two primary proteins, IPI00181997.7 and IPI00479677.3. Given equal likelihood for peptides VFH..., TQT..., and GSL..., selection of the two inferred proteins is arbitrary, because the primary proteins could also have been assigned as IPI00181997.7 & IPI00376351.2, or IPI00479677.3 & IPI00376351.2.
Ambiguity in assigning primary proteins
IPI00181997.7 and IPI00479677.3 inferred, but other equally likely inferences are possible

VFH = VFHHNAWDNVEWSEEQAAAER; HWL = HWLFTEFPELAPSQNQNHLK; NNE = NNEDGPGLIMEEQHK;
TQT = TQTPPVEENVTQK; ISD = ISDLEICADEFGSSATYR; GSL = GSLDIIILIFVLSAIVPDK; VYF = VFYFTQEEQDLFTTLTAGEK
Suppl. Figure S5. Protein inference is affected by joining multiple datasets in different ways.  

a. In a pooled analysis, peptides from different fractions are combined prior to protein inference. Pooling peptides and then performing inference yields the simplest solution with the smallest number of proteins, as shown at the right (Dataset 1B), but loses important information when analyzing fractions separately.  
b. In an aggregate analysis, peptides from LC-MS/MS datasets of different fractions from a chromatographically resolved sample are first analyzed by inference, then the proteins are combined.  
c. In a compare profile, IsoformResolver combines the strengths of pooled and aggregate analyses, by pooling the datasets to identify the minimal list proteins, and then displaying spectral counts for each individual dataset.
a. Pooled Analysis

Expt 1  Expt 2  pool  peptides  protein list
Expt N  peptides  protein list

KLP=KLPLMALSTTMAESFK; LSE=LSEEELPAILK; SLS=SLSSLDTALAELR; TPG=TPGTGSLAAAVETASGR; SPP=SPPETAAPVEDMAR; LEE=LEEAQAYLAAGQHDLPHYVIESIADLTEGLED

b. Aggregate Analysis

Expt 1  peptides  proteins in Expt 1
Expt 2  peptides  proteins in Expt 2
Expt N  peptides  proteins in Expt N

| Fraction | KLP | LSE | SLS | TPG | SPP | LEE |
|----------|-----|-----|-----|-----|-----|-----|
| fr# 15   | x   |     | x   |     |     |     |
| fr# 16   | x   | x   |     |     |     |     |
| fr# 17   |     |     | x   |     |     |     |
| fr# 22   | x   |     |     |     | x   |     |
| fr# 23   | x   |     |     |     | x   |     |
| fr# 24   |     | x   |     |     |     |     |

Proteins inferred:
- IPI00025340
- IPI00025340
- IPI00025340
- IPI00444788, IPI00445123, IPI00456744, IPI00743804
- IPI00444788, IPI00445123, IPI00456744, IPI00456745, IPI00743804

1° CDNA FLJ44925 FIS. CLONE BRAMY3014613. HIGHLY SIMILAR TO SJ3BP1.
2° CDNA FLJ4492 FIS. CLONE BLADE2002310. SIMILAR TO SH3BP1 (FRAGMENT).
2° ISOFORM 1 OF SH3 DOMAIN-BINDING PROTEIN 1
2° ISOFORM 2 OF SH3 DOMAIN-BINDING PROTEIN 1
1° 67 KDA PROTEIN.


c. Compare Profile

Expt 1  peptides  protein list
Expt 2  pool  peptides  protein list
Expt N  peptides  protein list

| Protein ID | Accession | Protein descriptor | Gene |
|------------|-----------|--------------------|------|
| 763        | IPI00444788.1 | CDNA FLJ44925 FIS. CLONE BRAMY3014613. HIGHLY SIMILAR TO HOMO SAPIENSSH3-DOMAIN BINDING PROTEIN 1. | PDXP |
| 764        | IPI00025340.3 | PYRIDOXAL PHOSPHATE PHOSPHATASE. | PDXP |
| a          | IPI00445123.1 | CDNA FLJ44592 FIS. CLONE BLADE2002310. HIGHLY SIMILAR TO HOMO SAPIENSSH3-DOMAIN BINDING PROTEIN 1 (FRAGMENT). | SH3BP1 |
| b          | IPI00456744.1 | ISOFORM 1 OF SH3-DOMAIN BINDING PROTEIN 1. | SH3BP1 |
| c          | IPI00456745.3 | ISOFORM 2 OF SH3-DOMAIN BINDING PROTEIN 1. | SH3BP1 |
| d          | IPI00743804.1 | 67 KDA PROTEIN. | PDXP |

