Proteomics

Laurent Gatto[1]
http://cpu.sysbiol.cam.ac.uk

CSAMA – 27 June 2014

[1] lg390@cam.ac.uk

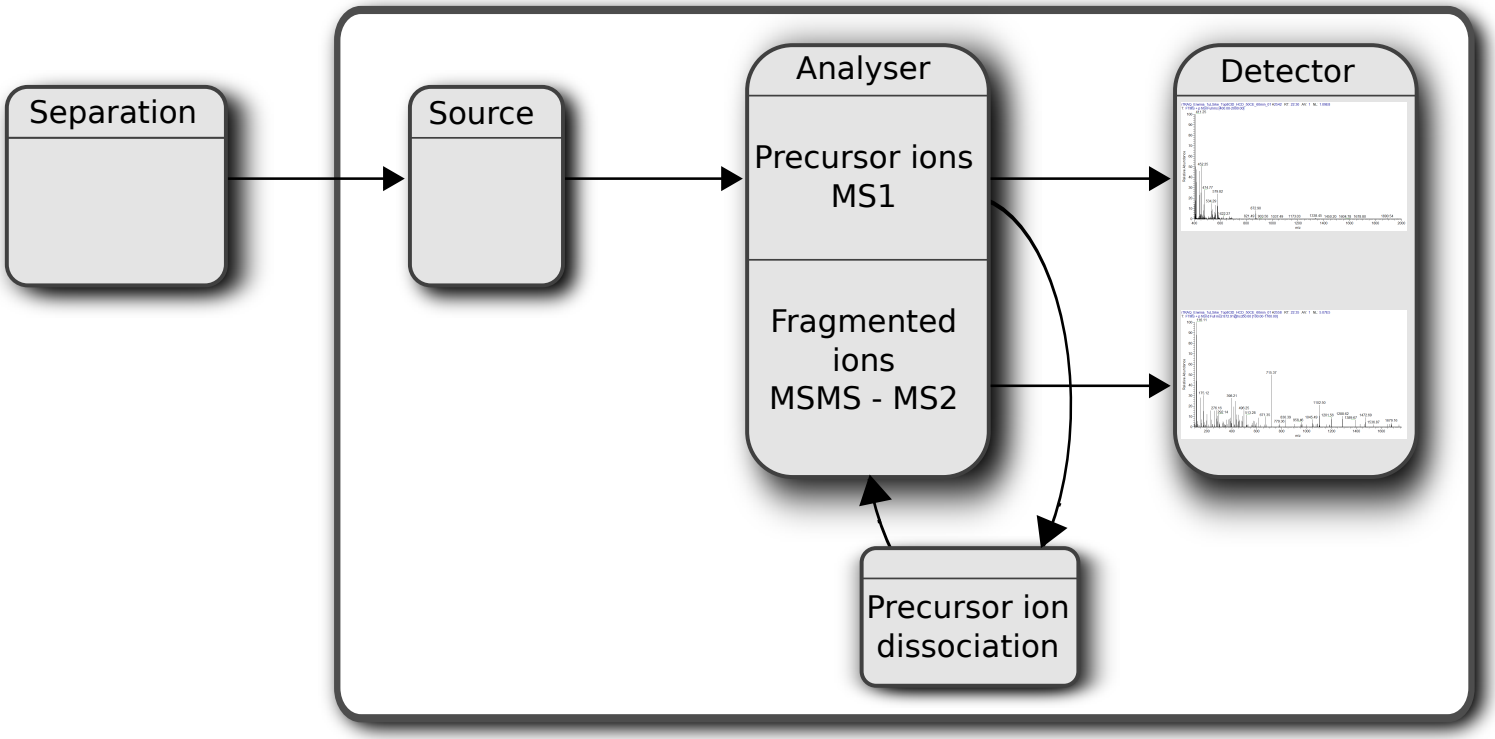
Outline

Proteomics and MS data

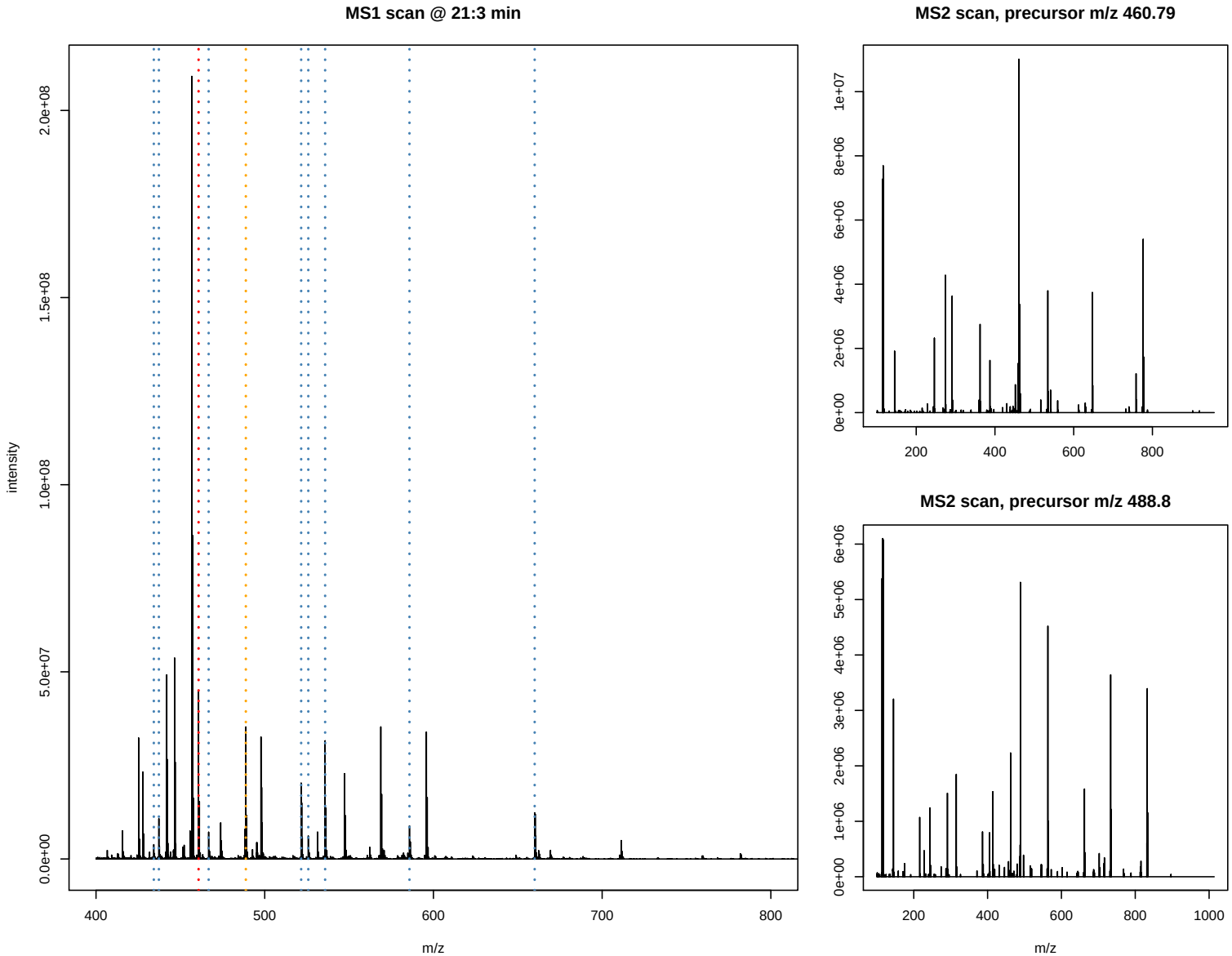
Ranges infrastructure

Application: spatial proteomics

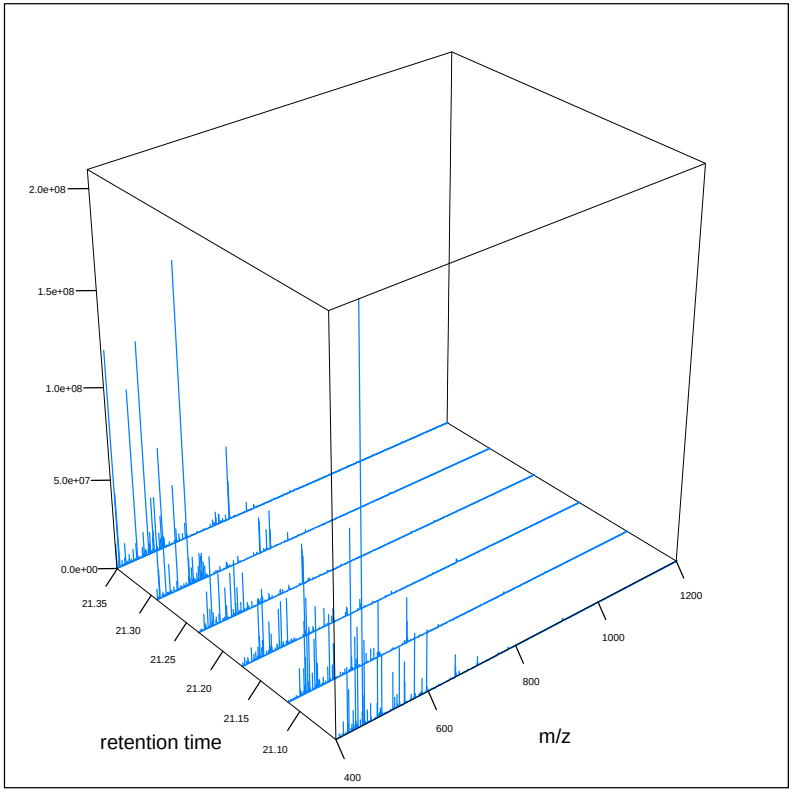
Mass-spectrometry (LC-MSMS)

