

Interactive visualisation of spatial proteomics data

L. Breckels and L. Gatto

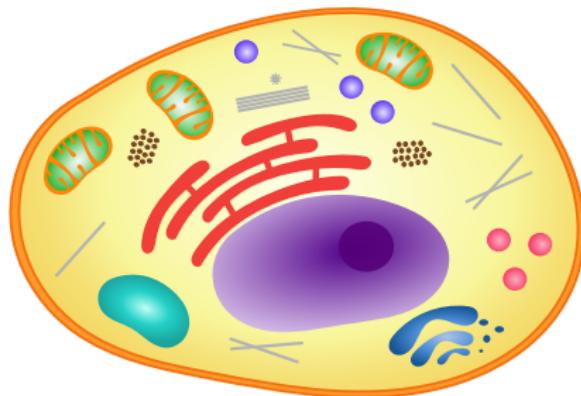
University of Cambridge, UK | de Duve Institute, Belgium
lms79@cam.ac.uk | laurent.gatto@uclouvain.be

Slides: DOI [10.5281/zenodo.1927886](https://doi.org/10.5281/zenodo.1927886)

<https://zenodo.org/record/1927886>

EuroBioC 2018, 6-7th December 2018

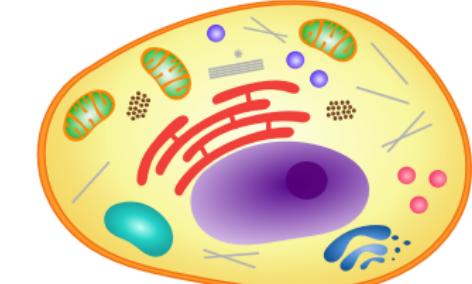
Cell organisation



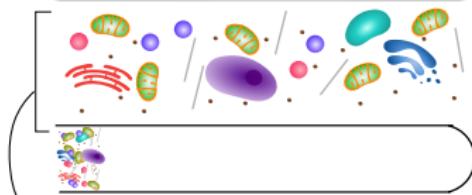
- ▶ Proteins are spatially organised according to function
- ▶ Significant correlation between disease classes and sub-cellular localisations
- ▶ Abnormal protein localisation leading to the loss of functional effects in diseases

Spatial proteomics is the systematic study of protein localisations.

Localisation maps from quantitative proteomics

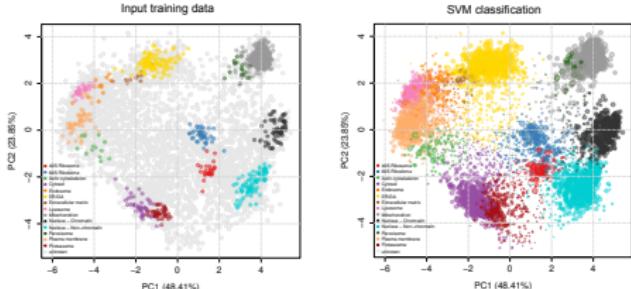
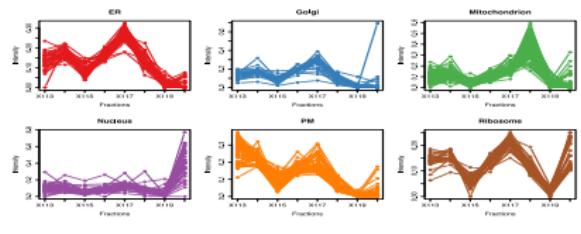


Cell lysis



- Fractionation/centrifugation
- Quantitation/identification
- Mass-spectrometry

	Fraction ₁	Fraction ₂	...	Fraction _m	markers
prot ₁	q _{1,1}	q _{1,2}	...	q _{1, m}	unknown
prot ₂	q _{2,1}	q _{2,2}	...	q _{2, m}	organelle ₁
prot ₃	q _{3,1}	q _{3,2}	...	q _{3, m}	unknown
prot ₄	q _{4,1}	q _{4,2}	...	q _{4, m}	organelle ₂
...
prot _i	q _{i,1}	q _{i,2}	...	q _{i, m}	...
...
prot _n	q _{n,1}	q _{n,2}	...	q _{n, m}	unknown



Goal: to pinpoint the sub-cellular localisation of proteins

A synergistic framework

Data analysis and machine learning

- ▶ MSnbase and pRoloc - dedicated packages providing robust and reproducible tools

