

# Labelled quantitative proteomics with MSnbase

Laurent Gatto

lg390@cam.ac.uk

Cambridge Centre For Proteomics  
University of Cambridge

European Bioinformatics Institute (EBI)

18<sup>th</sup> November 2010

# Plan

- 1 Introduction**
  - Motivation
  - Isobaric tags
- 2 Data structures**
- 3 Application**
  - A typical workflow
  - Use cases
- 4 QC with PRIDE**
- 5 Future work**

# Plan

- 1 Introduction**
  - Motivation
  - Isobaric tags
- 2 Data structures**
- 3 Application**
  - A typical workflow
  - Use cases
- 4 QC with PRIDE**
- 5 Future work**

## Motivation

- Many softwares are **black boxes** and *just* return values.
- There are some open softwares, but that do not (as far as I know) allow to perform exploratory data analysis.

## Goals

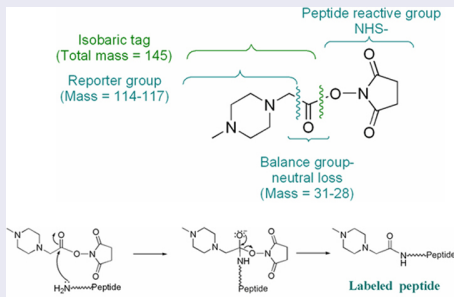
- Enable to load, process and quantify raw isobaric tag labelled MS<sup>a</sup> data in a controlled and reproducible way.
- Having a robust and annotation rich data structure for subsequent analyses.
- Have access to R and Bioc treasures by being eSet *et al.* compatible.

---

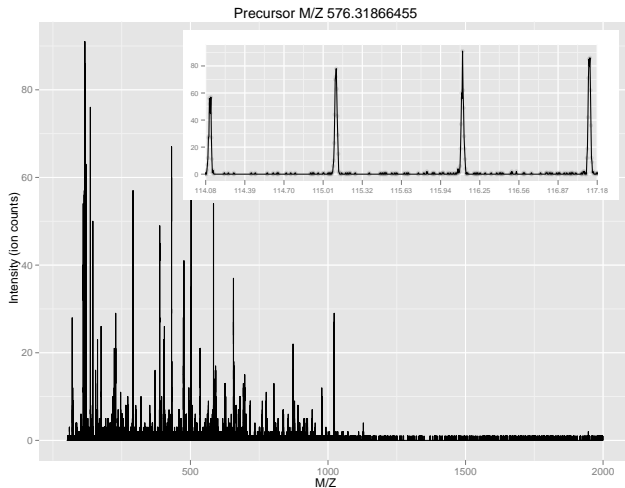
<sup>a</sup>but should also work for spectral counting

## Multiplexed tags

Samples (peptides) are labelled using different isobaric tags (below: iTRAQ4 – other: iTRAQ8, TMT6), combined and processed (separation - MSMS) together. The labelled peptides will behave in the same way until dissociation of the peptides where the reporter groups will be released and separate in MS2.



## Multiplexed tags - identification and quantification



# Plan

- 1 Introduction
  - Motivation
  - Isobaric tags
- 2 Data structures
- 3 Application
  - A typical workflow
  - Use cases
- 4 QC with PRIDE
- 5 Future work

## Classes

- MSnExp - MS(MS) experiment.
- Spectrum, Spectrum1 and Spectrum2 - mass spectra.
- ReporterIons defines reporter ions – data(iTRAQ4).
- MSnSet - quantified expression.
  