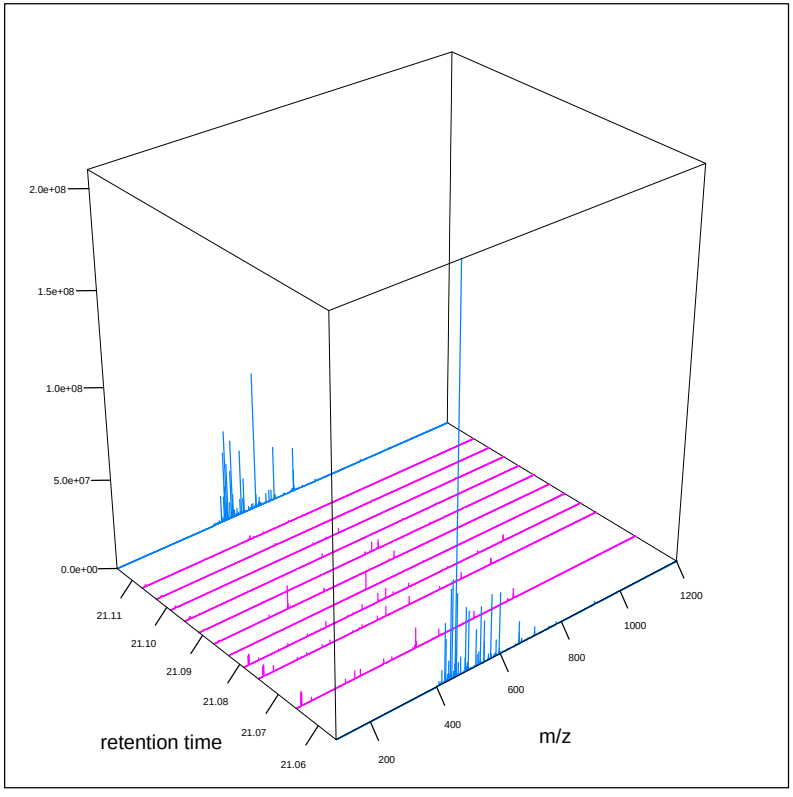


MS1 and MS2 spectra



MS1 and MS2 spectra





Proteomics data

- raw data
  - quantitation
  - identification
- protein database

Proteomics data

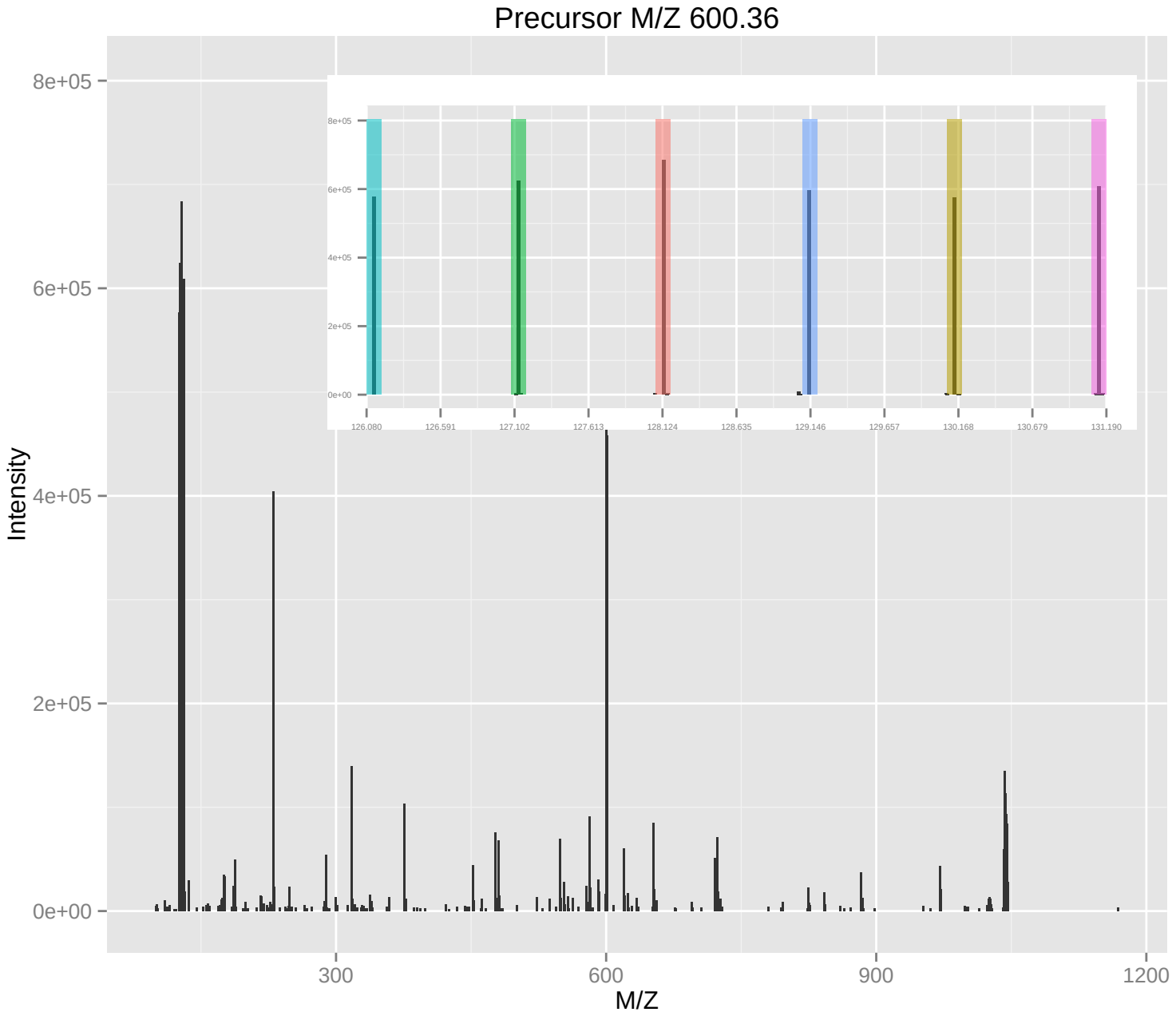
- raw data
  - quantitation
  - identification
- protein database

	Status	package
Raw (`mz*ML`)	✓	mzR
mzTab	✓	MSnbase
mgf	✓	MSnbase
mzIdentML	✓	mzID (mzR)
mzQuantML		(?mzR)

Example

```
library("MSnbase")
rx <- readMSData(f, centroided = TRUE)
rx <- addIdentificationData(rx, g)
rx <- rx[!is.na(fData(rx)$pepseq)]
plot(rx[[10]], reporters = TMT6, full=TRUE)
```