Data visualisation

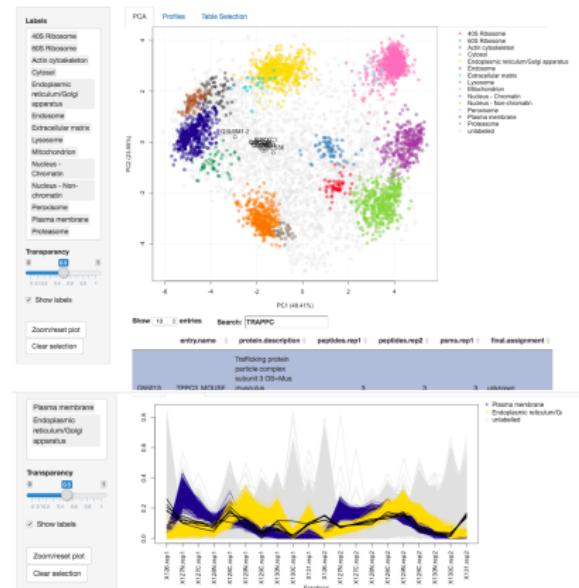
- ▶ pRolocGUI - interactive visualisation and data mining
- ▶ User driven: essential to transfer of findings easily between programmatic and graphical interfaces

The pRolocGUI package

1. The *main app* - for exploratory data analysis and features a searchable, clickable and zoomable PCA plot
2. The *comparison app* - for examining two replicate experiments, or two experiments from different conditions etc
3. The *aggregation app* - allows one to compare the effects of feature aggregation, for example, when combining peptides to proteins
4. The *classify app* - useful for viewing the sub-cellular class predictions output from a supervised machine learning analysis

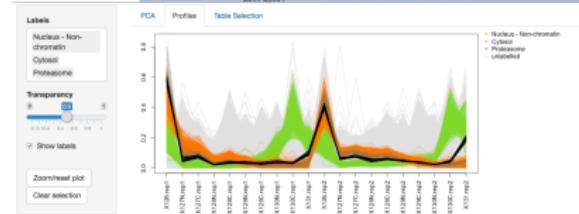
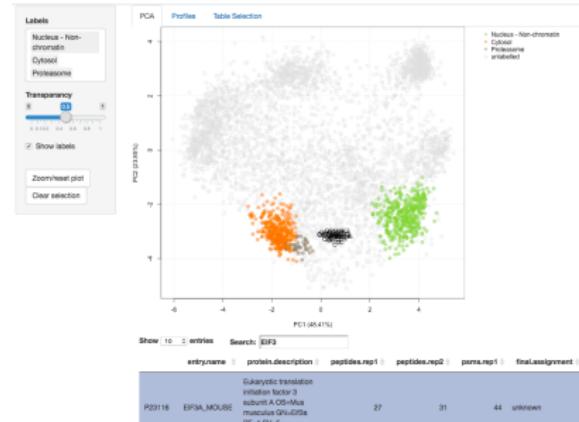
Main app - Example 1: TRAPP complex

- ▶ TRAPP complex (trafficking protein particle complex)
- ▶ Multi-localised: involved in trafficking between ER/GA and PM
- ▶ Note: one protein sits away, investigation shows it has a separate role as an activator of NF-kappa-B through increased phosphorylation