- Additional meta-data in MSnProcess and MIAPE.



```
> library("MSnbase")  
> getClass("MSnExp")
```

```
Class "MSnExp" [package "MSnbase"]
```

```
Slots:
```

```
Name:          spectra          process          fromFile  
Class:         list            MSnProcess      numeric
```

```
Name:          files            assayData      phenoData  
Class:         character        AssayData     AnnotatedDataFrame
```

```
Name:          featureData      experimentData  annotation  
Class: AnnotatedDataFrame      MIAxE         character
```

```
Name:          protocolData    __classVersion__  
Class: AnnotatedDataFrame      Versions
```

```
Extends:
```

```
Class "eSet", directly
```

```
Class "VersionedBiobase", by class "eSet", distance 2
```

```
Class "Versioned", by class "eSet", distance 3
```

```
> getClass("Spectrum2")
```

```
Class "Spectrum2" [package "MSnbase"]
```

```
Slots:
```

```
Name:          merged          ms1scan          precursorMz  
Class:         numeric         integer         numeric
```

```
Name: precursorIntensity  precursorCharge  scanindex  
Class:         numeric         integer         integer
```

```
Name:      collisionEnergy  msLevel          peaksCount  
Class:     numeric         integer          integer
```

```
Name:          rt      acquisitionNum  scanIndex  
Class:         numeric  integer          integer
```

```
Name:          mz          intensity  __classVersion__  
Class:         numeric  numeric          Versions
```

```
Extends:
```

```
Class "Spectrum", directly
```

```
Class "Versioned", by class "Spectrum", distance 2
```

```
> data(iTRAQ4)
> iTRAQ4

Object of class "ReporterIons"
iTRAQ4: '4-plex iTRAQ' with 4 reporter ions
- 114.13 +/- 0.05 (red)
- 115.13 +/- 0.05 (green)
- 116.13 +/- 0.05 (blue)
- 117.13 +/- 0.05 (yellow)

> iTRAQ4[1:2]

Object of class "ReporterIons"
iTRAQ4[1:2]: 'subset of 4-plex iTRAQ' with 2 reporter ions
- 114.13 +/- 0.05 (red)
- 115.13 +/- 0.05 (green)
```

```
> getClass("MSnSet")
```

```
Class "MSnSet" [package "MSnbase"]
```

```
Slots:
```

```
Name:          qual          process          files  
Class:         data.frame    MSnProcess    character
```

```
Name:    experimentData    assayData    phenoData  
Class:   MIAME             AssayData    AnnotatedDataFrame
```

```
Name:    featureData    annotation    protocolData  
Class:  AnnotatedDataFrame    character    AnnotatedDataFrame
```

```
Name:    .__classVersion__  
Class:   Versions
```

```
Extends:
```

```
Class "ExpressionSet", directly
```

```
Class "eSet", by class "ExpressionSet", distance 2
```

```
Class "VersionedBiobase", by class "ExpressionSet", distance 3
```

```
Class "Versioned", by class "ExpressionSet", distance 4
```

# Plan

- 1 **Introduction**
  - Motivation
  - Isobaric tags
- 2 **Data structures**
- 3 **Application**
  - A typical workflow
  - Use cases
- 4 **QC with PRIDE**
- 5 **Future work**

- 1 `readMzXMLData()` to create and `MSnExp` instance
- 2 `plot()` subset of `MSnExp` or `Spectrum`
- 3 Quality control (see later)
- 4 Processing: `removePeaks`, `bg.correct`, smoothing (TODO)
- 5 (although there is a `mergeSpectra` function, don't know yet how to combine MS2 spectra – see later)
- 6 Define `ReporterIons`
- 7 `quantify(MSnExp,ReporterIons)` to create an `MSnSet` instance and collect curve (peaks) statistics.

```
> f <- dir(system.file(dir = "extdata", package = "MSnbase"), full.names = TRUE)
> dummy <- readMzXMLData(f, verbose = FALSE)
> raw <- readMzXMLData("raw.mzXML", verbose = FALSE)

> raw

Object of class "MSnExp"
Object size in memory: 179.6 Mb
- - - Meta data - - -
Loaded from:
  raw.mzXML
- - - Processing information - - -
Data loaded: Thu Nov 18 01:26:31 2010
MSnbase version: 0.0.2
Xcms version: 1.25.1
- - - Spectra data - - -
MSn level(s): 2
Number of MS1 acquisitions: 759
Number of MS2 scans: 10910
Number of precursor ions: 10910
5393 unique MZs
Precursor MZ's: 400.2 - 1690.9
MSn M/Z range: 100 2069.28
MSn retention times: 1:20 - 60:0 minutes
```