Protein ID  Peptide sequence  fr#15  fr#16  fr#17  fr#22  fr#23  fr#24
763_764     LEEAQAYLAAGQHDLPHYVIESIADLTEGLED 2 1
763_a_b_c_d LSEEELPAILK 2 1
763_a_b_c_d SLSSLDTALAEK 2
763_a_b_d SPPETAAPVEDMAR 1 2
763_b_c KLPLMALSTTMAESFK 1
764 TPGTGSLAAAVETASGR 3 2 1

Suppl. Figure S5, Meyer-Arendt et al.
Suppl. Table S1. **Datasets used in these studies.** Data collection was carried out by LC-MS/MS on three samples of proteins from human cell lines.

| Dataset | Application | DTAs  | Peptides       | Data analysis                                                                                           |
|---------|-------------|-------|----------------|---------------------------------------------------------------------------------------------------------|
| **Dataset 1** WM115 cell line | | | | |
| Dataset 1A | Examine extent of problems introduced by protein inference | 929,171 | 372,642 (26,225 non-redundant, 1,784 bridge) | Searched against IPI v.3.27 separated target and decoy databases by Mascot and Sequest, and validated by MAE. Parent ion tolerance 1.2 Da, 8 amino acid minimum length. Rejected when > 10 ppm or < -5 ppm than predicted. |
| Dataset 1B | Compare pooled vs. aggregate dataset results | 462,445 | 181,460 (21,414 non-redundant, 1,321 bridge) | |
| Dataset 1C | Compare protein identification software | 41,659 | 15,929 (3,939 non-redundant, 155 bridge) | |
| Dataset 1D | Compare protein identification software | 5,489 | 1,867 (1,314 non-redundant, 63 bridge) | |
| **Dataset 2** K562 cell line | | | | |
| Replicate 1 (1 rawfile) | Compare replicate runs for protein overlap | 16,983 | | See Table 2 |
| Replicate 2 (1 rawfile) | | 17,349 | | |
| Replicate 3 (1 rawfile) | | 16,638 | | |
| **Dataset 3** K562 cell line | | | | |
| Lower depth (2 rawfiles) | Depth of sampling experiment | 29,907 | | See Table 3 |
| Higher depth (25 rawfiles) | | 252,205 | | |
| Simulated lower depth (chosen from 25 rawfiles) | | 24,338 | | |

"13"
### Suppl. Table S2. Comparison of protein inference programs.

| Factor                        | Isoform Resolver | IDPicker       | Panoramics  | Phenyx                               | Scaffold                               | TPP Protein Prophet |
|-------------------------------|------------------|----------------|-------------|--------------------------------------|----------------------------------------|---------------------|
| Cost                          | Free             | Free           | Free        | Three versions, one is free          | Fee required                           | Free                |
| Search methods supported      | Mascot, Sequest, X!Tandem, X!Hunter | MyriMatch, X!Tandem, Sequest, Mascot | Mascot | OLAF, Mascot, Sequest, Scaffold, X!Tandem | Sequest, Mascot, Sequest, Scaffold, X!Tandem | Sequest, Mascot, SpectraST, X!Tandem |
| Program description           | Command line     | HTML           | HTML        | HTML or downloaded GUI program      | Downloaded GUI program                 | HTML or command line |
| Ease of use                   | Moderate         | Moderate       | Very easy   | Easy to moderate                     | Easy                                   | Moderate            |
| Determining accuracy, validation | Allows filtering by physico-chemical properties | By probability | No tuning parameters | Various probability and proprietary scores | By probability | By probability |
| Bridge peptide handling       | Uses protein groups and avoids replication | Uses protein groups and avoids replication | Replicates bridge peptides | Replicates bridge peptides | Replicates bridge peptides | Replicates bridge peptides |
| Protein reporting             | Primary, representative and secondary proteins | Representative only (not all primary and no secondary proteins) | Primary only (no representative or secondary proteins) | Representative and secondary proteins (not all primary) | Primary only (no representative or secondary proteins) | Primary and secondary proteins (no representative) |
| Protein databases             | DAT or FASTA format; converts into peptide-centric format | Requires target decoy database in FASTA format | FASTA format | FASTA format, limited in the free version | FASTA format | FASTA format |
| Output display                 | Spreadsheet      | HTML           | HTML        | HTML or spreadsheet                  | HTML or spreadsheet                     | XML or spreadsheet |
| Compare profiles?             | Yes, for any grouping | Yes, for any grouping, even hierarchical | No         | Limited to small datasets in the free version | Yes, by sample | No, although ProteinProphet in CPAS does |
| Throughput capabilities       | Good with large datasets | Limited with large datasets | One rawfile at a time | Limited in the free version | Good with large datasets | Limited with large datasets |
| Quantification?               | Yes              | Yes            | No          | Yes                                  | Yes                                    | Yes                |
Suppl. Table S3. Examples of ISD groups found in common by six protein inference programs. The five examples illustrate cases shown in Fig. 9. Proteins inferred by each program are shown in the columns on the left, while spectral counts reported by each program are shown in columns on the right. The programs tested were TPP ProteinProphet (TPP); Panoramics (Pan); Scaffold (Sca); IsoformResolver, in the default, primary protein selecting mode (IRP); IsoformResolver in the representative protein selecting mode (IRR); Phenyx (Phe); and IDPicker (IDP). For more detail, all 112 annotated ISD protein groups can be viewed in Suppl. Worksheet_3.xlsx.

