Hidden Alliances: RNA-Dependent Protein Interactions in Cancer Cells

Baureis, J., Ferdin, J., Nicklas, B., Wintel, L.
Data Analysis Project Molecular Biotechnology SS2025

1 Identifying RNA-Dependent Proteins from Proteomic Data

Why should we observe RNA-dependent proteins?

- **Key Regulators:** RBPs control RNA metabolism & gene expression.
- **Disease Links:** Missregulation is tied to cancer & neurodegeneration.
- **Functional Clues:** New RBPs hint at RNA's role in specific pathways.
- **Molecular Insights:** Deepens the understanding of cell cycle and cellular behavior

Key Characteristics of Our Dataset

- Dataset was created using the R-Deep approach and non-synchronized HeLa cells
- Contains 4765 proteins and their intensity values under normal (control) and RNase treated conditions in 25 sucrose density fractions, each fraction measured in triplicates

1

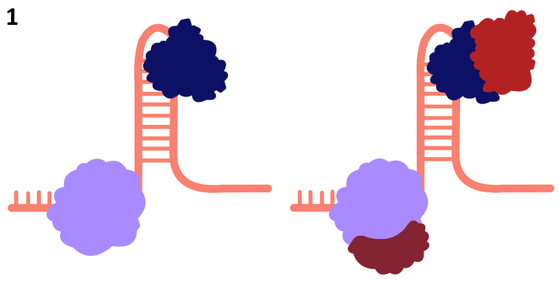
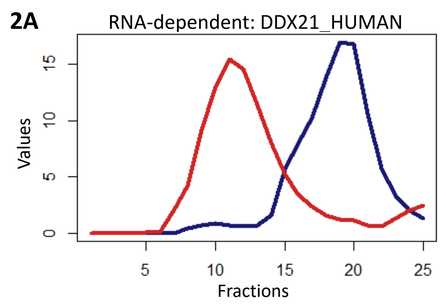


Fig. 1 Schematic illustration of RNA-dependency.

A protein is considered RNA-dependent if its interactome is dependent on RNA. It is either directly or indirectly attached to RNA & its functionality in biological context is associated with RNA

2A



B

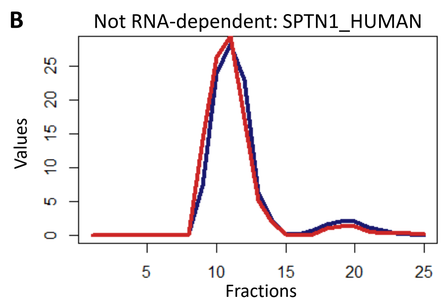


Fig. 2 Plot of proteins in data set. The data has been cleaned and normalized beforehand.

A) An RNA-dependent protein exhibits a shift for its intensity values between control (red) and RNase-treated (blue) group. **B)** A nor RNA-dependent protein exhibits no such shift.

2 Our Approach and Results

3

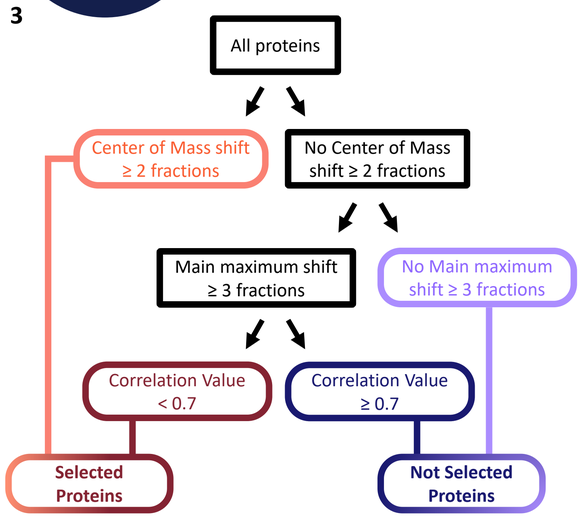
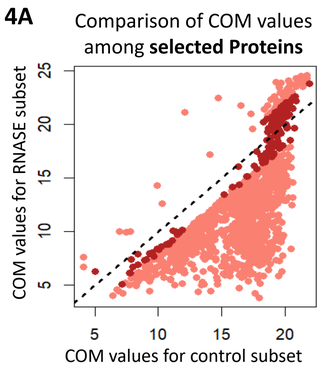


Fig. 3 Criteria for the classification of proteins regarding RNA-dependency

Selected (RNA-dependent):
- Center of mass (COM) shift ≥ 2 fractions *(orange)*
- or: no COM shift, but main peak shift ≥ 3 fractions and correlation < 0.7 *(red)*

Not selected (not RNA-dependent):
- No COM shift ≥ 2 and no main peak shift ≥ 3 *(purple)*
- or: peak main shift without COM shift and correlation ≥ 0.7 *(dark blue)*

4A



B

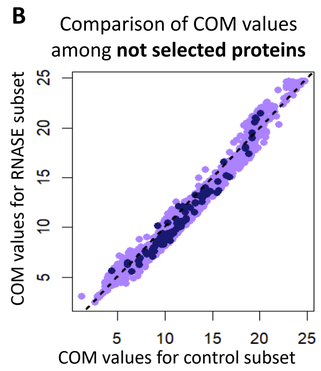


Fig. 4 Center-of-mass (COM) shifts between control and RNase conditions.