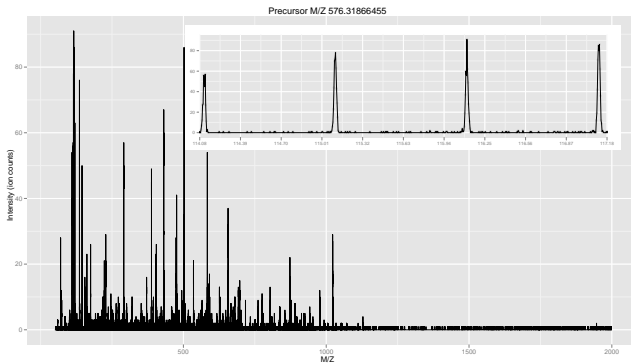
```
> raw[64]

Object of class "MSnExp"
  Object size in memory: 0.1 Mb
- - - Meta data - - -
  Loaded from:
    raw.mzXML
- - - Processing information - - -
Data loaded: Thu Nov 18 01:26:31 2010
MSnbase version: 0.0.2
Xcms version: 1.25.1
- - - Spectra data - - -
MSn level(s): 2
Number of MS1 acquisitions: 1
Number of MS2 scans: 1
Number of precursor ions: 1
  1 unique MZs
Precursor MZ's: 404.23 - 404.23
MSn M/Z range: 100 838
MSn retention times: 15:47 - 15:47 minutes
```



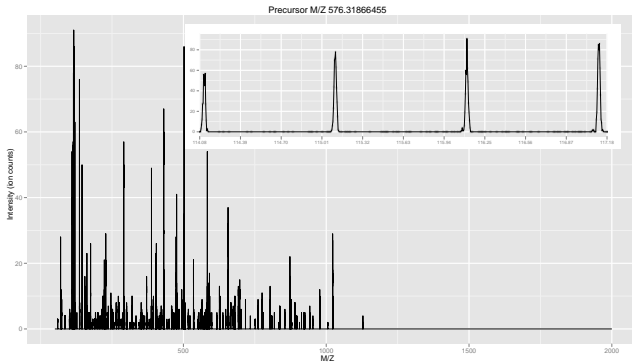
```
> print(plot(spectra(dummy)[[64]], full = TRUE, reporters = iTRAQ4))
```

GRID.VP.2

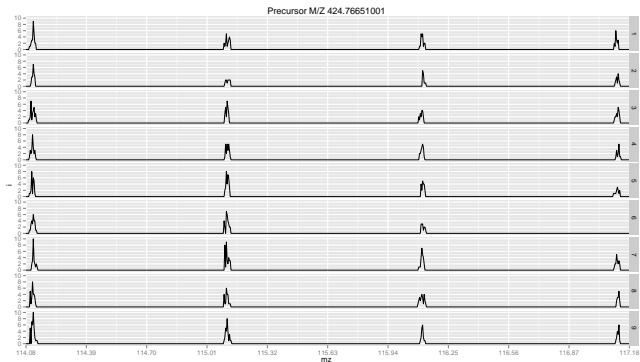


```
> dummy <- removePeaks(dummy, t = 3, verbose = FALSE)  
> print(plot(spectra(dummy)[[64]], full = TRUE, reporters = iTRAQ4))
```