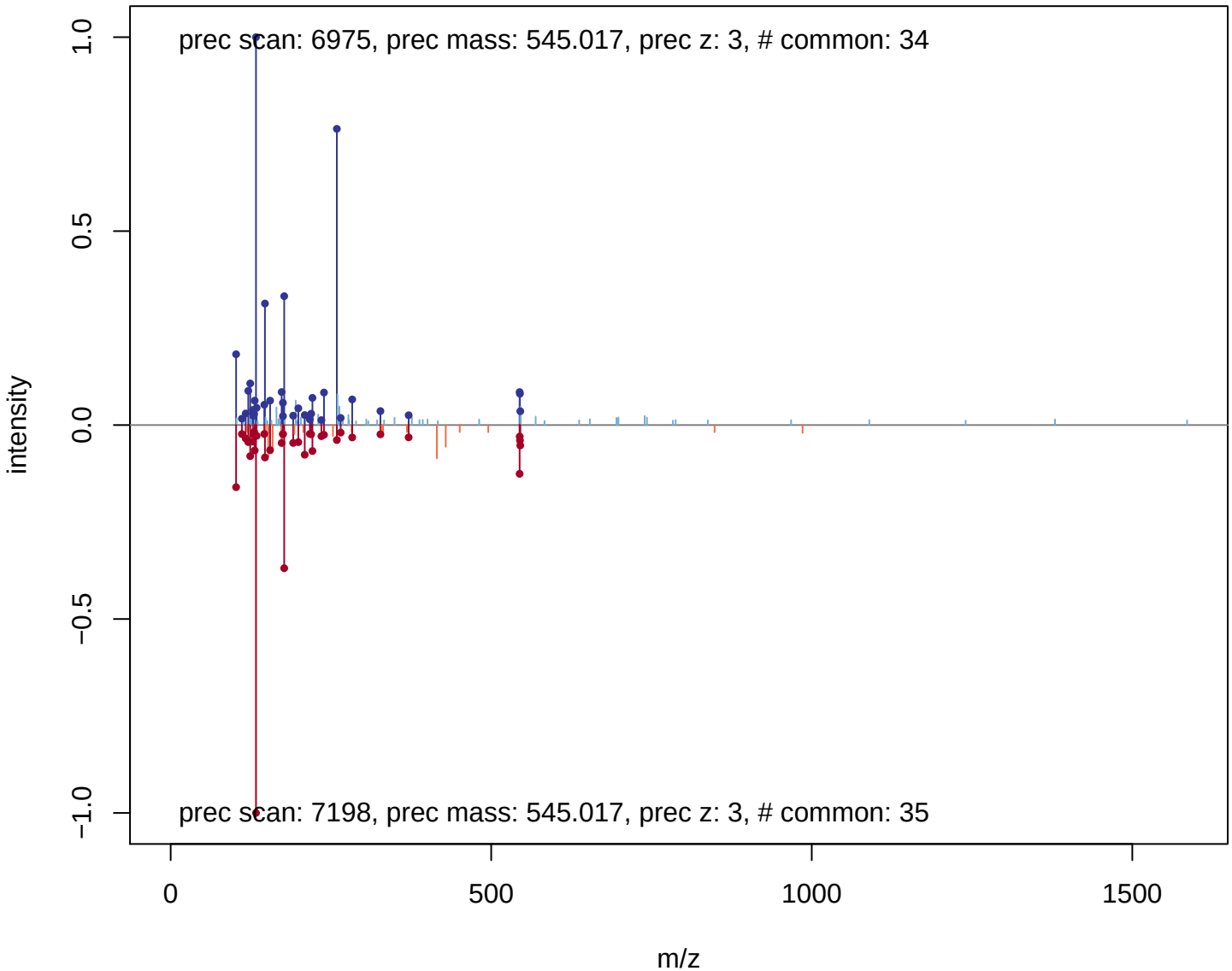
Example


Precursor M/Z 600.36
Intensity
8e+05
6e+05
4e+05
2e+05
0e+00
300
600
900
1200
M/Z
126.080
126.591
127.102
127.613
128.124
128.635
129.146
129.657
130.168
130.679
131.190

Example

```
library("MSnbase")
rx <- readMSData(f, centroided = TRUE)
rx <- addIdentificationData(rx, g)
rx <- rx[!is.na(fData(rx)$pepseq)]
plot(rx[[10]], reporters = TMT6, full=TRUE)
plot(rx[[4730]], rx[[4929]])
```

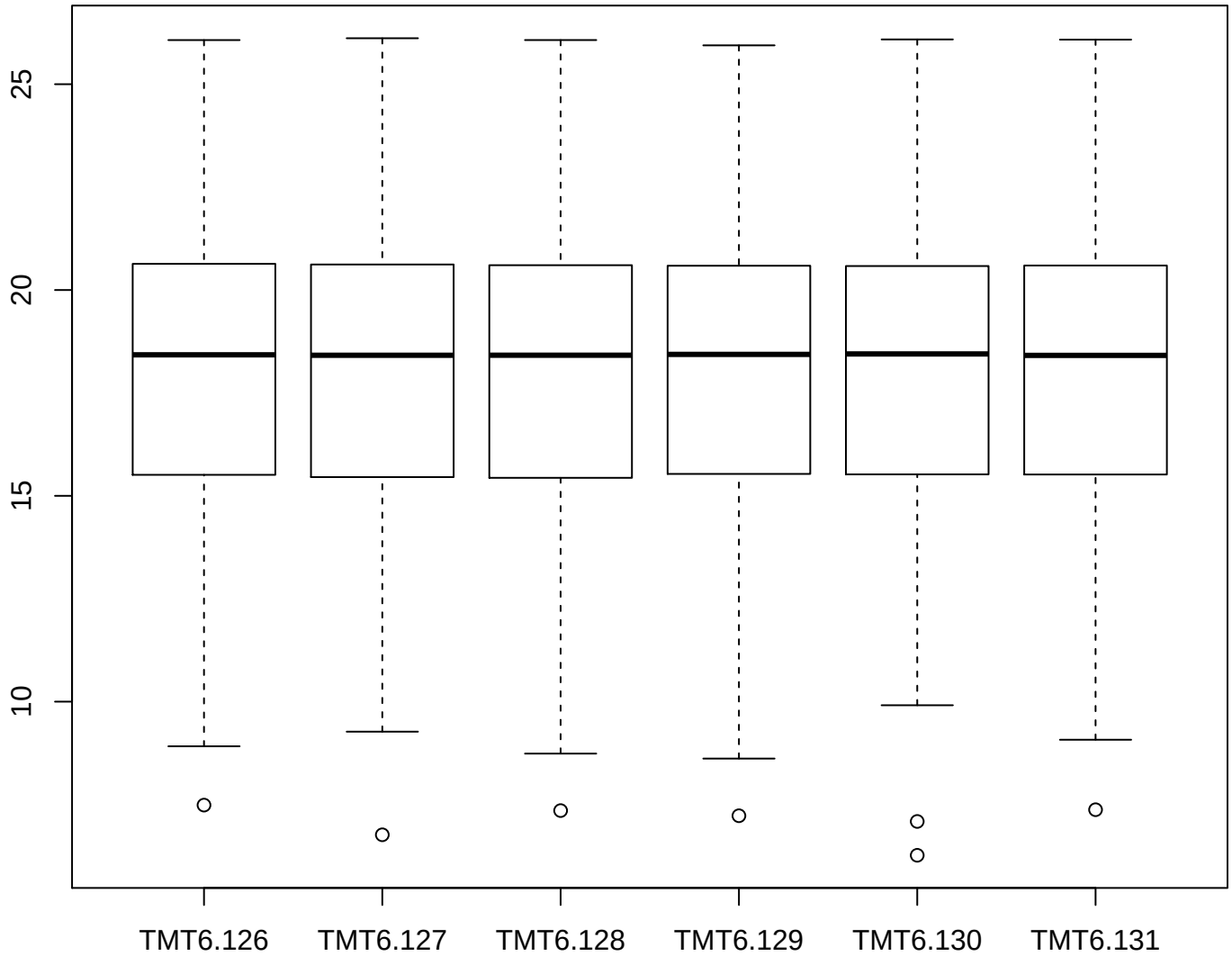
Example


prec scan: 6975, prec mass: 545.017, prec z: 3, # common: 34
prec scan: 7198, prec mass: 545.017, prec z: 3, # common: 35
intensity
1.0
0.5
0.0
−0.5
−1.0
0
500
1000
1500
m/z

Example

```
library("MSnbase")
rx <- readMSData(f, centroided = TRUE)
rx <- addIdentificationData(rx, g)
rx <- rx[!is.na(fData(rx)$pepseq)]
plot(rx[[10]], reporters = TMT6, full=TRUE)
plot(rx[[4730]], rx[[4929]])
qt <- quantify(rx, reporters = TMT6, method = "max")
## qt <- readMSnSet(f2)
nqt <- normalise(qt, method = "vsn")
boxplot(exprs(nqt))
```

