

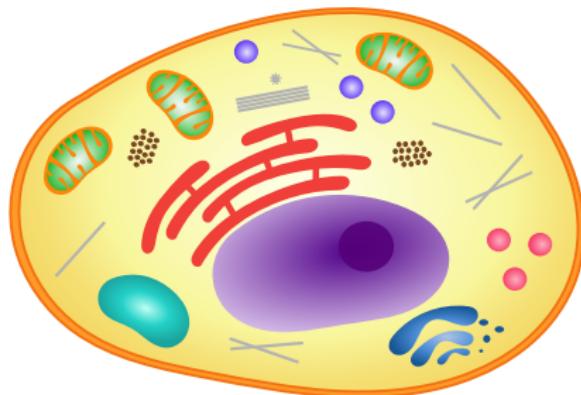
# Interactive visualisation of spatial proteomics data

L. Breckels and L. Gatto

University of Cambridge, UK | Institut de Duve, Belgium  
[lms79@cam.ac.uk](mailto:lms79@cam.ac.uk) | [laurent.gatto@uclouvain.be](mailto:laurent.gatto@uclouvain.be)

EuroBioC 2018, 6-7th December 2018

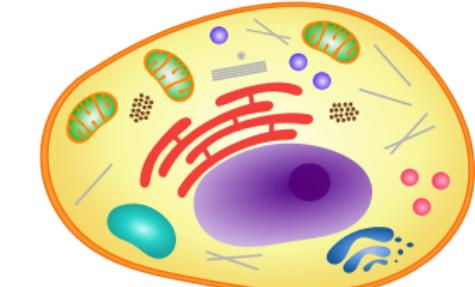
# Cell organisation



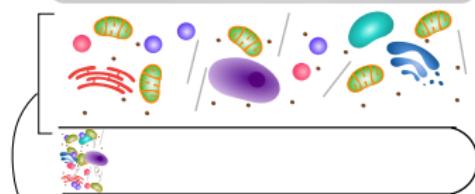
- ▶ Proteins are spatially organised according to function
- ▶ Significant correlation between disease classes and sub-cellular localisations
- ▶ Abnormal protein localisation leading to the loss of functional effects in diseases

**Spatial proteomics** is the systematic study of protein localisations.

# Localisation maps from quantitative proteomics

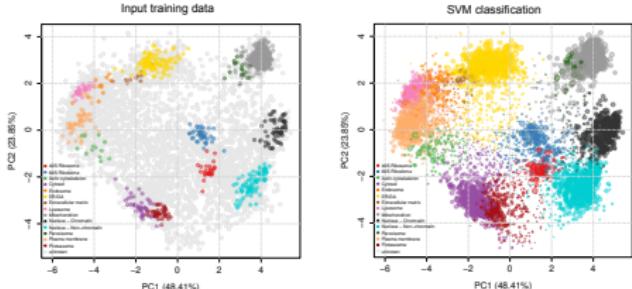
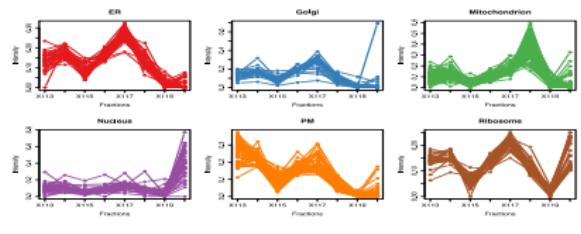


Cell lysis



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

	Fraction <sub>1</sub>	Fraction <sub>2</sub>	...	Fraction <sub>m</sub>	markers
prot <sub>1</sub>	q <sub>1,1</sub>	q <sub>1,2</sub>	...	q <sub>1, m</sub>	unknown
prot <sub>2</sub>	q <sub>2,1</sub>	q <sub>2,2</sub>	...	q <sub>2, m</sub>	organelle <sub>1</sub>
prot <sub>3</sub>	q <sub>3,1</sub>	q <sub>3,2</sub>	...	q <sub>3, m</sub>	unknown
prot <sub>4</sub>	q <sub>4,1</sub>	q <sub>4,2</sub>	...	q <sub>4, m</sub>	organelle <sub>2</sub>
...	...	...	...	...	...
prot <sub>i</sub>	q <sub>i,1</sub>	q <sub>i,2</sub>	...	q <sub>i, m</sub>	...
...	...	...	...	...	...
prot <sub>n</sub>	q <sub>n,1</sub>	q <sub>n,2</sub>	...	q <sub>n, m</sub>	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