GRID.VP.6



```
> print(plot(extractPrecSpectra(dummy, precursorMz(dummy)[1]),  
+ reporters = iTRAQ4))
```



```
> qnty <- quantify(dummy, reporters = iTRAQ4, method = "trap",
+   verbose = FALSE)
> qnty

MSnSet (storageMode: lockedEnvironment)
assayData: 98 features, 4 samples
  element names: exprs
protocolData: none
phenoData
  rowNames: iTRAQ4.114 iTRAQ4.115 iTRAQ4.116 iTRAQ4.117
  varLabels: mz reporters
  varMetadata: labelDescription
featureData
  featureNames: 424.76651001 424.76651001.1 ... 539.99688721.8 (98
  total)
  fvarLabels: index file ... collision.energy (9 total)
  fvarMetadata: labelDescription
experimentData: use 'experimentData(object)'
Annotation:
- - - Processing information - - -
Data loaded: Thu Nov 18 01:25:29 2010
Curves <= 3 set to '0': Thu Nov 18 03:01:11 2010
Curves <= 3 set to '0': Thu Nov 18 03:01:15 2010
iTRAQ4 quantification by trapezoidation: Thu Nov 18 03:01:35 2010
MSnbase version: 0.0.2
Xcms version: 1.25.1
```

## Number of times a precursor ion has been selected

Optimise MS paramters.

```
> allPrecs <- precursorMz(raw)
> number.selection <- c()
> ms1scanNums <- ms1scan(raw)
> for (mp in unique(allPrecs)) number.selection <- c(number.selection
+   length(unique(ms1scanNums[allPrecs == mp])))
> names(number.selection) <- unique(allPrecs)
> print(table(number.selection))
```

```
number.selection
```

```
  1    2    3    4
5337  52    2    2
```

## Assessing incomplete dissociation

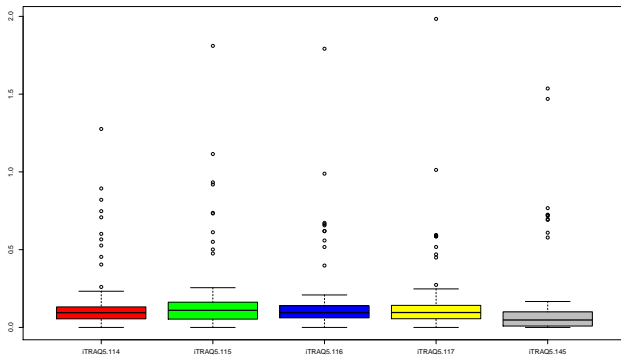
We can easily create a new set of reporter ions, including the 4 iTRAQ tags and a virtual tag at MZ 145, which corresponds to partially fragmented reporter tags and balance groups and quantify:

```
> data(iTRAQ5)  
> iTRAQ5
```

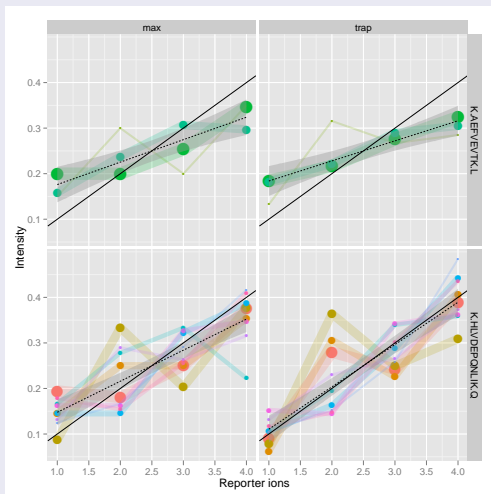
Object of class "ReporterIons"

```
iTRAQ4: '4-plex iTRAQ with isobaric tag' with 5 reporter ions  
- 114.13 +/- 0.05 (red)  
- 115.13 +/- 0.05 (green)  
- 116.13 +/- 0.05 (blue)  
- 117.13 +/- 0.05 (yellow)  
- 145.13 +/- 0.05 (grey)
```

```
> foo <- quantify(dummy, iTRAQ5, "trap", verbose = FALSE)  
> boxplot(exprs(foo), col = iTRAQ5@col)
```