Example

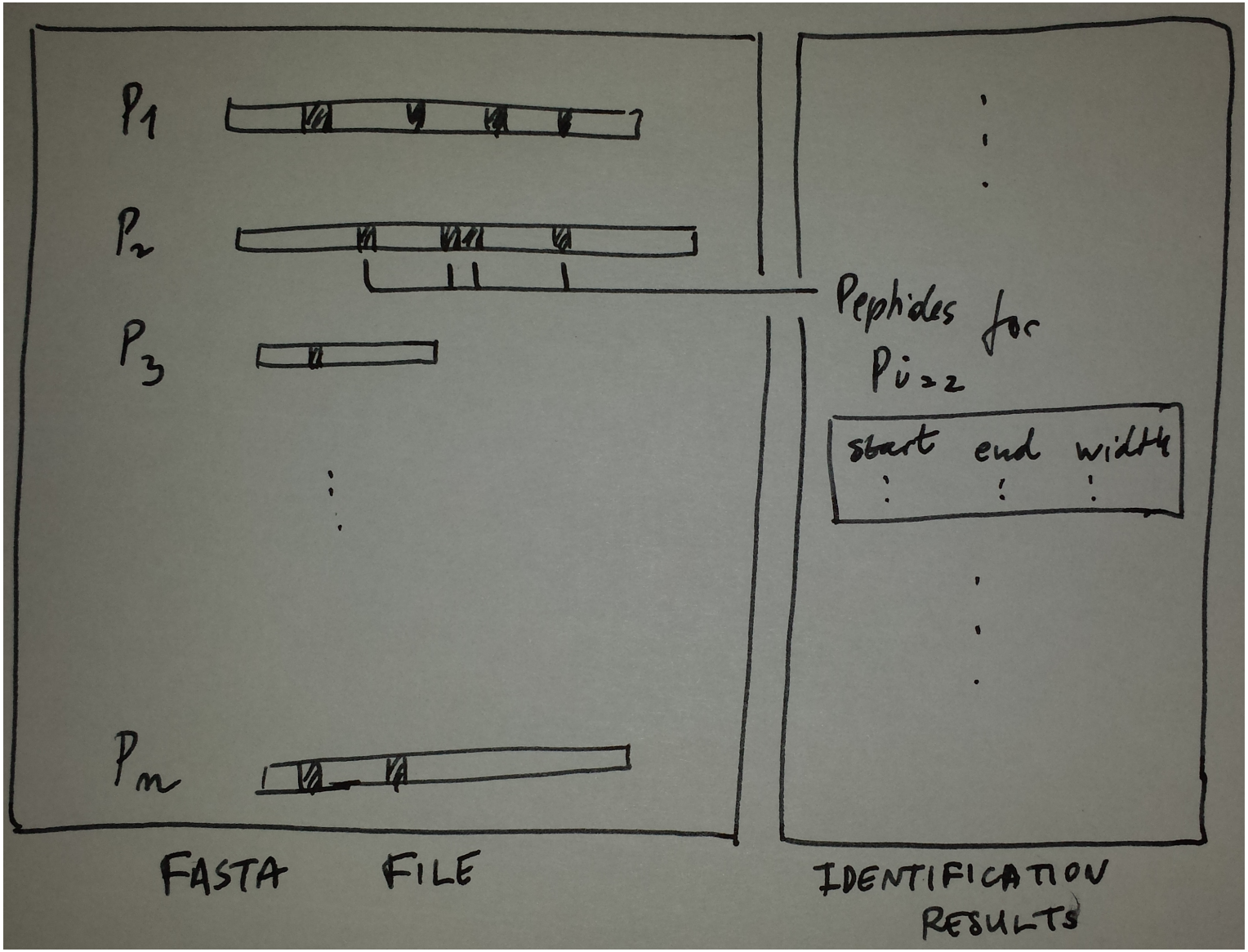

25
20
15
10
TMT6.126
TMT6.127
TMT6.128
TMT6.129
TMT6.130
TMT6.131

More

```
library("BiocInstaller")
biocLite("RforProteomics")
```

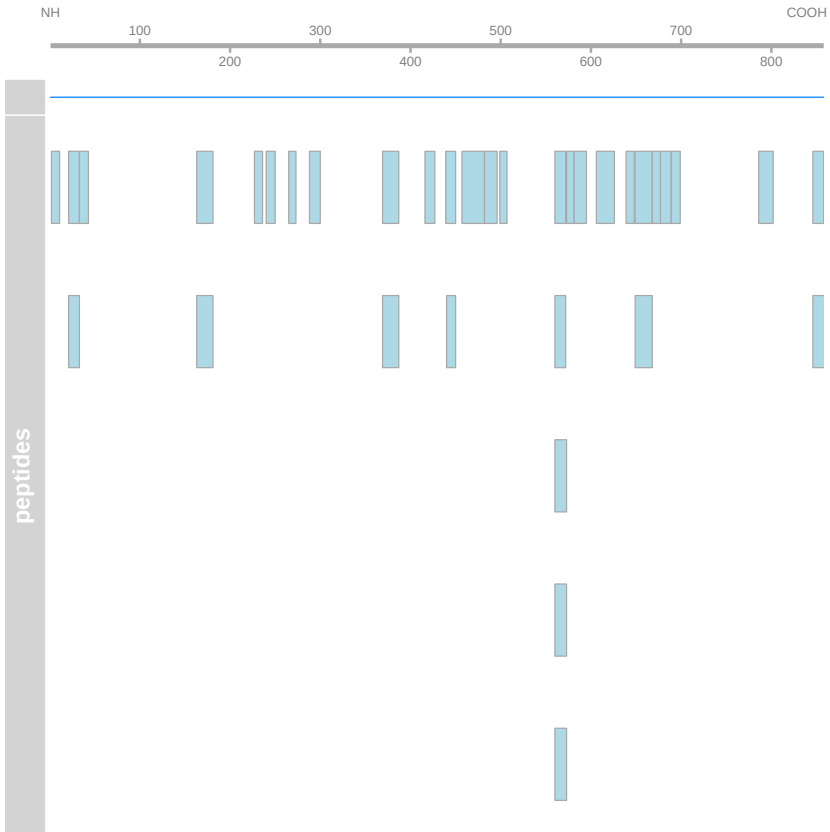
- raw data
  - quantitation
  - identification
- protein database

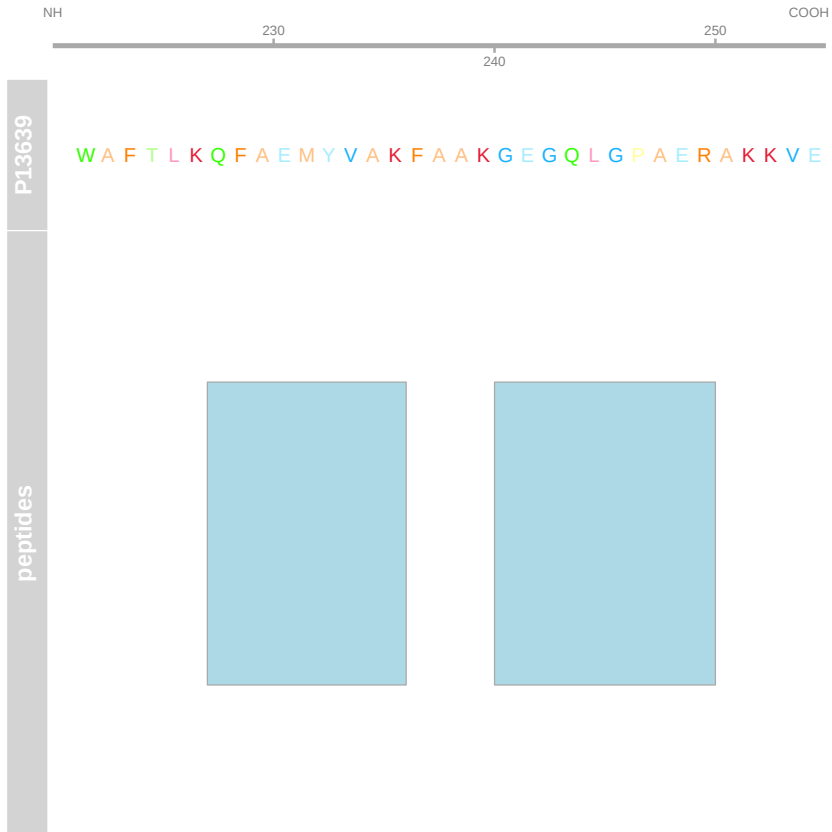
Ranges infrastructure



Pbase package

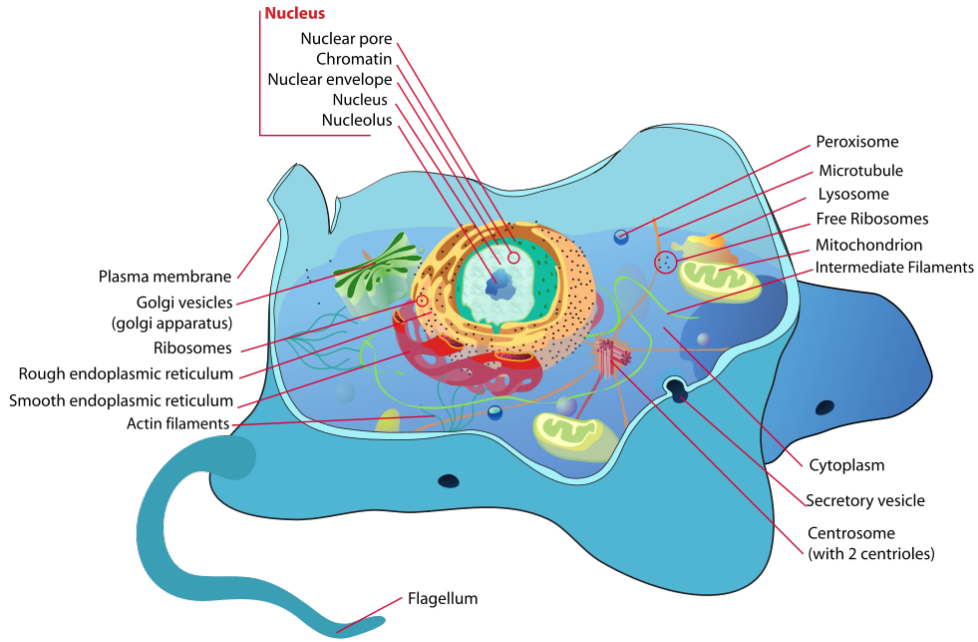
```
library("Pbase")
p <- Proteins(db)
p <- addIdentificationData(p, id)
aa(p) ## AAStringSet
pranges(p) ## IRangesList
i <- which(acols(p)[, "EntryName"] == "EF2_HUMAN")
plot(p[i])
plot(p[i], from = 155, to = 185)
```


NH
COOH
100
200
300
400
500
600
700
800
peptides


NH
COOH
230
240
250
P13639
W A F T L K Q F A E M Y V A K F A A K G E G Q L G P A E R A K K V E
peptides

Spatial proteomics

- The cellular sub-division allows cells to establish a range of distinct microenvironments, each favouring different biochemical reactions and interactions and, therefore, allowing each compartment to fulfil a particular functional role.
- Localisation and sequestration of proteins within subcellular niches is a fundamental mechanism for the post-translational regulation of protein function.

