isobar for developers

Florian P. Breitwieser, Jacques Colinge

January 7, 2015

Contents

1 Introduction

This documents highlights the structure of the S4 classes and methods in the `isobar` pacakge.

```
> library(isobar)
```

2 Classes

2.1 IBSpectra

```
> getClass("IBSpectra")

Virtual Class "IBSpectra" [package "isobar"]

Slots:

Name:          proteinGroup    reporterTagNames   reporterTagMasses
Class:          ProteinGroup           character             numeric

Name:    isotopeImpurities                 log          assayData
Class:             matrix              matrix          AssayData

Name:            phenoData         featureData       experimentData
Class: AnnotatedDataFrame AnnotatedDataFrame                 MIAxE
```

```
Name:          annotation        protocolData  .__classVersion__
Class:          character AnnotatedDataFrame        Versions

Extends:
Class "eSet", directly
Class "VersionedBiobase", by class "eSet", distance 2
Class "Versioned", by class "eSet", distance 3


Known Subclasses:
Class "iTRAQSpectra", directly
Class "TMTSpectra", directly
Class "iTRAQ4plexSpectra", by class "iTRAQSpectra", distance 2
Class "iTRAQ8plexSpectra", by class "iTRAQSpectra", distance 2
Class "TMT2plexSpectra", by class "TMTSpectra", distance 2
Class "TMT6plexSpectra", by class "TMTSpectra", distance 2
Class "TMT6plexSpectra2", by class "TMTSpectra", distance 2
Class "TMT10plexSpectra", by class "TMTSpectra", distance 2
```

identifications and quantitative values. Spectrums are identified as stemming from distinct peptides, and quantitative information of each spectrum are extracted from a certain m/z region.

IBSpectra class holds this qualitative and quantitative information. It is a virtual class. It extends eSet from Biobase to store meta-information of spectrum identifications and quantitative information (m/z and intensity) of reporter tags. eSet is extended by slots for protein grouping, tag names, tag masses and isotope impurity correction matrix.

ProteinGroup objects store the mapping and grouping of peptide level identifications to protein identifications.

IBSpectra is a virtual class. Currently used isobaric tagging kits iTRAQ 4plex and 8plex, and TMT 2plex and 6plex are implemented in the iTRAQ4plexSpectra, iTRAQ8plexSpectra, TMT2plexSpectra, TMT6plexSpectr and TMT10plexSpectr, respectively. These are subclasses of iTRAQSpectra and TMTSpectra, resp. which in turn are virtual subclasses of IBSpectra.

2.2   ProteinGroup

```
>   getClass("ProteinGroup")

Class "ProteinGroup" [package "isobar"]

Slots:

Name:          spectrumToPeptide                spectrumId
Class:                    character                data.frame


Name:          peptideSpecificity            peptideNProtein
Class:                    data.frame                     matrix
```

```
Name:   indistinguishableProteins        proteinGroupTable
Class:                    character                 data.frame

Name:          overlappingProteins     isoformToGeneProduct
Class:                       matrix                data.frame

Name:                  proteinInfo              peptideInfo
Class:                   data.frame               data.frame

Name:           .__classVersion__
Class:                    Versions

Extends:
Class "VersionedBiobase", directly
Class "Versioned", by class "VersionedBiobase", distance 2
```

mapped back to proteins. This mapping leads to protein groups, which explain the observed peptides according to the parsimony law.

A `ProteinGroup` object is generated when a `IBSpectra` object is created by `readIBSpectra`. Protein to peptide to spectrum mapping is extracted from a suitable identication format[1]

2.3 NoiseModel

```
>   getClass("NoiseModel")

Virtual Class "NoiseModel" [package "isobar"]

Slots:

Name:          na.region    low.intensity                f       parameter
Class:           numeric          numeric         function         numeric

Name:  .__classVersion__
Class:          Versions

Extends:
Class "VersionedBiobase", directly
Class "Versioned", by class "VersionedBiobase", distance 2

Known Subclasses: "ExponentialNoANoiseModel", "ExponentialNoiseModel", "InverseNoiseModel",
"InverseNoANoiseModel", "GeneralNoiseModel"
```

in the spectrum-level ratios of a certain experimental setup.

---

[1]IBSpectra CSV, and MzIdentML format. Mascot DAT and Phenyx pidres.xml format converters to IB-Spectra format are provided.

3   Session Information

The version number of R and packages loaded for generating the vignette were:

- R version 3.1.2 (2014-10-31), `x86_64-unknown-linux-gnu`

- Locale: `LC_CTYPE=en_US.UTF-8`, `LC_NUMERIC=C`, `LC_TIME=en_US.UTF-8`,
  `LC_COLLATE=C`, `LC_MONETARY=en_US.UTF-8`, `LC_MESSAGES=en_US.UTF-8`,
  `LC_PAPER=en_US.UTF-8`, `LC_NAME=C`, `LC_ADDRESS=C`, `LC_TELEPHONE=C`,
  `LC_MEASUREMENT=en_US.UTF-8`, `LC_IDENTIFICATION=C`

- Base packages: base, datasets, grDevices, graphics, methods, parallel, stats, utils

- Other packages: Biobase 2.26.0, BiocGenerics 0.12.1, isobar 1.12.2, plyr 1.8.1

- Loaded via a namespace (and not attached): Rcpp 0.11.3, SweaveListingUtils 0.6.2,
  distr 2.5.3, sfsmisc 1.0-27, startupmsg 0.9, tools 3.1.2