Case 1. All programs identify the same protein. Not all programs identified all peptides, but all programs identified the same protein in the group (IPI0025019.3).

| Proteins inferred by each program | Protein | Spectral Counts |
|----------------------------------|---------|----------------|
| | | | TPP | Pan | Sca | IRP | IRR | Phe | IDP |
| | IPI0025019.3 PROTEASOME SUBUNIT TYPE 1 PRECURSOR. Gene=PSMB1 | 1 | 1 | 1 | 1 | 1 | 1 |
| a | IPI00556291.1 PROTEASOME SUBUNIT 1 SUBUNIT VARIANT (FRAGMENT). No gene annotation. | 1 | 1 | 1 | 1 | 1 | 1 |
| 1 | AGGSASAMLQPLDNQVGFK | 1 | 1 | 1 | 1 | 1 | 1 |
| 1 | NMQNEHVELQLD | 2 | 2 | 2 | 2 | 2 | 2 |
| 1_a | AMTTGAIAMLSTILYSR | 1 | 1 | 1 | 1 | 1 | 1 |
| 1 | GAVYSFDPVGYSQR | 1 | 1 | 1 | 1 | 1 | 1 |

Case 2. Disagreement between programs which select primary protein vs. representative proteins. The programs that select and report primary proteins (TPP ProteinProphet, Panoramics, Scaffold, IsoformResolver) all selected the same three primary proteins, and the programs that select representative proteins (IsoformResolver in representative protein selection mode, Phenyx, IDPicker) selected the same representative protein. However, none of the peptides provided distinguishing evidence for any specific isoform of Drebrin-like protein. The primary protein selecting programs retained this ambiguity while the representative selecting programs selected one of the three arbitrarily.

| Proteins inferred by each program | Protein | Spectral Counts |
|----------------------------------|---------|----------------|
| | | | TPP | Pan | Sca | IR | Phe | IDP |
| | IPI00101968.3 ISOFORM 3 OF DREBRIN-LIKE PROTEIN. Gene=DBNL | 1 | 1 | 1 | 1 | 1 | 1 |
| | IPI00396437.3 ISOFORM 2 OF DREBRIN-LIKE PROTEIN. Gene=DBNL | 1 | 1 | 1 | 1 | 1 | 1 |
| | IPI00456925.3 ISOFORM 1 OF DREBRIN-LIKE PROTEIN. Gene=DBNL | 1 | 1 | 1 | 1 | 1 | 1 |
| 2_a | AEEDVEPECIMEK | 1 | 1 | 1 | 1 | 1 | 1 |
| 2_a | FQDVGPQAPVGYQK | 1 | 1 | 1 | 1 | 1 | 1 |
| 2_a | GYPDGHFMFPANYVELIE | 1 | 1 | 1 | 1 | 1 | 1 |
| 2_a | VAGTGELEGMEVSN | 1 | 1 | 1 | 1 | 1 | 1 |
Case 3. Disagreement between programs which select representative proteins. Each program identified the same peptide, and the same primary proteins were selected by TPP ProteinProphet, Panoramics, Scaffold, and IsoformResolver. However, there was no agreement between representative protein selecting programs (IsoformResolver in representative protein selection mode, Phenyx, IDPicker) as to which of the three indistinguishable proteins should be reported.