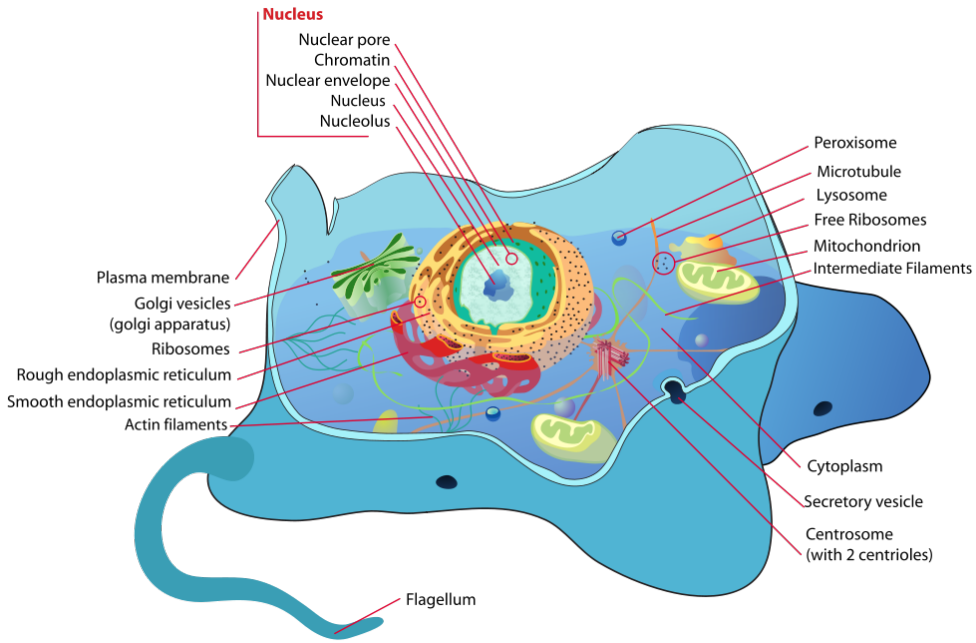


Spatial proteomics is the systematic study of protein localisations.

Spatial proteomics

Disruption of the targeting/trafficking process alters proper sub-cellular localisation, which in turn perturb the cellular functions of the proteins.

- Abnormal protein localisation leading to the loss of functional effects in diseases (Laurila et al. 2009)
- Disruption of the nuclear/cytoplasmic transport (nuclear pores) have been detected in many types of carcinoma cells (Kau et al. 2004).



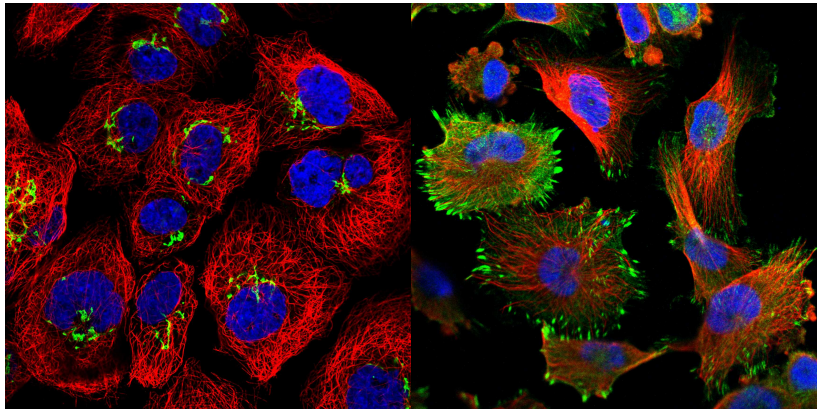


Figure: **Immunofluorescence**: ZFPL1, Golgi (left) and FHL2, mainly localized to actin filaments and focal adhesion sites. Also detected in the nucleus (right). (from the Human Protein Atlas)

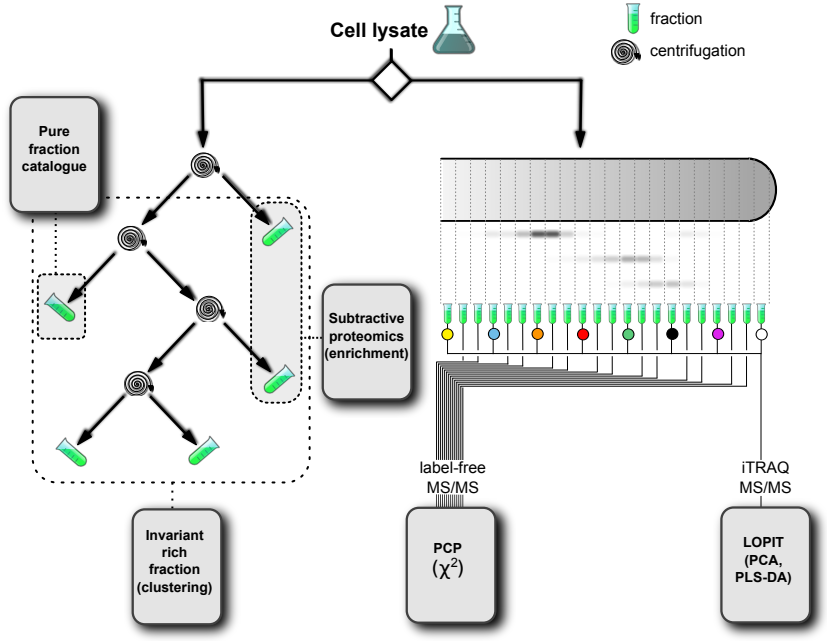
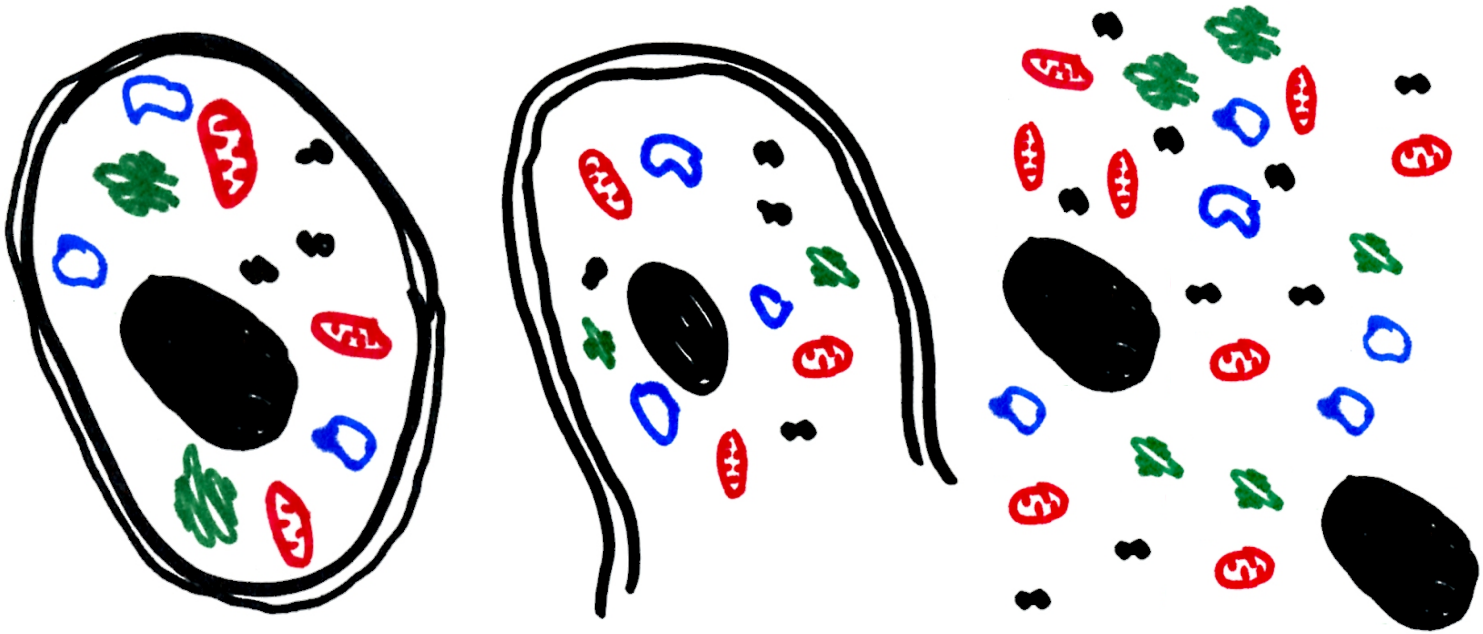


