SwathXtend *

An R package for generating SWATH assay libraries and perform statistical analysis

Jemma Wu, Dana Pascovici and Xiaomin Song
APAF, Australia
jwu@proteome.org.au

May 3, 2016

SwathXtend is an R package aimed to facilitate extended assay library generation and stastical data analysis for SWATH data. This package contains the major functions decribed in Wu et al. 2016. This vignette describes how to use the funcitons in SwathXtend.

Extended libraries are built from a locally-generated seed library and some other libraries including in-house archived assay libraries or externally acquired entire proteome repository libraries.

Introduction

The first integrated DIA and quantitative analysis protocol, termed SWATH was shown to offer accurate, reproducible and robust proteomic quantification (Gillet et al 2012). An important concept in DIA analysis is use of a LC-retention time referenced spectral ion library to enable peptide identification from DIA generated multiplexed MS/MS spectra. SwathXtend is an R based software package to facilitate the generation of extended assay libraries for SWATH data extraction.

Package installation

To install the SwathXtend package the following commands can be executed within R.

```
> source("http://www.bioconductor.org/biocLite.R")
> biocLite("SwathXtend")
```

Typically the workspace is cleared and the SwathXtend pakcage is loaded.

```
> rm(list=ls())
> library(SwathXtend)
```

*MCP

The example data, that is included in the package, consists of six assay libraries. The libraries can be loaded using the following commands. Library format can be "PeakView" or "openSWATH" format which is in a tab-delimitated .txt or comma-delimitated .csv file. The parameter *clean* in function *readLibFile* specifies if the library to be clearned, which will be describe later.

```
> filenames <- c("Lib2.txt", "Lib3.txt")
> libfiles <- paste(system.file("files",package="SwathXtend"),filenames,sep="/")
> Lib2 <- readLibFile(libfiles[1], clean=TRUE)
> Lib3 <- readLibFile(libfiles[2], clean=TRUE)
```

Building extended assay library

To build an extended library using SwathXtend, a seed library and one or two add-on libraries are needed. The seed library is usually a local assay library which was generated with SWATH data using the same instrument and the same chromatography condition. Add-on libraries can be local archived assay libraries or external libraries downloaded from public data repositories such as SWATHAtlas(Biology IfS 2014).

Library cleaning

All candidate assay libraries were first subject to a cleaning process which removes low confident peptides and low intensity ions by user-defined thresholds. The default values for these two thresholds are 99% for peptide confidence and 5 for ion intensity. The clearning process can also opt to remove peptides with modifications for miss cleavages. The clearning process can be done separately using function *cleanLib* or as part of the library reading process as shown above.

```
> Lib2 <- cleanLib(Lib2, intensity.cutoff = 5, conf.cutoff = 0.99, nomod = FALSE, nomc = F
> Lib3 <- cleanLib(Lib3, intensity.cutoff = 5, conf.cutoff = 0.99, nomod = FALSE, nomc = F
```

Matching quality checking

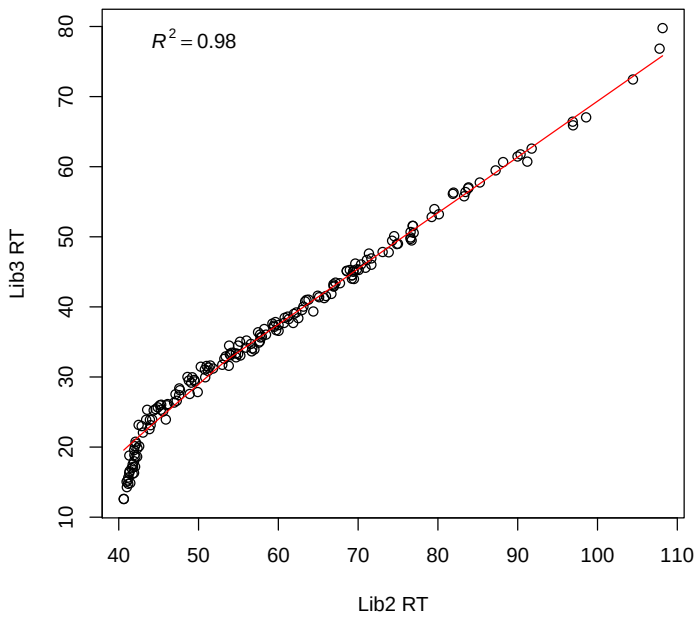
It is very important to check the matching quality between the seed and add-on libraries before building the extended library. Function *checkQuality* can be used to perform the library matching quality check based on the retention time and the relative ion intensity. Three measurements, including the retentiont time correlation, the predicted average error of the RT and the relative ion intensity correlation, will be returned.

```
> checkQuality(Lib2, Lib3)

$RT.corsqr
[1] 0.9778811

$RT.RMSE
[1] 0.9949378

$RII.cormedian
[1] 0.9
```