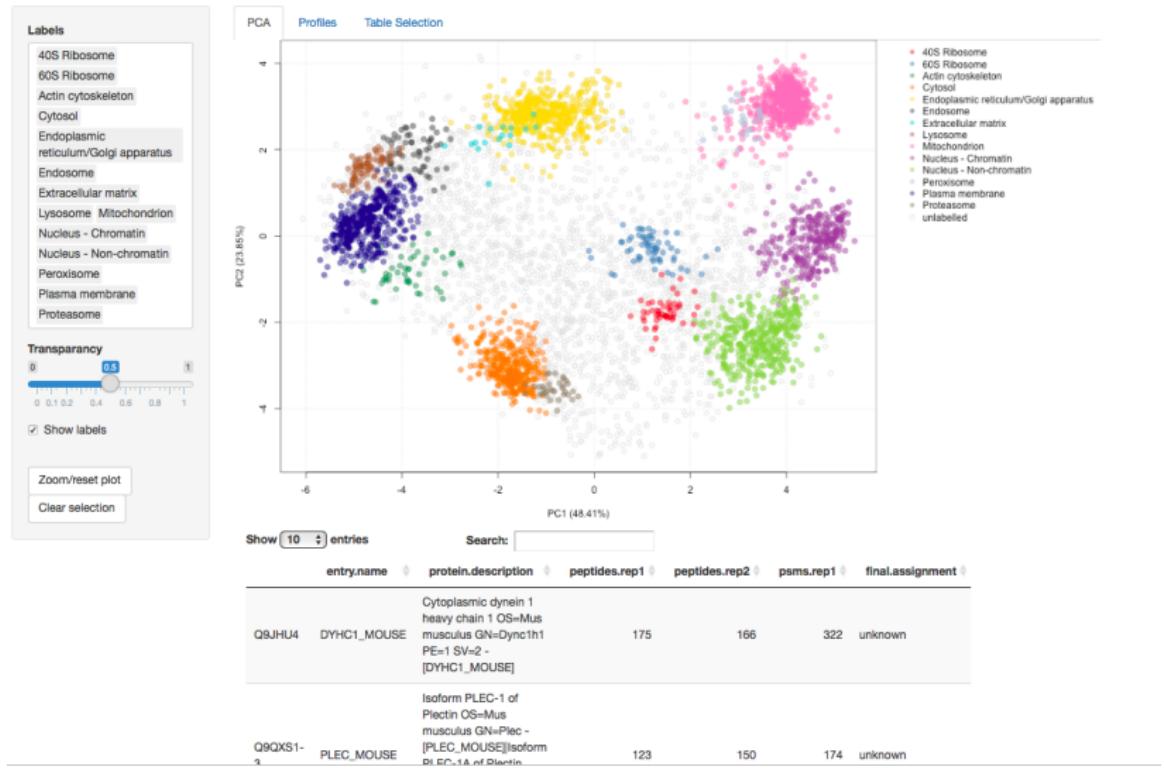
- ▶ pRolocGUI - interactive visualisation and data mining
- ▶ User driven: essential to transfer of findings easily between programmatic and graphical interfaces