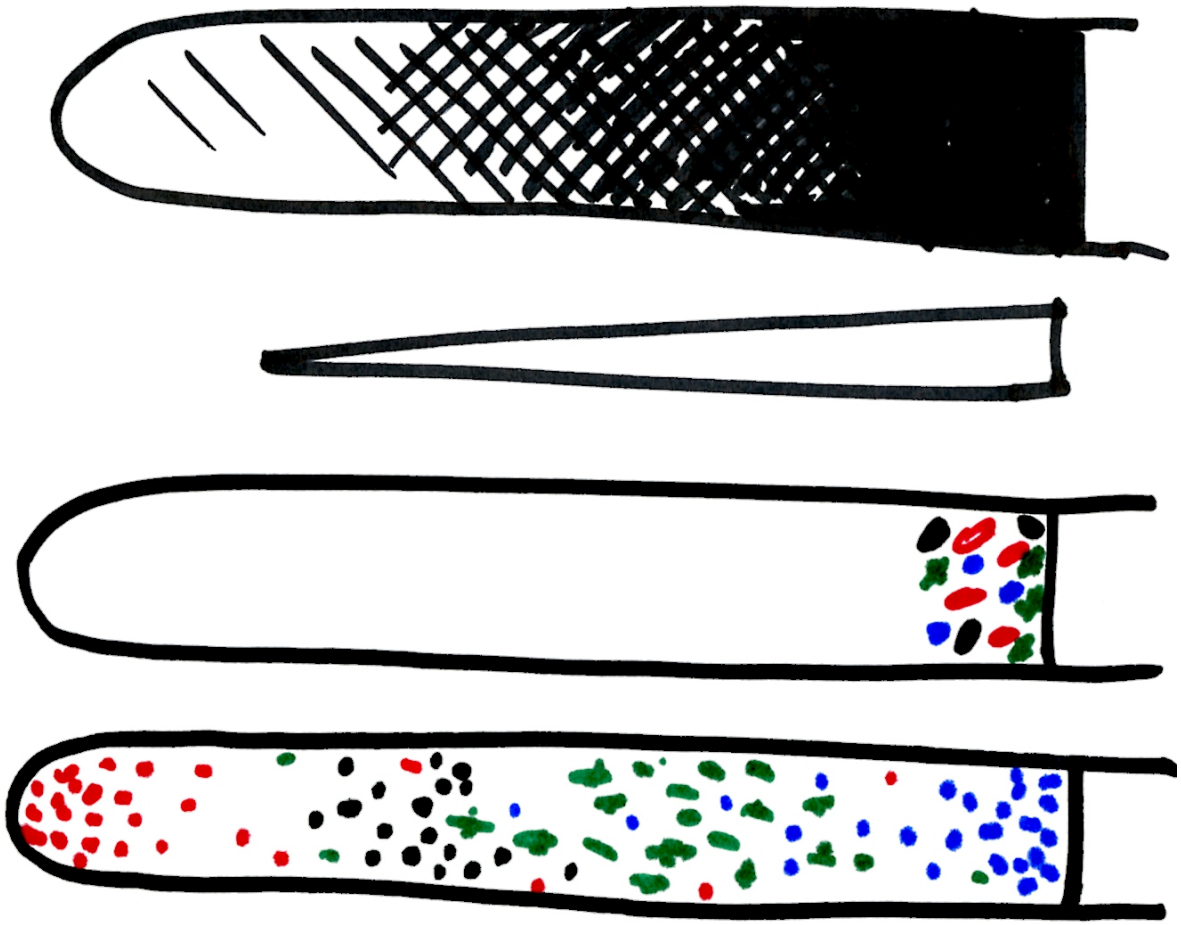
Figure: Mass spectrometry-based approaches based on density gradient subcellular fractionation.

Cell membrane lysis

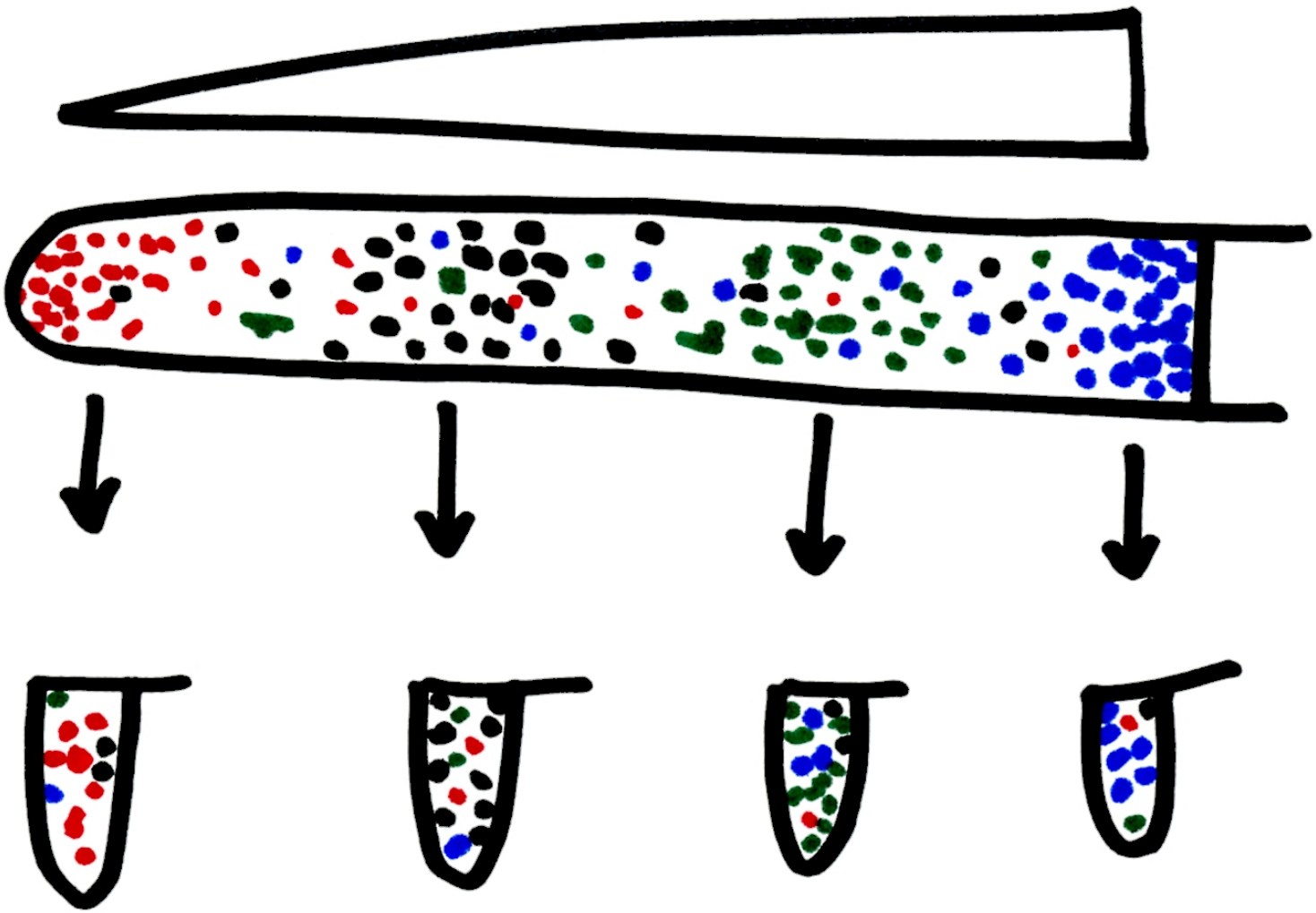
Mechanical or buffer-induced lysis of the plasma membrane with minimal disruption to intracellular organelles followed by subcellular fractionation.



Density gradient separation



Quantitation by LC-MSMS



Data

	Fraction_1	Fraction_2	...	Fraction_m	markers
p_1	$q_{1,1}$	$q_{1,2}$	...	$q_{1,\,m}$	unknown
p_2	$q_{2,1}$	$q_{2,2}$	...	$q_{2,\,m}$	loc_1
p_3	$q_{3,1}$	$q_{3,2}$	...	$q_{3,\,m}$	unknown
p_4	$q_{4,1}$	$q_{4,2}$	...	$q_{4,\,m}$	loc_k
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
p_n	$q_{n,1}$	$q_{n,2}$	...	$q_{n,\,m}$	unknown