## MS2 spectra are not equal





# Plan

- 1 **Introduction**
  - Motivation
  - Isobaric tags
- 2 **Data structures**
- 3 **Application**
  - A typical workflow
  - Use cases
- 4 **QC with PRIDE**
- 5 **Future work**

## Quality control

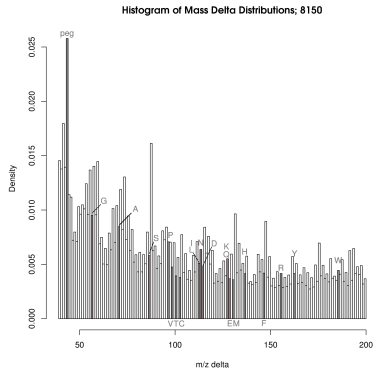
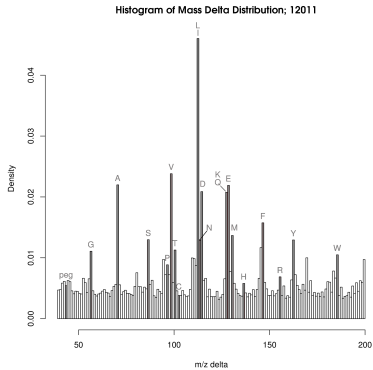
- Spectrum-level QC
- Use PRIDE<sup>a</sup> as reference<sup>b</sup>. (Query and create MSnExp objects.)

---

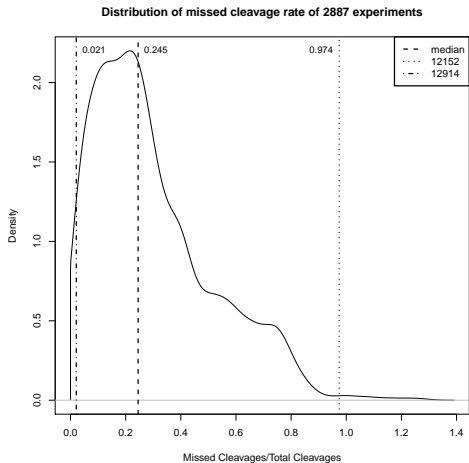
<sup>a</sup>PRoteomics IDentifications database

<sup>b</sup>In collaboration with Joe Foster from EBI. – Foster, Gatto et al., submitted

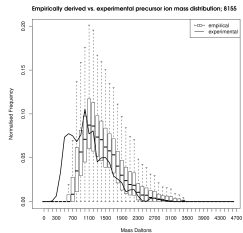
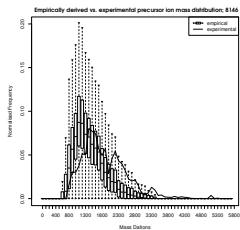
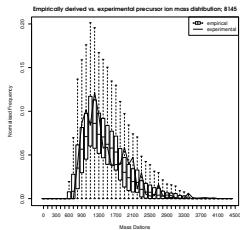
## Experiment-wide MS2 spectra



## Assessing digestion efficiency



## Precursor ions distribution



# Plan

- 1 **Introduction**
  - Motivation
  - Isobaric tags
- 2 **Data structures**
- 3 **Application**
  - A typical workflow
  - Use cases
- 4 **QC with PRIDE**
- 5 **Future work**

- <http://github.com/lgatto/MSnbase>
- Managing identification data (see RpepXML)
- MS2 spectra → peptides → proteins
- Peak picking, smoothing, centroiding...
- **Finish documentation, testing, release**
- **Use it!**

## Acknowledgement

- CCP team, especially Mike Deery, Phil Charles, Arnoud Groen and of course Kathryn Lilley.
- Joe Foster from PRIDE at EBI.

## Funding

 **BBSRC** Tools and Resources Development Fund Award.



Thank you for you attention.