specL: prepare peptide spectrum matches for use in targeted proteomics

Christian Trachsel <christian.trachsel@fgcz.uzh.ch>
Christian Panse <cp@fgcz.ethz.ch>
Jonas Grossmann <jg@fgcz.ethz.ch>
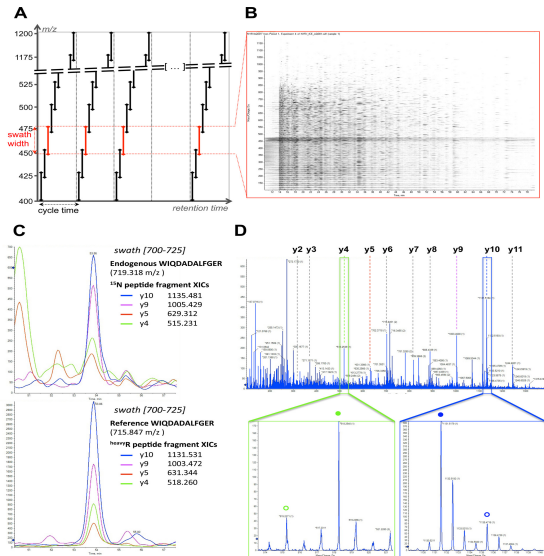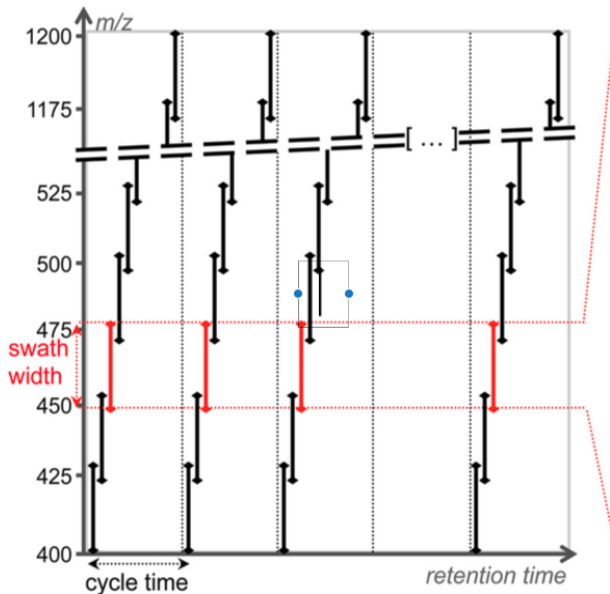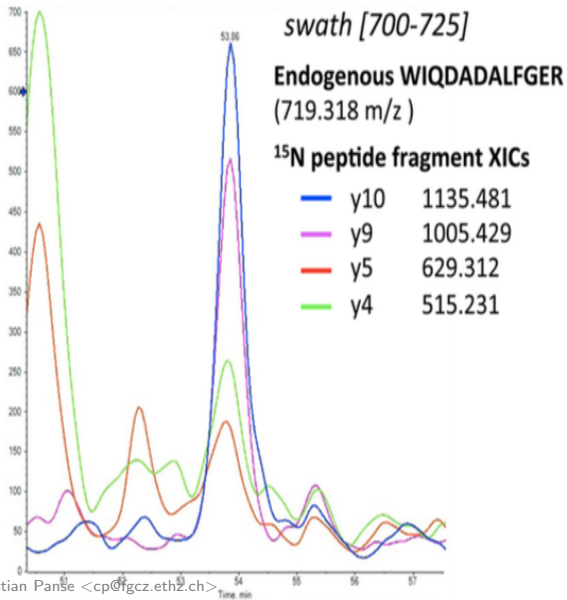
Bioconductor European Developers Workshop, 2015-01-13, EMBL