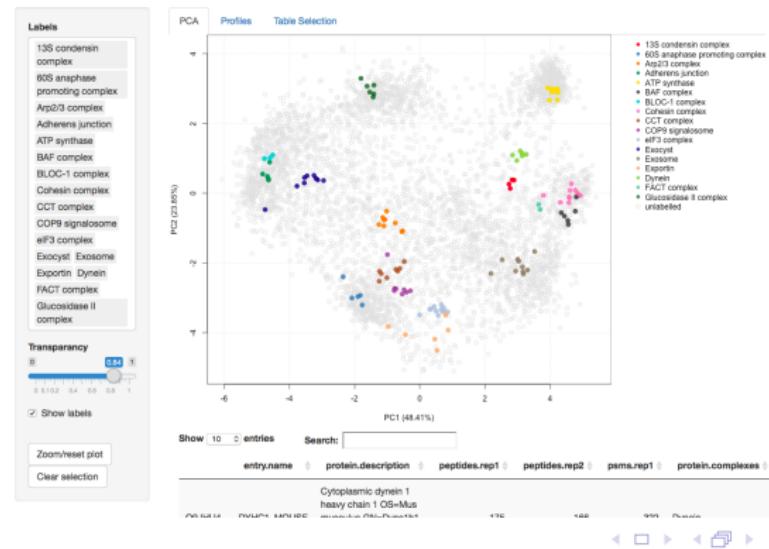
Main app - Example 2: EIF3 complex

- ▶ Eukaryotic initiation factor 3 (eIF3) is a multiprotein complex that functions during the initiation phase of translation
 - ▶ eIF3 stimulates nearly all steps of translation initiation
 - ▶ location and profiles highly correlated -> well characterised complex



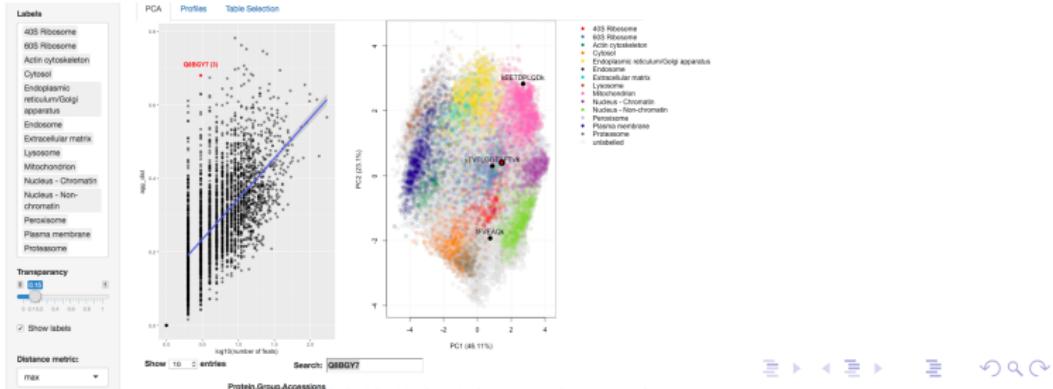
Main app - Example 3: Visualisation of many complexes

- ▶ The application is well suited to visualisation of multiple organelles/complexes/sub-cellular niches
- ▶ The handy drop down menu allows one to easily click on/off compartments - avoid the regeneration of multiple plots etc.



Aggregation app - Example 4: Protein FAM210A

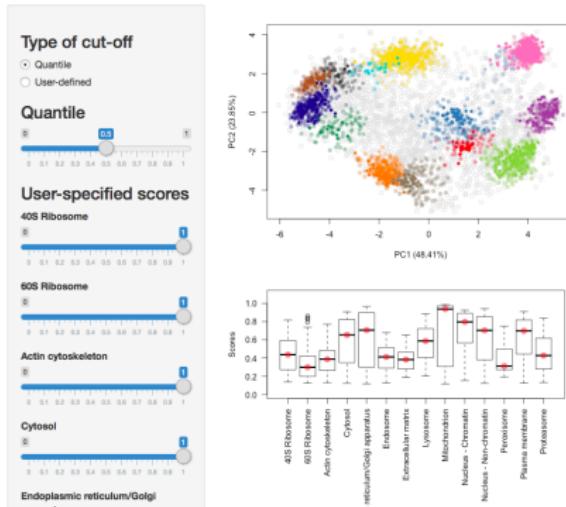
- ▶ This is an interesting protein found from looking at the outliers of the aggvar plot on the LHS
- ▶ 3 PSMs available for quantitation all with very different localisations on the plot
- ▶ Not much was known about this protein's function until this year (see <http://www.pnas.org/content/115/16/E3759>)
- ▶ It is expressed in muscle mitochondria and cytoplasm
- ▶ Thought to strongly influence the structure and strength of both muscle and bone and a new target for osteoporosis



Classify app - Example 5: Visualisation of ML results

- ▶ Load the SVM classification results from the mouse stem cell dataset
- ▶ Explain how users set thresholds
- ▶ See the effects on classification of setting thresholds