# 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

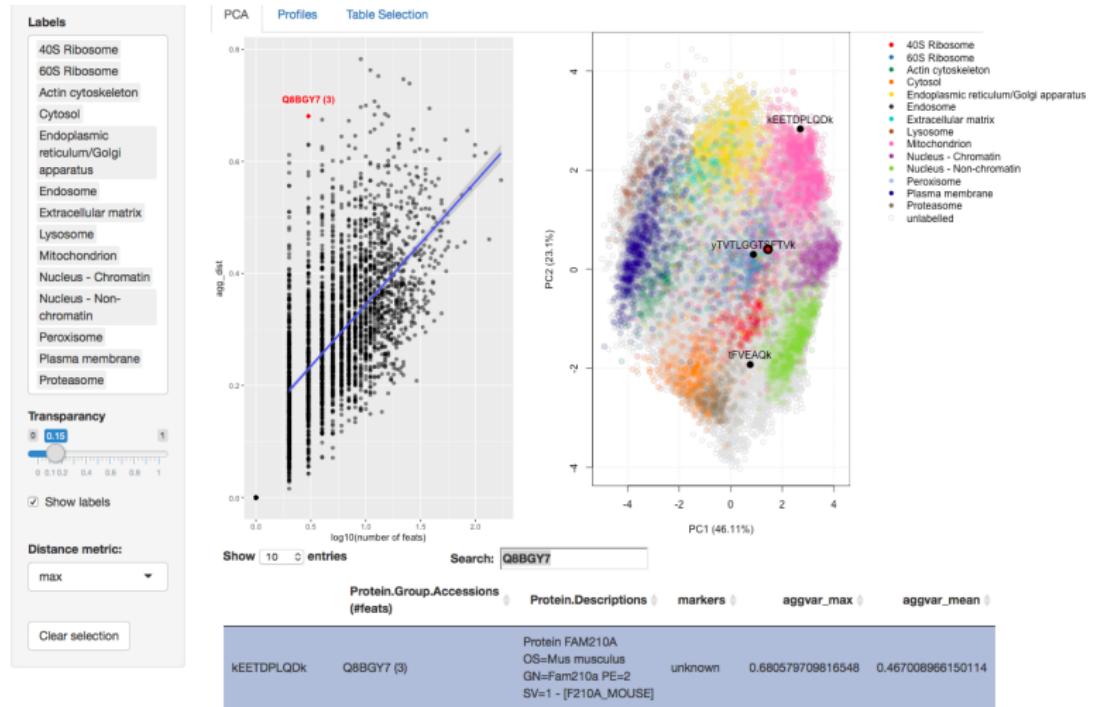
# The main application

## Use case: hyperLOPIT on mouse embryonic stem cells



# The aggregation application

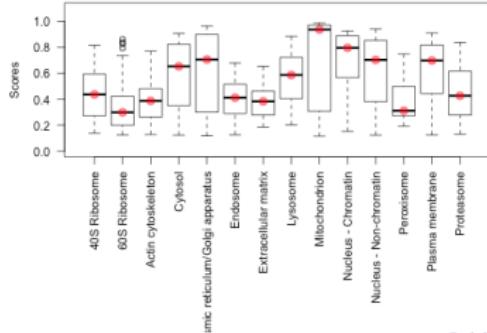
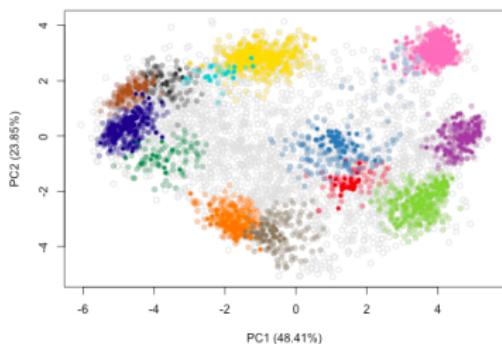
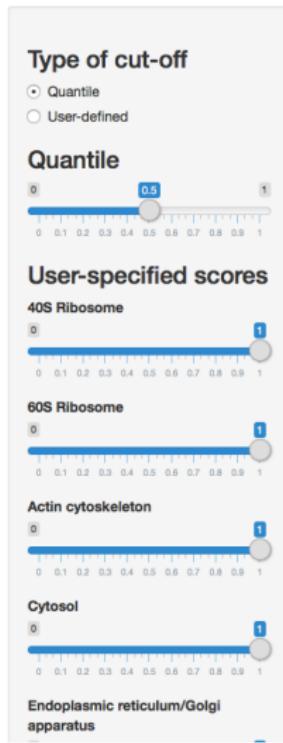
## Use case: Protein and peptide localisation



# The classify application

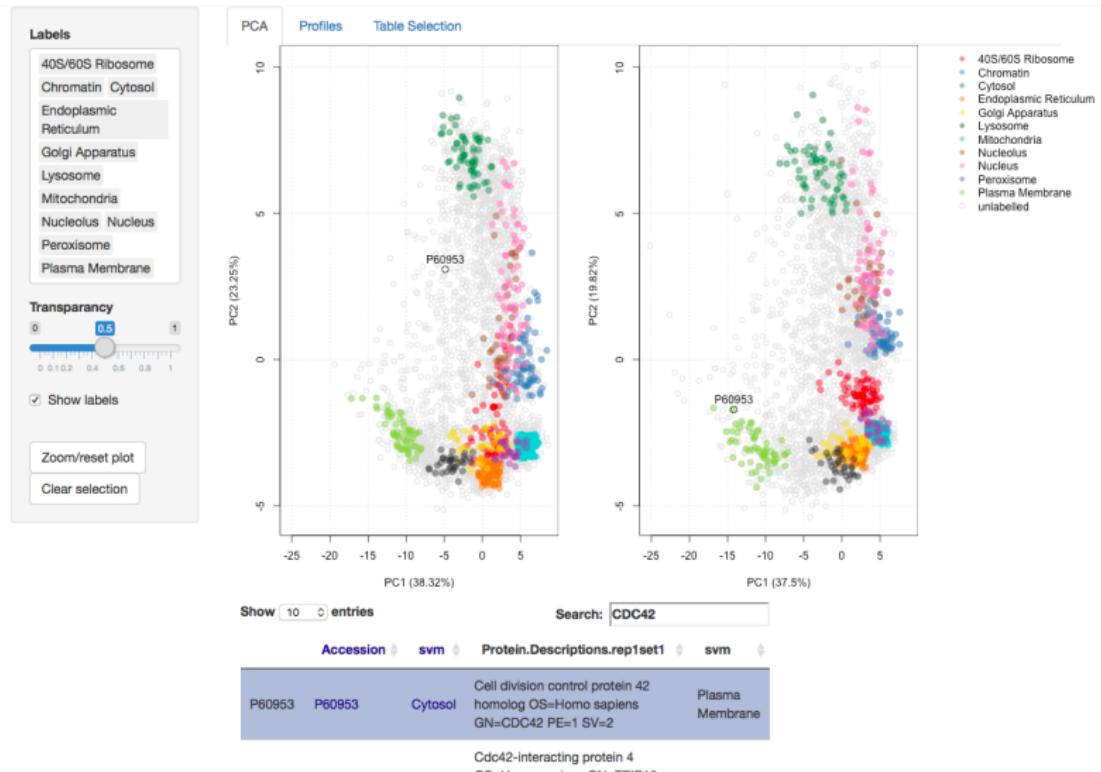
Use case: Setting thresholds for protein location by SVM