Fig. 1. **SWATH MS data-independent acquisition and targeted data analysis.** *A*, the data-independent acquisition method consists of the consecutive acquisition of high resolution, accurate mass fragment ion spectra during the entire chromatographic elution (retention time) range by repeatedly stepping through 32 discrete precursor isolation windows of 25-Da width (*black double arrows*) across the 400–1200 *m/z* range. The series of isolation windows acquired for a given precursor mass range and across the LC is referred to as a "swath" (*e.g.*, series of
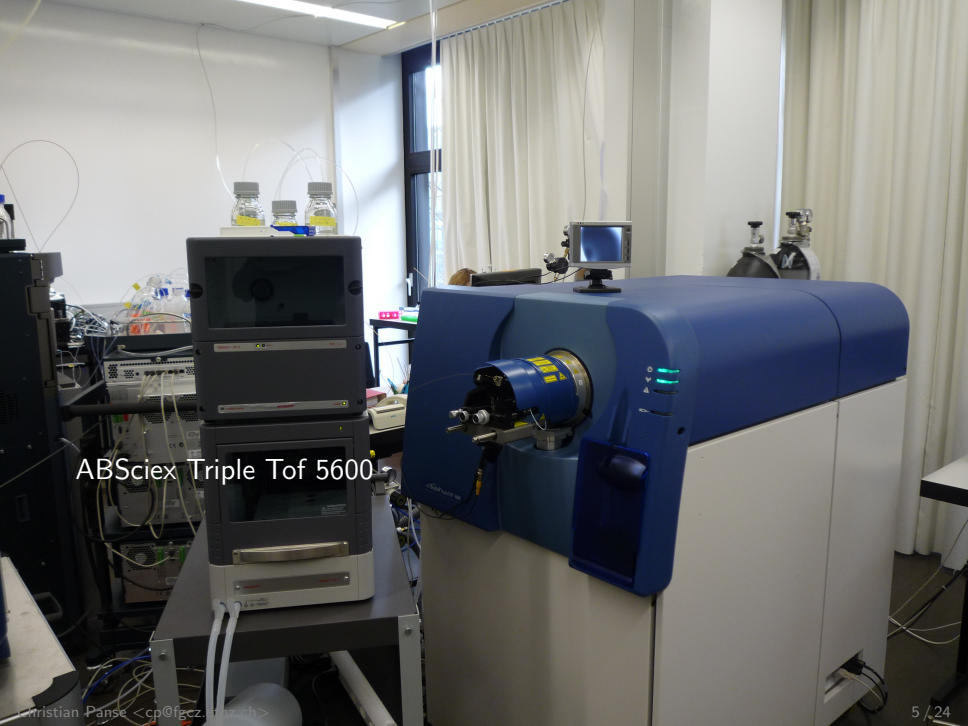
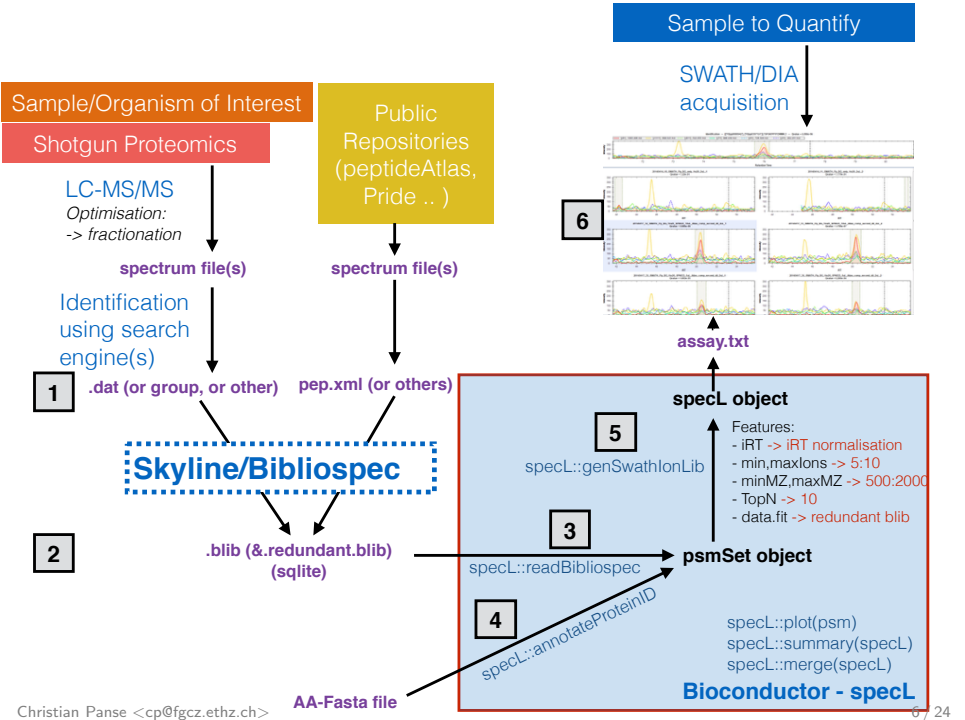*swath [700-725]*