| Proteins inferred by each program | Protein | Spectral Counts |
|----------------------------------|---------|-----------------|
|                                  |         | TPP  Pan  Sca  IRP  IRR  Phe  IDP |
|                                  |         | √       √       √       √       √       √       |
|                                  |         | 3* IPI00384745.3 HSR1 PROTEIN. No gene annotation. |
|                                  |         | √       √       √       √       √       √       |
|                                  |         | 3* IPI00396387.3 GUANINE NUCLEOTIDE BINDING PROTEIN-LIKE 1. Gene= GNL1 |
|                                  |         | √       √       √       √       √       √       |
|                                  |         | 3* IPI00797329.1 GUANINE NUCLEOTIDE BINDING PROTEIN-LIKE 1. No gene annotation. |
|                                  |         | √       √       √       √       √       √       |
|                                  |         | 3* EVYLOPVSAELLELDIR |

| Proteins inferred by each program | Protein | Spectral Counts |
|----------------------------------|---------|-----------------|
|                                  |         | TPP  Pan  Sca  IR  Phe |
|                                  |         | √       √       1       1       1       1       1 |
|                                  |         | 3* IPI00384745.3 HSR1 PROTEIN. No gene annotation. |
|                                  |         | √       √       1       1       1       1       1       |
|                                  |         | 3* IPI00396387.3 GUANINE NUCLEOTIDE BINDING PROTEIN-LIKE 1. Gene= GNL1 |
|                                  |         | √       √       1       1       1       1       1       |
|                                  |         | 3* IPI00797329.1 GUANINE NUCLEOTIDE BINDING PROTEIN-LIKE 1. No gene annotation. |
|                                  |         | √       √       1       1       1       1       |
|                                  |         | 3* EVYLOPVSAELLELDIR |

Case 4. Disagreement among programs which select primary proteins. Each program identified the same peptide, and the same representative proteins were selected by IsoformResolver in representative protein selection mode, Phenyx, and IDPicker. However, programs which select primary proteins (TPP ProteinProphet, Panoramics, Scaffold, and IsoformResolver) disagreed with each other.

| Proteins inferred by each program | Protein | Spectral Counts |
|----------------------------------|---------|-----------------|
|                                  |         | TPP  Pan  Sca  IRP  IRR  Phe  IDP |
|                                  |         | √       √       √       √       √       √       |
|                                  |         | 4* IPI00220834.8 ATP-DEPENDENT DNA HELICASE 2 SUBUNIT 2. Gene=XRCC5 |
|                                  |         | √       √       √       √       √       √       |
|                                  |         | 4* IPI00792121.1 PROTEIN. Gene=XRCC5 |
|                                  |         | √       √       √       √       √       √       |
|                                  |         | 4* IPI00795088.1 HYPOTHETICAL PROTEIN XRCC5. Gene=XRCC5 |
|                                  |         | √       √       √       √       √       √       |
|                                  |         | 4* YAPTEAQLNAVDALIDSMSLAK |

Case 5. Disagreement among programs which select primary and representative proteins. Each of the 6 programs identified the same peptide, but programs within each group failed to reach agreement about which protein to infer.

| Proteins inferred by each program | Protein | Spectral Counts |
|----------------------------------|---------|-----------------|
|                                  |         | TPP  Pan  Sca  IRP  IRR  Phe  IDP |
|                                  |         | √       √       √       √       √       √       |
|                                  |         | 5* IPI00339320.5 ANKYRIN REPEAT DOMAIN-CONTAINING PROTEIN 15. Gene=ANKRD15 |
|                                  |         | √       √       √       √       √       √       |
|                                  |         | 5* IPI00479846.1 ANKYRIN REPEAT DOMAIN-CONTAINING PROTEIN 15. Gene=ANKRD15 |
|                                  |         | √       √       √       √       √       √       |
|                                  |         | 5* IPI00551051.4 ANKYRIN REPEAT DOMAIN 15. Gene=ANKRD15 |
|                                  |         | √       √       √       √       √       √       |
|                                  |         | 5* IPI00646764.2 ANKYRIN REPEAT DOMAIN PROTEIN 15 ISOFORM B. Gene=ANKRD15 |
|                                  |         | √       √       √       √       √       √       |
|                                  |         | 5* AGYTPIMLAALAAVEAEK |

16