## Classification results



# The comparison application

Use case: hyperLOPIT on THP1 monocytic cancer cells after 12 hours LPS stimulation



## Software

Infrastructure: MSnbase

ML: pRoloc

Visualisation: pRolocGUI

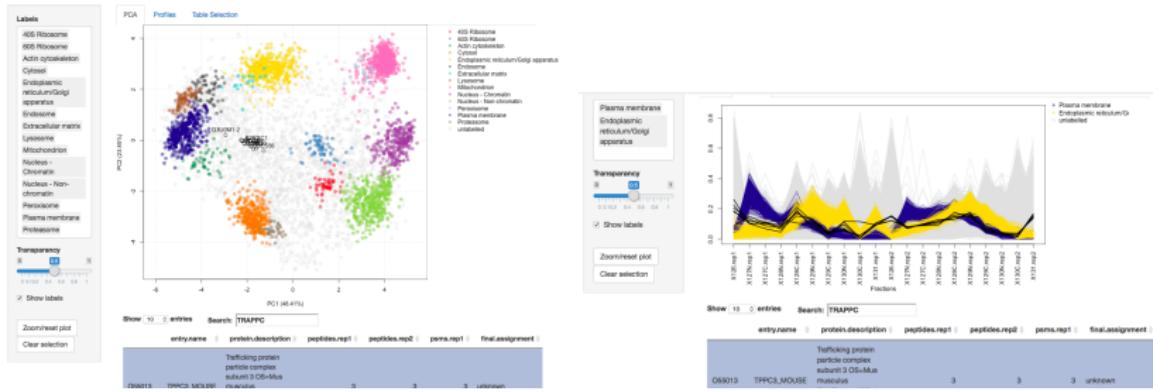
Data: pRolocdata.