Endogenous WIQDADALFGER
(719.318 m/z)

^{15}N peptide fragment XICs

▬	y10	1135.481
▬	y9	1005.429
▬	y5	629.312
▬	y4	515.231

ABSciex Triple Tof 5600

Get data in

reading from bibliospec (sqlite)

```sql
1 SELECT numPeaks, peakMZ, peakIntensity, peptideSeq,
       precursorCharge, precursorMZ, retentionTime,
       peptideModSeq, score, SpectrumSourceFiles.fileName
2 FROM SpectrumSourceFiles, RefSpectraPeaks, RefSpectra
3 WHERE RefSpectra.id=RefSpectraPeaks.RefSpectraID
4     and SpectrumSourceFiles.id = RefSpectra.fileID;
```

Get data in
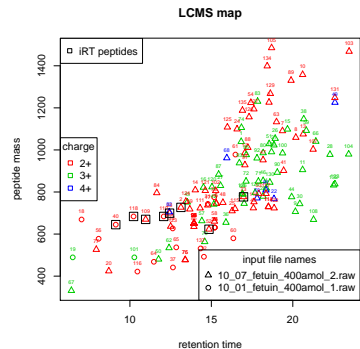reading from bibliospec (sqlite)

```
5 SELECT numPeaks, peakMZ, peakIntensity, peptideSeq,
      precursorCharge, precursorMZ, retentionTime,
      peptideModSeq, score, SpectrumSourceFiles.fileName
6 FROM SpectrumSourceFiles, RefSpectraPeaks, RefSpectra
7 WHERE RefSpectra.id=RefSpectraPeaks.RefSpectraID
8     and SpectrumSourceFiles.id = RefSpectra.fileID;
```

```
> library(specL)

> peptideStd <- read.bibliospec("peptideStd.blib")
```
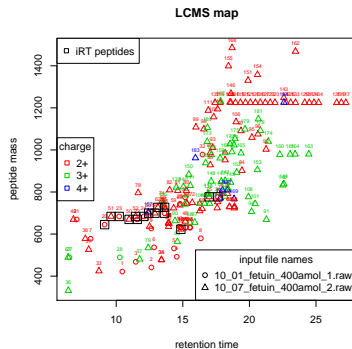
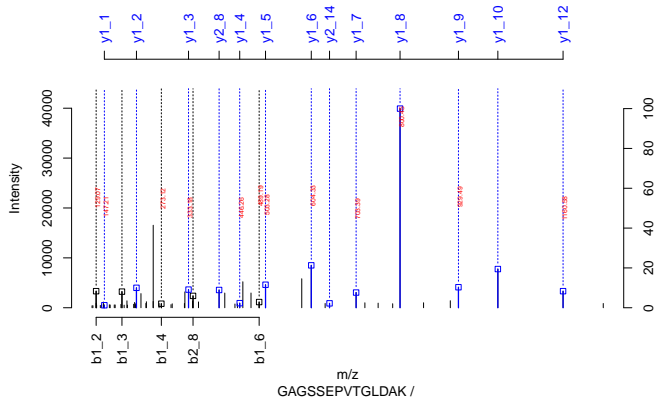# Input: Redundant plus non-redundant blib files

> `plot(peptideStd)`

> `plot(peptideStd.redundant)`

Display peptide spectrum match (psm)

`protViz::peakplot`

```
> res.plot <- plot(peptideStd[[40]], ion.axes=TRUE)
```

Ion Library Generation

Generation of the spec Library with default settings.

```
> res.ionLib <- genSwathIonLib(data=peptideStd,
+     data.fit=peptideStd.redundant)
```

