Mass spectrometry Proteomics and MIAPE

Florian Breitwieser - Research Center of Molecular Medicine, Vienna, Austria

Nov 18, 2010

Outline

BioC-devel meeting Europe 17.-18. 11. 2010

- MS-MS Proteomics
- Standard classes
- Protein Grouping

Detection of proteins using mass spectrometer

- Workflow
  1. Sample: Proteins
  2. Digest to: Peptides (Less diversity, more complexity)
  3. Mass-Spectrometer detects: m/z & Intensity (MS) + Spectrum (MS/MS)
  4. Search-engine identifies: Peptides and Proteins
- Difficulties
  - Nature of data
  - Standardisation: Formats and minimal information (reproducible research)
  - Infrastructure: R and Bioconductor

HUPO PSI and MIAPE

- HUPO Proteomics Standards Initiative
  - Founded in April 2002
  - http://psidev.info/
  - Aims:
    - Create minimum reporting standards
    - Enable easier transfer of proteomics data
- MIAPE
  - Minimum Information about a proteomics experiment:
    - Mass Spectrometry and Mass Spectrometry Informatics
    - Gel Electrophoresis
    - Liquid Chromatography
  - Formats for MIAPE:MS (2.24) / MIAPE:MSI (1.1):
    - mzML for spectrum level information
    - mzIdentML for protein identification

Proteomics Formats

- Spectrum formats:
  - dta, pkl, mgf
  - ...
  - Seattle Proteome Center: mzXML
  - Proteomics Standard Initiative: mzData,mzML
- Identification formats
  - Mascot: DAT files
  - Phenyx: PIDRES XML files
  - X Tandem: XML files
  - ...
  - SPC: pepXML, protXML
  - PSI: mzIdentML

Classes: MIAPE:MS Experiment Information

```
setClass("MIAPE",
         representation=representation(
           creationDate="character",
           contact="character",

           software="character",
           software.version="character",
           software.contact="character",
           software.constomisations="character",
           software.uri="character",

           spectra.source.files="character",
           spectra.source.format="character",

           search.database="character",
           search.database.version="character",
           search.database.nSeq="character",
           search.database.filters="character",
           search.database.nSeqSearched="character",

           search.enzymes="character",
           search.enzymes.missedcleavages="numeric",
           search.enzymes.aditionalParams="character",

           search.modif.fixed="character",
           search.modif.variable="character",

           search.param.fragmentTolerance="character",
           search.param.parentTolerance="character",
           search.threshold.protein="list",
           search.threshold.peptide="list",

           search.additionalParams="list")
         )
```

Classes: MIAPE:MS ProteinGroup and MzIdent

```
setClass("ProteinGroup",
         representation(
                        proteinDescription = "data.frame",
                        ##AC,Description,Validation Status,#different peptide seq, coverage%
                        peptideToProtein = "matrix",
                        peptideDetails = "data.frame",
                        indistinguishableProteins = "character",
                        proteinGroups = "data.frame"
                        )
         )

setClass("MzIdent",
         contains = "eSet",
         representation(
           proteinGroup = "ProteinGroup",
           assayData = "list",                  # eSet
           ## spectra-matrices: columns m/z, intensity, charge
           phenoData = "AnnotatedDataFrame",    # eSet
           featureData = "AnnotatedDataFrame", # eSet
           ## retention time, peptide sequence, peptide modif, scores,
           ##  charge state, calculatedMassToCharge, experimentalMassToCharge, ...
           experimentData = "MIAPE",
           annotation = "character",            # eSet
           protocolData="AnnotatedDataFrame"    # eSet
           )
)
```

Protein Grouping

- Problem
  - Peptides are detected, not proteins
  - Peptides are often shared between proteins
    - especially between splice variants!
  - Create a minimal set: Occams Razor
- Algorithm
  1. For each protein: assemble list of peptides with which it is identified
  2. Indistinguishable proteins: Proteins detected with the same peptides
  3. Master proteins: Proteins with peptides specific to them
  4. Group proteins with no specific peptides to master proteins
  5. Classify

Proteomics in Bioconductor

- Good class representations
  - MIBBI: Minimum Information for Biological and Biomedical Investigations
- Multiple experiments?
- Quantitation
- Importers:
  - mzML and mzIdentML
  - pepXML, mgf, ...

–> Statistics
–> Applications