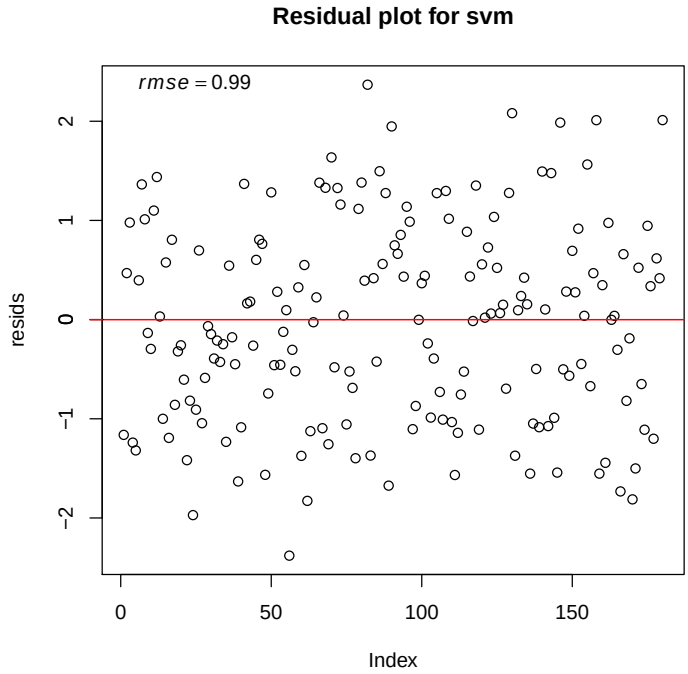
The first two outputs, RT.corsqr and RT.RMSE, represent the R^2 of the retention time correlation between the two libraries and the root of mean squared error of the RT prediction. We recommend if RT.corsqr is greater than 0.92 and the RT.RMSE less than 2 minutes, the retention time matching quality is good. The second output, The third output, RII.cormedian, repesents the median spearman correlation rou of relative ion intensity (RII) correlation. We suggest if it is greater than 0.6, these two libraries have good matching quality. We suggest the integration of libraries should be performed only when the RT and RII matching quality are good.

We can visualise the retention time correlation, prediction residual and relative ion intensity correlation using function *plotRTCor*, *plotRTResd* and *plotRIICor* respectively.

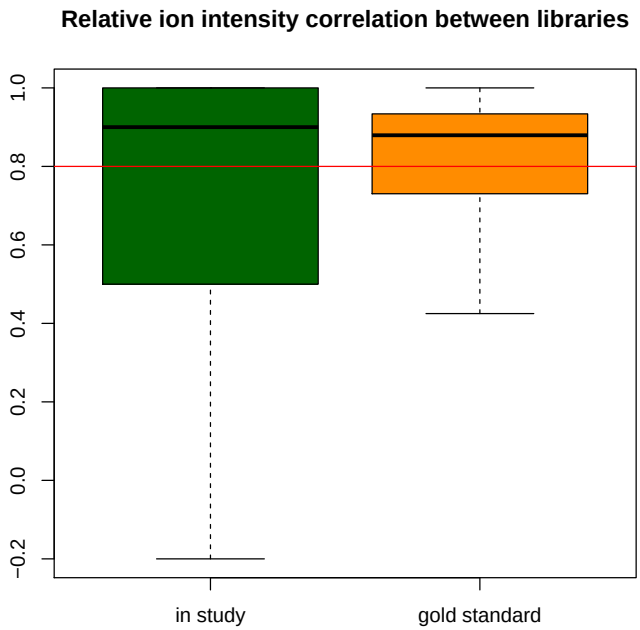
```
> plotRTCor(Lib2, Lib3, "Lib2", "Lib3")
```




```
> plotRTResd(Lib2, Lib3)
```




```
> plotRIICor(Lib2, Lib3)
```