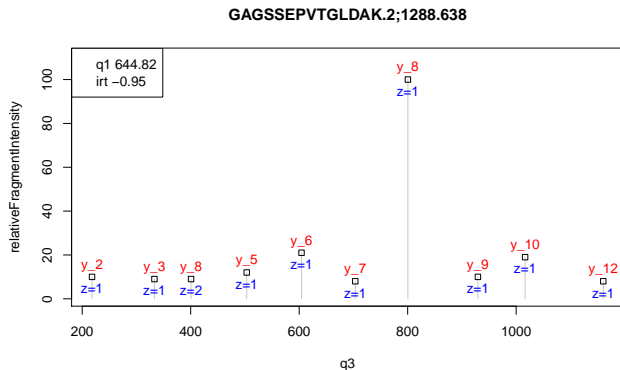# Display specL objects

```
> ionlibrary(res.ionLib)[[40]]

An "specL" object.


content:
group_id = GAGSSEPVTGLDAK.2;1288.638
peptide_sequence = GAGSSEPVTGLDAK
proteinInformation =
q1 = 644.8219
q1.in_silico = 1288.638
q3 = 800.4497 604.3285 1016.522 503.2805 929.4925 218.1497 400.7282 333.176
1160.581 703.3948
q3.in_silico = 800.4512 604.3301 1016.526 503.2824 929.4938 218.1499 400.7295
333.1769 1160.579 703.3985
decoy = NA NA NA NA NA NA NA NA NA NA
prec_z = 2
frg_type = y y y y y y y y y y
frg_nr = 8 6 10 5 9 2 8 3 12 7
frg_z = 1 1 1 1 1 1 2 1 1 1
relativeFragmentIntensity = 100 21 19 12 10 10 9 9 8 8
irt = -0.95
peptideModSeq = GAGSSEPVTGLDAK
mZ.error = 0.001514 0.00156 0.003685 0.001914 0.001318 0.000207 0.001313
0.000856 0.001846 0.003686
\ctrachse_20140910_Nuclei_diff_extraction_methods\20140910_01_fetuin_400amol_1.raw
size:
Memory usage: 4600 bytes
```

Display specL objects

```
> plot(ionlibrary(res.ionLib)[[40]])
```


GAGSSEPVTGLDAK.2;1288.638

```
> summary(res.ionLib)

Summary of a "specLSet" object.

Parameter:
mascotIonScoreCutOFF=20
proteinIDPattern=
max.mZ.Da.error=0.1
ignoreMascotIonScore=TRUE
topN=10
fragmentIonMzRange=200
  fragmentIonMzRange=2000
fragmentIonRange=2
  fragmentIonRange=100

Number of precursor (q1 and peptideModSeq) = 137
Number of unique precursor
(q1.in-silico and peptideModSeq) = 126
Number of iRT peptide(s) = 8
Number of transitions frequency:
6 7
7 6
8 11
9 10
10 103

Number of annotated precursor = 0
Number of file(s)
2

Number of precursors in Filename(s)
_methods\20140910_01_fetuin_400amol_1.raw 21
_methods\20140910_07_fetuin_400amol_2.raw 116

Misc:

Memory usage = 645600 bytes
```
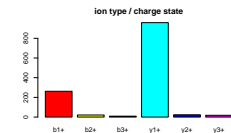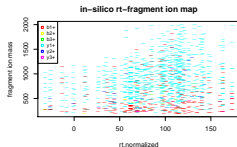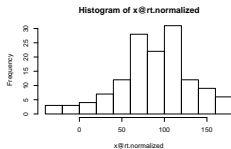
```
> op<-par(mfrow=c(4, 1), mar=c(4, 4, 3, 1));
> plot(res.ionLib)
```

iRT normalization

Problem
Retention times are not easy to transfer between different reverse phase columns or HPLC systems.

iRT normalization

Problem
Retention times are not easy to transfer between different reverse phase columns or HPLC systems.

Solution
iRT Normalization based on specific iRT / HRM peptides

Write Output
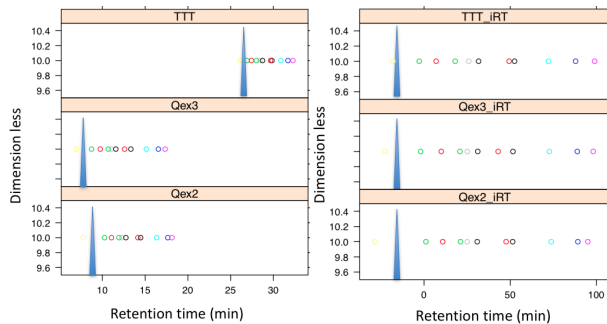to spectronaut format

```
> write.spectronaut(res.ionLib, file="spectronaut.csv")

NULL
```