## Acknowledgements

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

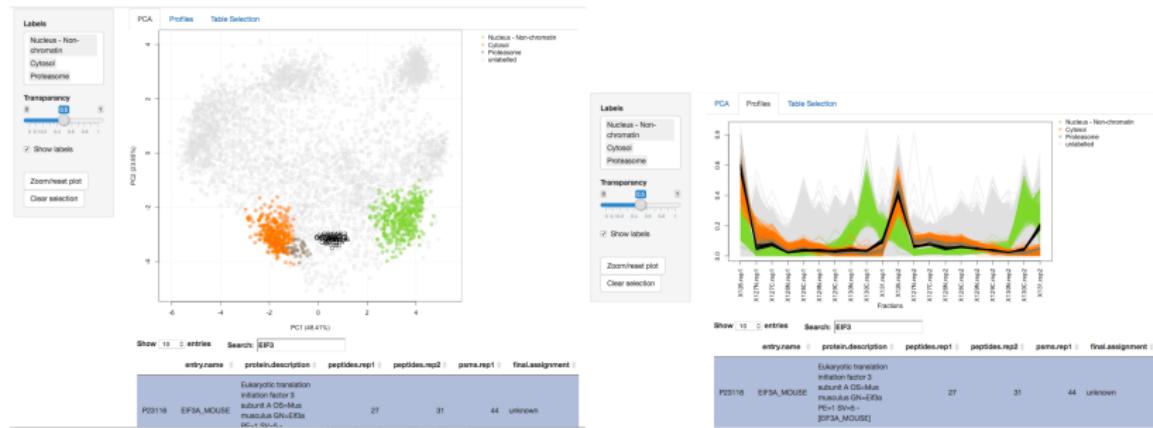
## Main app - Example 1: TRAPP complex

- ▶ 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



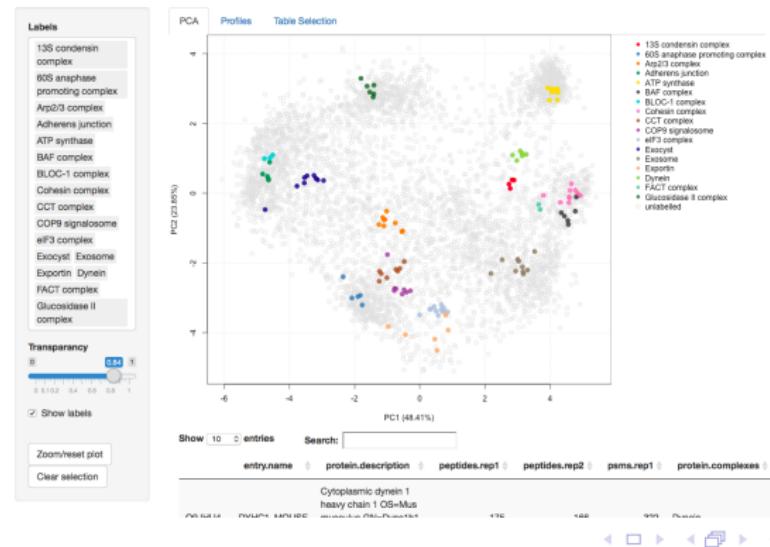
## Main app - Example 2: EIF3 complex

- ▶ 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



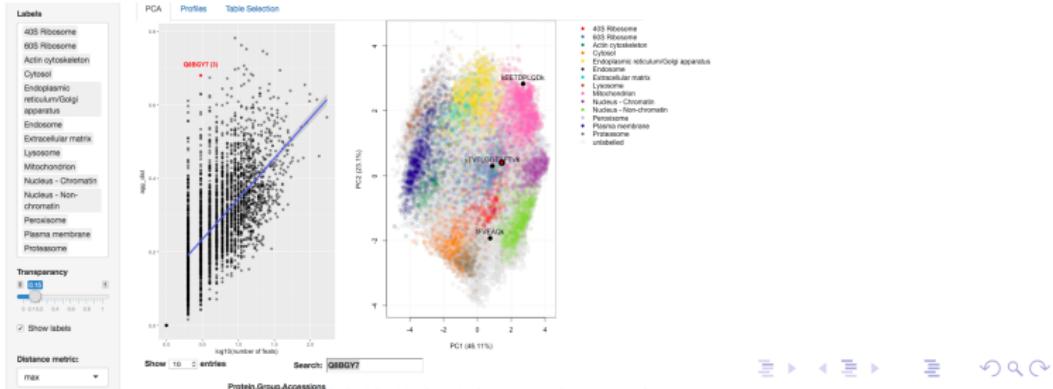
# Main app - Example 3: Visualisation of many complexes

- ▶ 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.



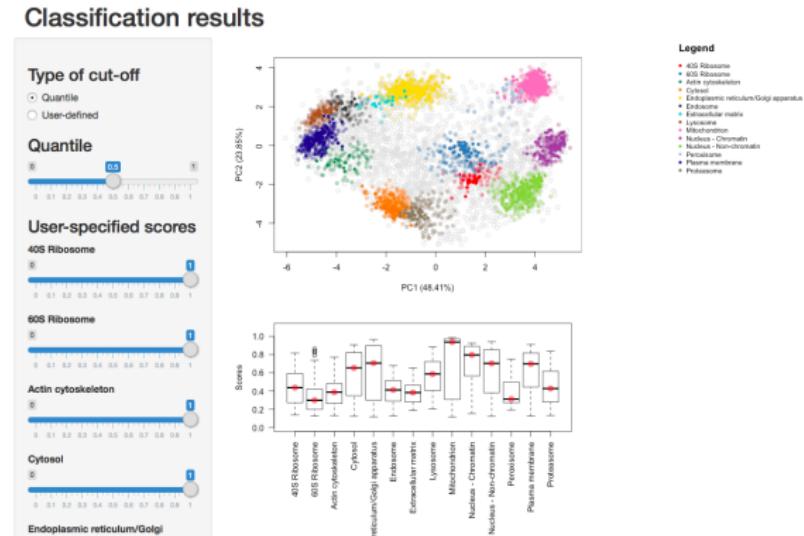
# Aggregation app - Example 4: Protein FAM210A

- ▶ 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



## 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)