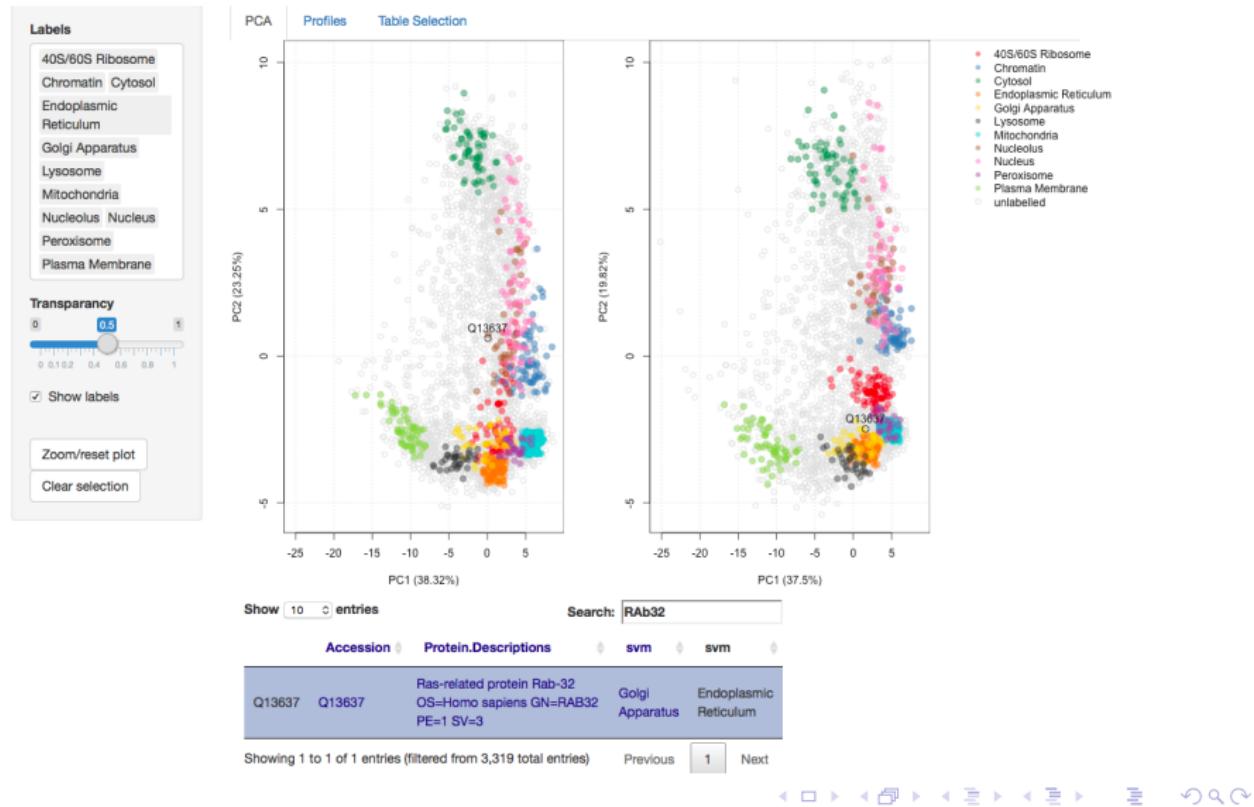
Classification results



Compare app - Example 6: Human THP data (a monocytic cancer cell line)

- ▶ hyperLOPIT experiment on a THP1 cells, 3 x 20 fractions (two 10 plex TMT each replicate)
- ▶ Unstimulated (LHS) and cells treated with Lipopolysaccharide (LPS) (12 hrs stimulation) (RHS)
- ▶ Many proteins re-localise after 12 hours LPS stimulations
- ▶ Example 1 - CDC42. In active state binds to a variety of effector proteins to regulate cellular responses at the plasma membrane (PM)
- ▶ Example 2 - RABs. Another GTPase regulates membrane trafficking to regulate cellular responses at the plasma membrane (PM) (RAB32 → GA to ER, RAB12 → PM to GA, RAB6B → GA to PM)

Compare app - Example 6: Human THP data (a monocytic cancer cell line)



Software

Infrastructure: MSnbase

ML: pRoloc

Visualisation: pRolocGUI

Data: pRolocdata.

Acknowledgements

- ▶ Prof. Kathryn Lilley (Cambridge Centre for Proteomics)
- ▶ Funding: BBSRC

Slides

- ▶ <https://zenodo.org/record/1927886>