Each dot represents one protein. X: COM in control, Y: COM in RNase. **A)** Selected proteins. **B)** Not selected proteins

5

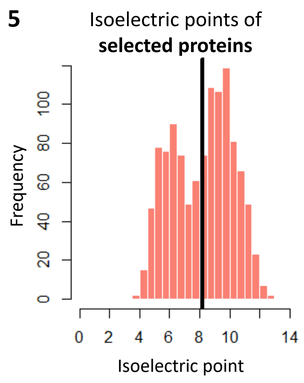
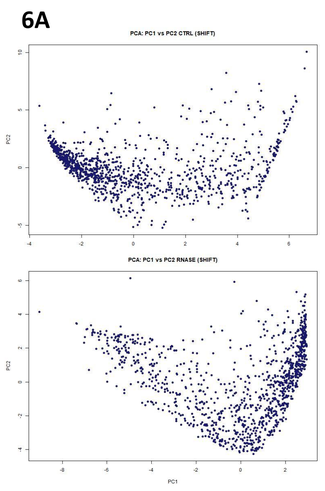


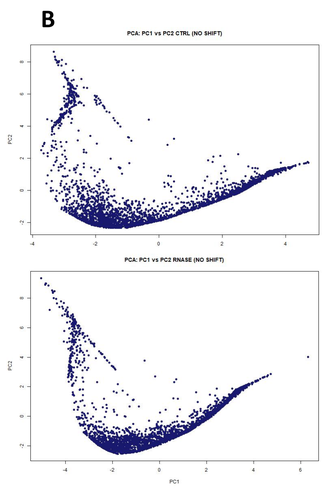
Fig 5. Isoelectric point (pI) distribution of selected proteins.

Mean pI = 8.16 (black line); one-sided t-test against pI = 7.0, p-value = 2.2e-16, indicates that selected proteins have significantly higher isoelectric points.

6A



B



C

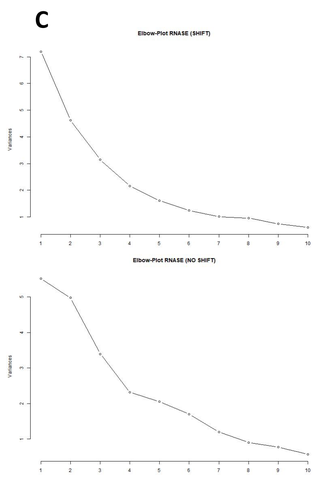
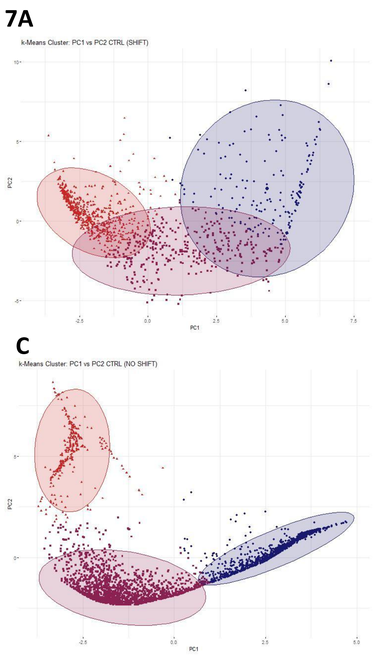


Fig. 6 Principal Component Analyses and Elbowplots of the selected and the non-selected proteins. RNASE and CTRL are plotted separately for comparison in the PCA. The Elbowplots shows only RNASE.

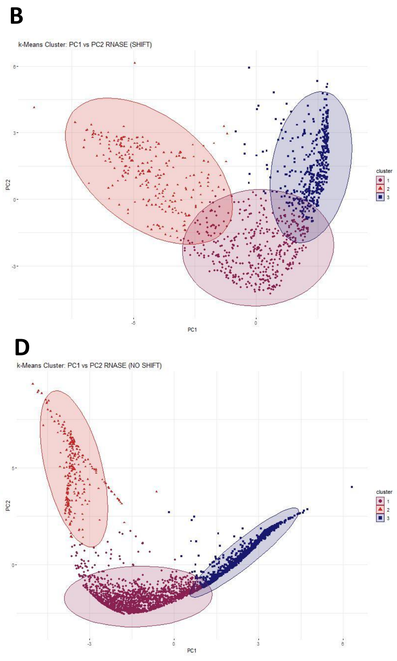
A) PCA of the selected proteins. The data points of the RNASE compared to the CTRL make up an overall similar shape, but a shift is visible in the density of the points. **B)** PCA of the non-selected proteins. The points of the RNASE and CTRL form mostly the same structure. **C)** Elbow-Plot of the RNASE. The knick of the elbow is between 3 and 4. To compare selected and non-selected proteins, we decided to use 3 clusters in the kmeans clustering.