Various statics about the two libraries can be plotted and exported into a multi-tab spreedsheet using *plotAll* function. These include barplots of the number of proteins and peptides of the seed library, add-on library and their relationship (including overlapping proteins, peptides, retention time scatter plots and spearman correlation coefficient boxplots)

```
> plotAll(Lib2, Lib3, file="allplots.xlsx")
```

Build the extended library

If the seed and add-on libraries have good matching quality (our recommendation is $R^2 > 0.92$ and $RMSE < 2minsandrelativeionintensitycorrelationSpearmanrou > 0.7$), we can generate an extended library by integrating them.

To build an extended assay library with SwathXtend, two functions are available. Function buildSpectraLibPair works for extending the seed library with on add-on library, and function buildSpectraLibTriple for extending with two add-on libraries.

```
> Lib2_3 <- buildSpectraLibPair(libfiles[2], libfiles[3], clean=T,
+                                 nomc=T, nomod=T, plot=F,
+                                 outputFormat = "PeakView",
+                                 outputFile = "Lib2_3.txt")
>
```

SwathXtend provides two methods of retention time alignment: time-based and hydrophobicity-based. If the retention time correlation between the seed and addon libraries are good (e.g., $R^2 > 0.8$), time-based method is recommended. Otherwise, hydrophobicity-based method can be tried. The hydrophobicity index for peptides can be calculated using SSRCalc(Krokhin 2006). The format of a hydrophobicity index file should include three columns, Sequence, Length and Hydrophobicity.

```
> hydroFile <- paste(system.file("files",package="SwathXtend"),"hydroIndex.txt",sep="/")
> hydro <- readLibFile(hydroFile, type="hydro")
> head(hydro)

                 Sequence Length Hydrophobicity
1                 AGIQLSPK      8          18.66
2              DASAGIQLSPK     11          22.84
3                 DLVEHVAK      8          15.74
4 DPANLPWGSSNVDIAIDSTGVFK     23          47.65
5                FVMGVNEEK      9           24.4
6       GIEEGLMTTVHSLTATQK     18          34.76
```

To build extended libraries using hydrophobicity-based retention time alignment, we can use the following command. The "method" can also be "hydrosequence" which will the combination of hydrophobicity index and the peptide sequence when building the model.

```
> Lib2_3 <- buildSpectraLibPair(libfiles[2], libfiles[3], clean=T,
+                                 nomc=T, nomod=T, plot=F,
```

```
+                                    method="hydro",
+                                    outputFormat = "PeakView",
+                                    outputFile = "Lib2_3.txt")
```

Export the library

The output of the library format can be ”PeakView” and ”OpenSwath”.

```
> outputLib(Lib2_3, filename="Lib2_3.txt", format="openswath")
```

References

Biology IfS (2014) SWATHAtlas.

Gillet LC, Navarro P, Tate S, R?st H, Selevsek N, Reiter L, Bonner R and Aebersold R (2012) Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Molecular and Cellular Proteomics 11.

Krokhin OV (2006). Sequence-specific retention calculator. Algorithm for peptide retention prediction in ion-pair RP-HPLC: application to 300-and 100-? pore size C18 sorbents. Analytical chemistry 78:7785-7795.

Wu JX, Song X, Pascovici D, Zaw T, Care N, Krisp C and Molly M. SWATH mass spectrometry performance using extended peptide MS/MS assay libraries. Molecular and Cellular Proteomics (Under review 2016).