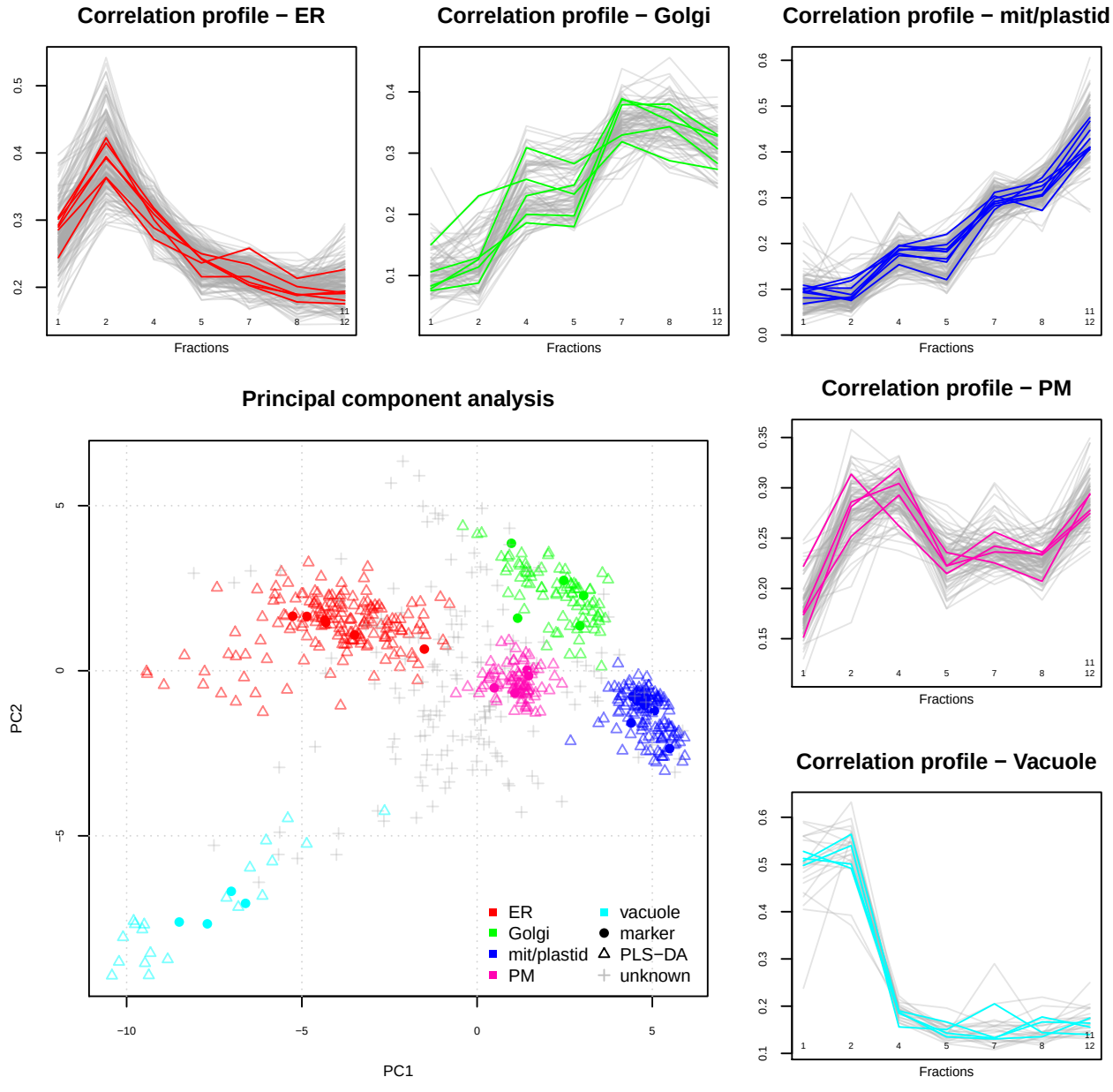


Figure: From Gatto et al. (2010), data from Dunkley et al. (2006).

2009 vs 2013

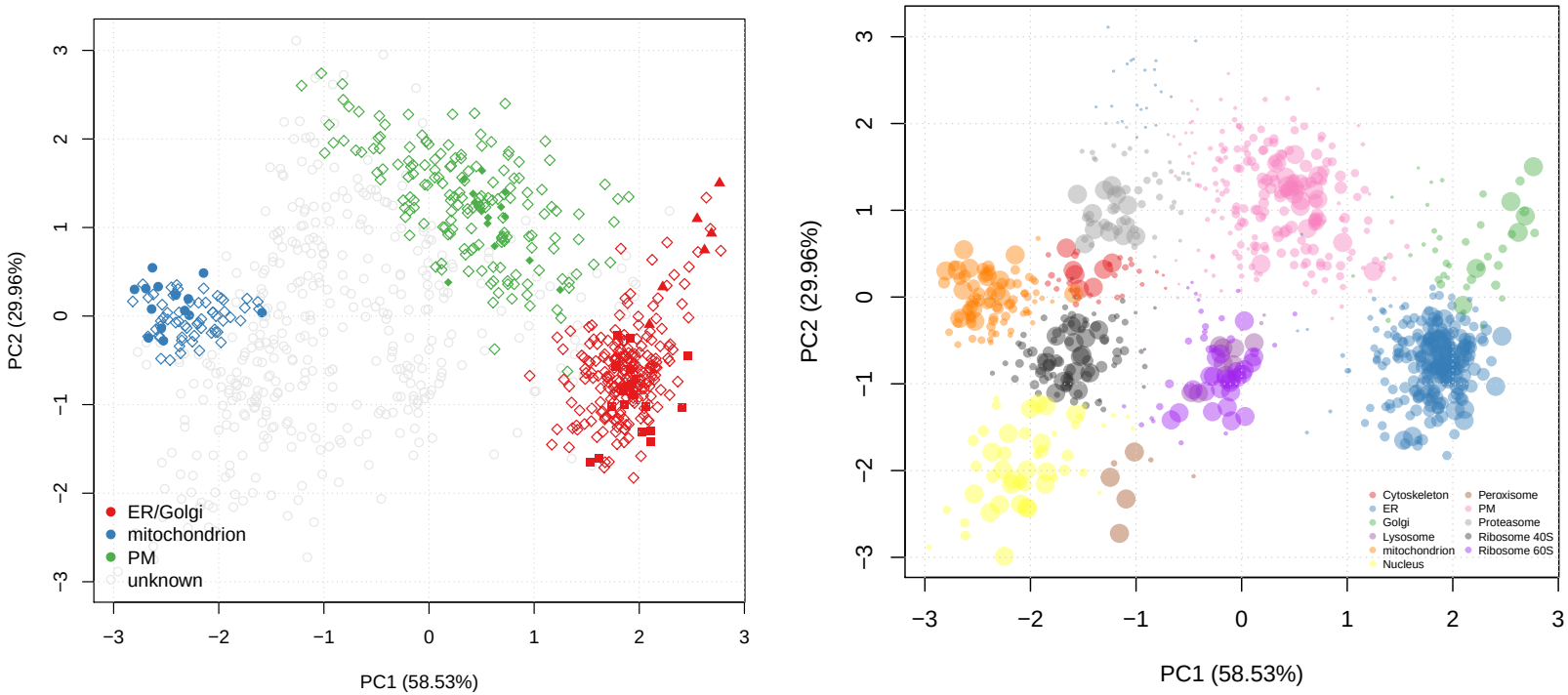
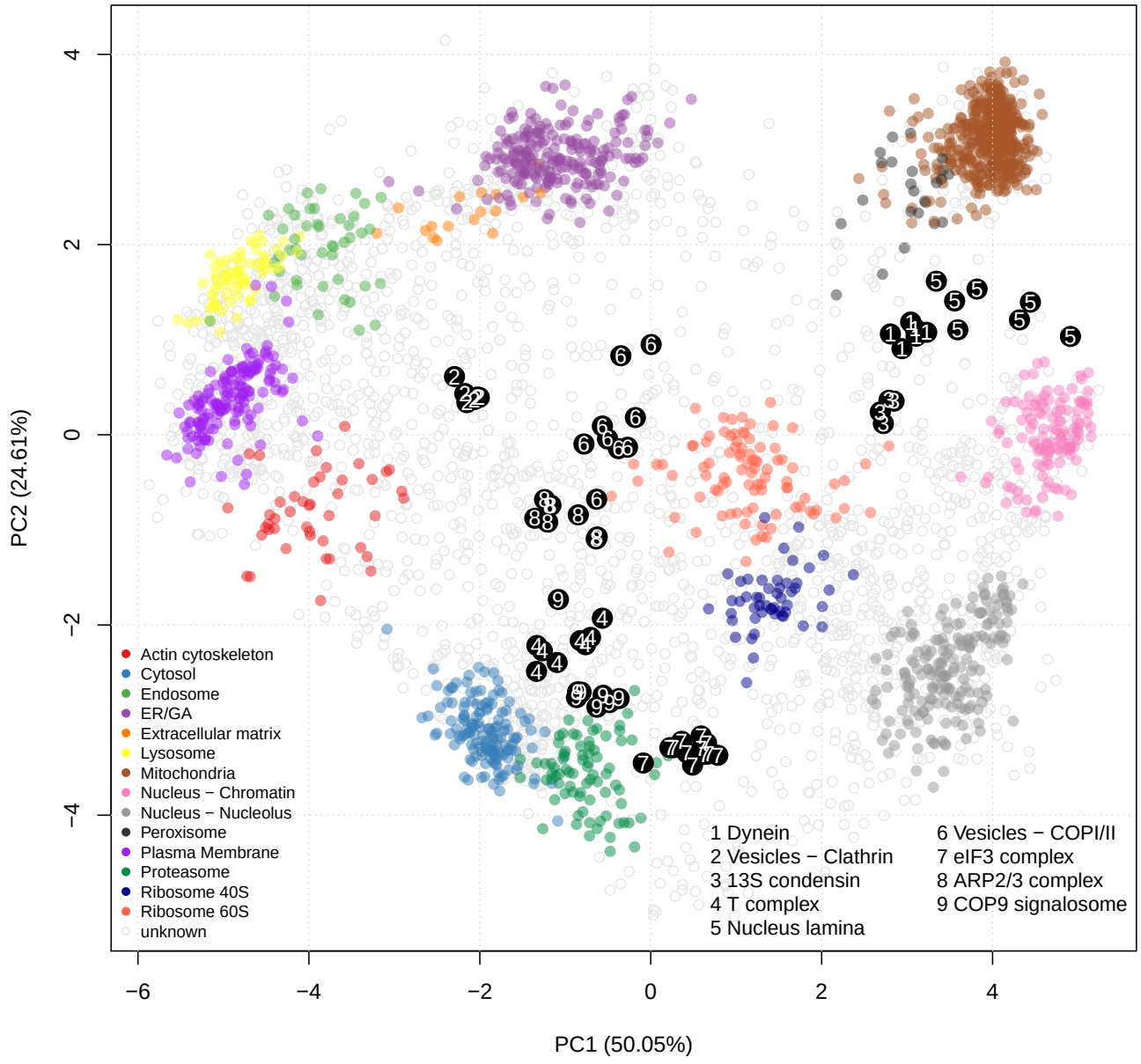


Figure: `pRoloc` package. Semi-supervised approach Breckels et al. (2013). Data from Tan et al (2009).