Benchmark

Org	fasta			blib			runtime	
	#proteins	#peptides	file size	#specs	#mods	file size	annotate	generate
TAIR10	71032	3423196	39M	39648/118268	16554/47547	51M	79min	19sec
TAIR10	71032	3423196	39M	65018/136963	28565/64792	120M	130min	30sec
HUMAN[1]				256908/3060421		4.4G		≈5min

processed on

- ▶ 12 core XEON Server (X5650 @ 2.67GHz) running Linux Debian wheezy
- ▶ R version 3.1.1 (2014-07-10) , specL 1.1.2, and BiocParallel 1.0.0
- ▶ The default setting of BiocParallel has used eight cores.
- ▶ As FASTA we used a TAIR10 retrived from http://www.arabidopsis.org/.

---

[1] Rosenberger et al. in Scientific Data (doi:10.1038/sdata.2014.31)

Outlook
Building Consensus Spectral Libraries

```
> ionLib <-
+ genSwathIonLib(peptideStd.redundant)
> ionLib.consensus <-
+ generate.consensus(ionLib)

Combining specs having the same group_id:
10, 13, 16 -> 17
11, 14, 17 -> 18
12, 15, 18 -> 19
25, 53, 54 -> 25
102, 103, 104 -> 71
115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 1
161, 162, 163, 164 -> 100
```
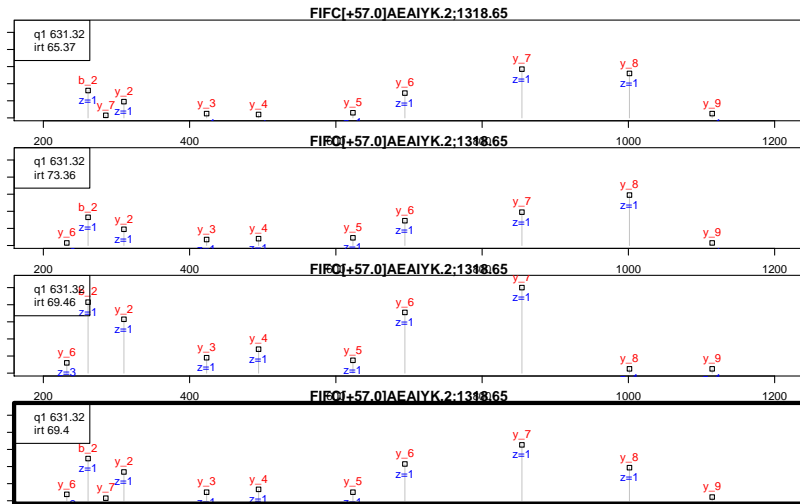
Outlook

Building Consensus Spectral Libraries

```
> op<-par(mfrow=c(4, 1), mar=c(1, 1, 1, 1));
> dump<-lapply(c(10,13,16), function(x){
+     plot(ionlibrary(ionLib)[[x]], xlim=c(200,1200)) })
> plot(ionlibrary(ionLib.consensus)[[17]], xlim=c(200,1200)); box(lwd=4)
```

pending

- importer for peakview csv format; enable `compare.specLSet(object0, object1)`
- new option for `specL::genSwathIonLib`; Exclude fragment ions from precursor `window = TRUE, FALSE`
- new option for `specL::genSwathIonLib`; Predict transitions for heavy labeled peptides using information from light peptides `predictHeavy = TRUE,FALSE, LabelFile = "fileWithHeavyAA"`
- new export function into TraML format for compatibility with OpenSWATH (pmid24727770)
- replace seqinr `read.fasta` by using Biostrings `readAAStringSet` to handle fasta files
- ~~add varMods to specL class~~
- replace mascot score by a generic score

Acknowledgement

users

Sira Echevarría Zomeño (Swiss Federal Institute of Technology in Zurich)
Tobias Kockmann (Swiss Federal Institute of Technology in Zurich)
Stephan Michalik (Ernst-Moritz-Arndt-Universität Greifswald, Germany)

CRAN & bioconductor people

Laurent Gatto (University of Cambridge, UK)

friends & colleagues

Witold E Wolski (Imagic Bildverarbeitung AG)
Lukas Reiter (BiognoSYS AG Zürich Area, Switzerland)
Natalie Selevsek (FGCZ, Switzerland)
Simon Barkow-Oesterreicher (now @ uberchord.com)

funding