3 Discussion

7A



B



C

D

Fig. 7 kmeans clustering of the selected and not-selected proteins

A) Shows the 3 clusters of the CTRL of the selected proteins. **B)** Shows the 3 clusters of the RNASE of the selected proteins. A significant shift in form and location of the clusters is noticeable. **C)** Shows the 3 clusters of the CTRL of the not-selected proteins. **D)** Shows the 3 clusters of the RNASE of the not-selected proteins. No shift in form and location is noticeable.

Fig. 8 Linear regression analyses between the selected proteins and the not-selected proteins each, with global maxima of the selected CTRL proteins as target variable

A) The regression analysis for the selected proteins describes the target variable well. **B)** The analysis of the not-selected proteins does not describe the target variable well. This proves, that there is a difference between the selected and not-selected proteins, and therefore, that the selection criteria worked.

8A

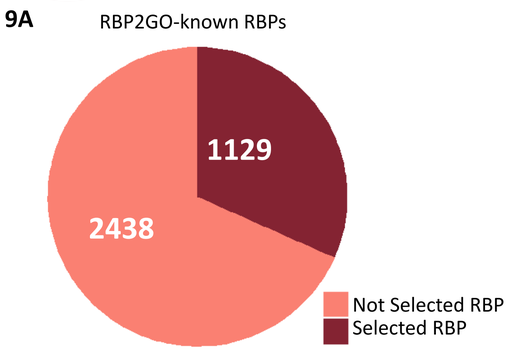
```
Call:
lm(formula = target ~ ctrl_1 + rnase_1, data = df_regression_selected_11)

Residuals:
    Min      1Q  Median      3Q     Max
-15.254  -0.896  -0.091   1.100  12.043

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 16.27942    0.05997 271.443  < 2e-16 ***
ctrl_1      -1.50827    0.02997 -50.321  < 2e-16 ***
rnase_1      0.14991    0.02935   5.108 3.79e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.061 on 1178 degrees of freedom
Multiple R-squared:  0.7648,    Adjusted R-squared:  0.7644
F-statistic:  1916 on 2 and 1178 DF,  p-value: < 2.2e-16
```

B

```
Call:
lm(formula = target ~ ctrl_ns_1 + rnase_ns_2, data = df_regression_not_selected_12)

Residuals:
    Min      1Q  Median      3Q     Max
-15.404  -3.423   1.761   2.808   8.949

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 16.2759487  0.0707927 229.910   <2e-16 ***
ctrl_ns_1    0.0006054  0.0302818   0.020   0.9841
rnase_ns_2  -0.0607672  0.0320809  -1.894   0.0583 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.238 on 3581 degrees of freedom
Multiple R-squared:  0.001019,  Adjusted R-squared:  0.000461
F-statistic: 1.826 on 2 and 3581 DF,  p-value: 0.1612
```

4 Our Achievements

9A



B

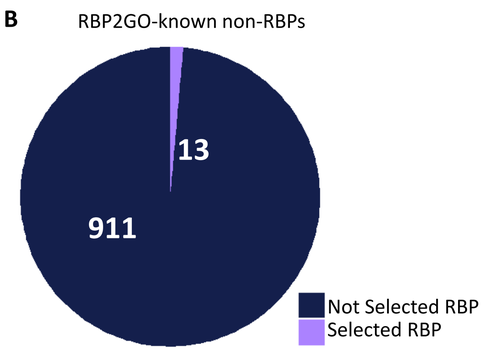


Fig. 9 Selection Status of RBPs and Non-RBPs According to the RBP2GO Database

A) Pie chart of RBP2GO-annotated RBPs in the dataset, showing selected (orange) versus not selected (red). **B)** Pie chart of RBP2GO-annotated Non-RBPs in the dataset, showing wrongly selected (purple) versus not selected (blue).

5 References

- Sternburg *et al.,* Global Approaches in Studying RNA-Binding Protein Interaction Networks, 2020, Trends in Biochemical Sciences
- Caudron-Herger *et al.,* R-DeeP Proteome-wide and Quantitative Identification of RNA-Dependent Proteins by Density Gradient Ultracentrifugation, 2019, Molecular Cell
- Caudron-Herger *et al.,* Identification, quantification and bioinformatic analysis of RNA-dependent proteins by RNase treatment and density gradient ultracentrifugation using R-DeeP, 2020, Nature Protocols
- Corley *et al.*, How RNA-Binding Proteins Interact with RNA Molecules and Mechanisms, 2020, Molecular Cell
- Caudron-Herger *et al.,* RBP2GO: a comprehensive pan-species database on RNA-binding proteins, their interactions and functions, 2021, Nucleic Acids Research

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

IPMB