PC2 (24.61%)
PC1 (50.05%)
4
2
0
−2
−4
−6
−4
−2
0
2
4
Actin cytoskeleton
Cytosol
Endosome
ER/GA
Extracellular matrix
Lysosome
Mitochondria
Nucleus − Chromatin
Nucleus − Nucleolus
Peroxisome
Plasma Membrane
Proteasome
Ribosome 40S
Ribosome 60S
unknown
1 Dynein
2 Vesicles − Clathrin
3 13S condensin
4 T complex
5 Nucleus lamina
6 Vesicles − COPI/II
7 eIF3 complex
8 ARP2/3 complex
9 COP9 signalosome

Dynamic

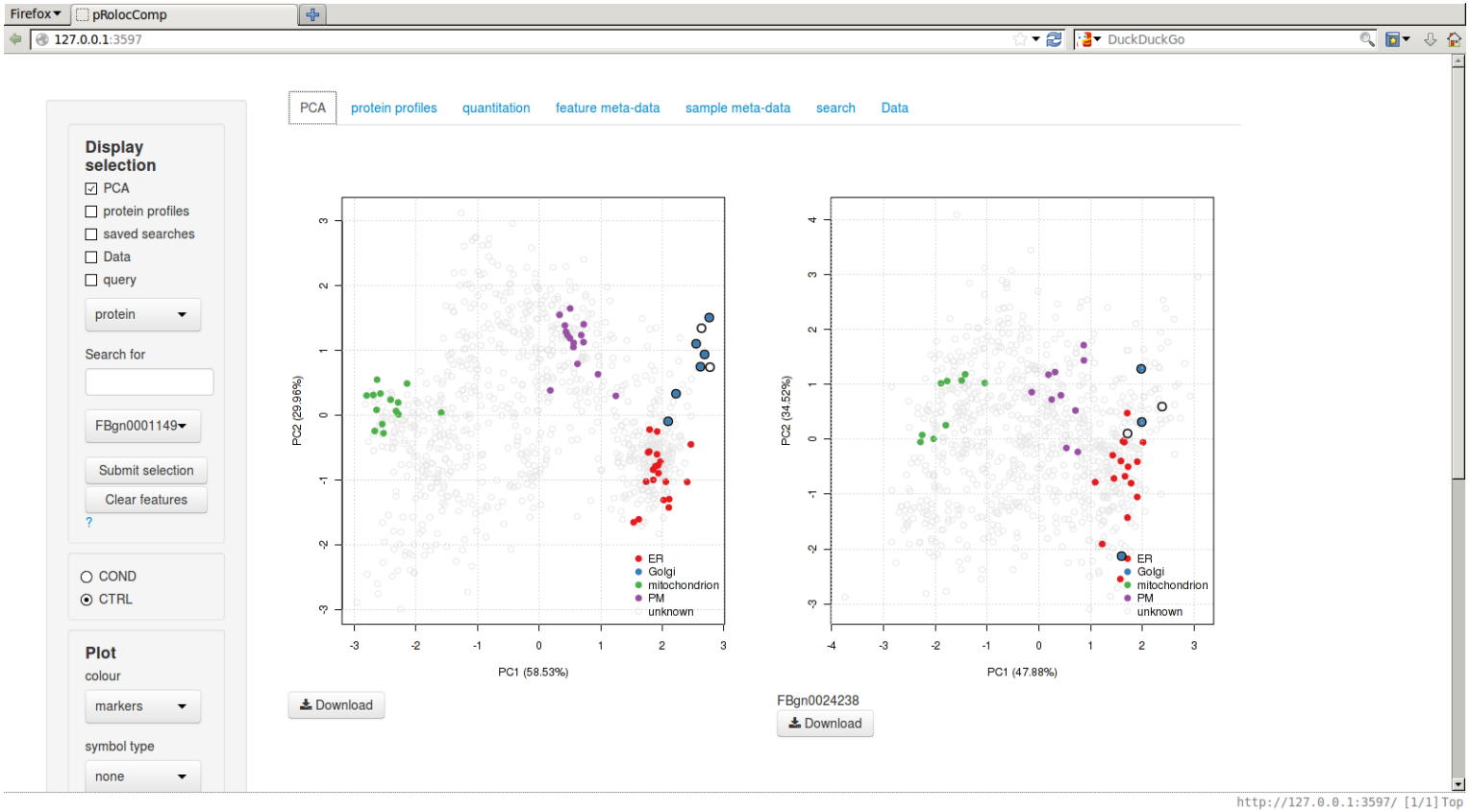


Figure: `pRolocGUI` package.

Dual localisation

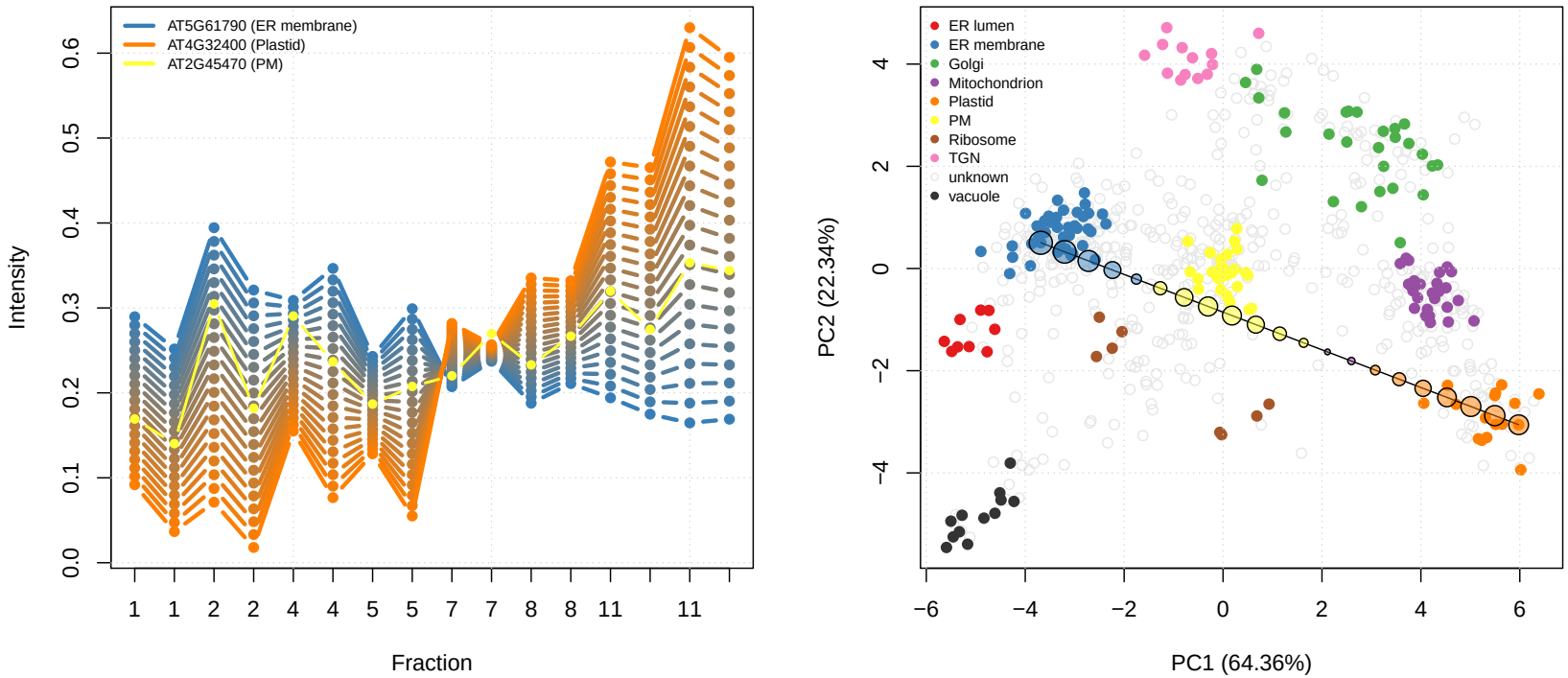


Figure: Proteins may be present simultaneously in several organelles (**dual localisation**, trafficking) vs. *no man's land*. (Gatto et al. 2014)

Acknowledgement

CPU

- Lisa Breckels
- Sebastien Gibb
- Thomas Naake
- Kathryn Lilley (CCP)

Computational Proteomics Unit
Cambridge Centre for Proteomics
Cambridge System Biology Centre
Department of Biochemistry
University of Cambridge
http://cpu.sysbiol.cam.ac.uk
@lgatt0

Software Sustainability Institute
http://software.ac.uk