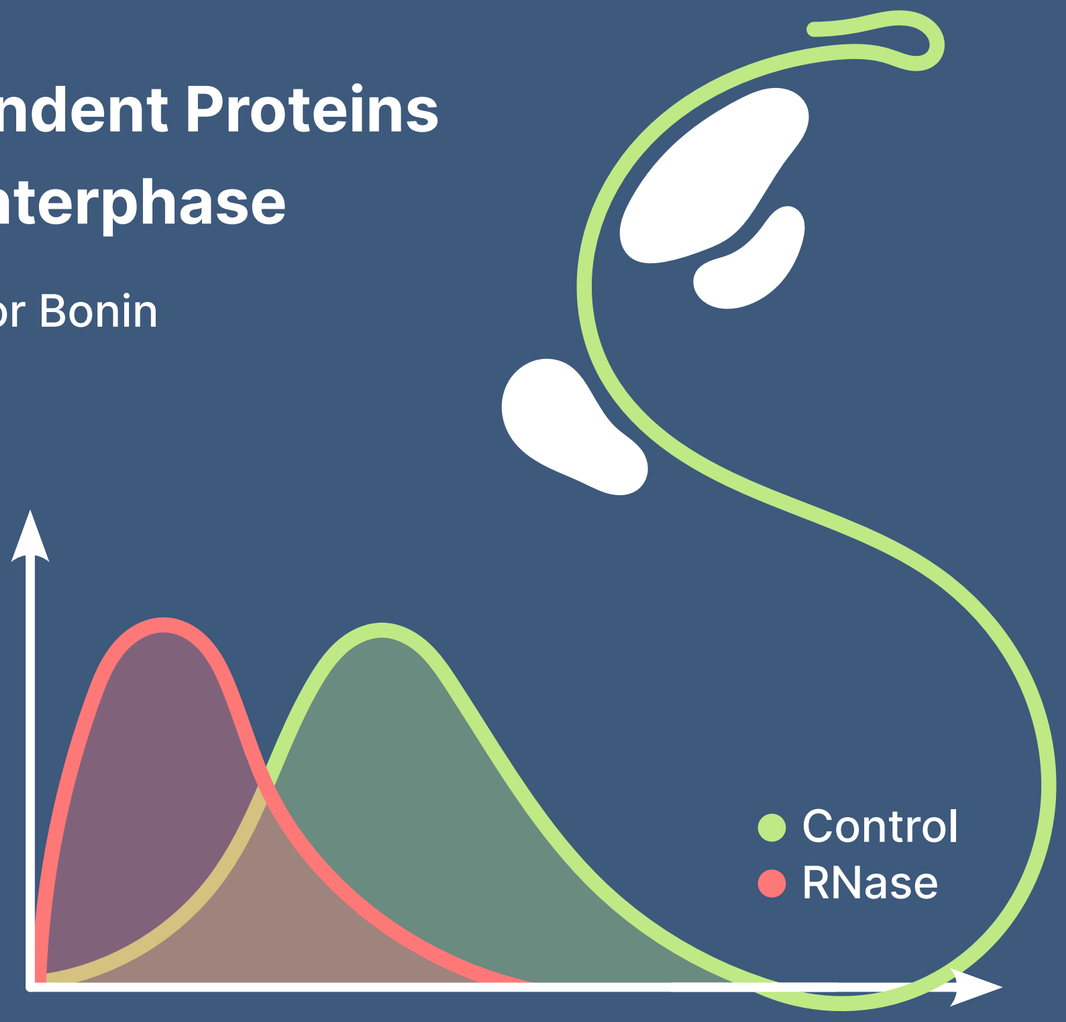
Project Proposal

Proteome-wide Screen for RNA-dependent Proteins

Team 3: HeLa Cells Synchronized in Interphase

Hannah Brehm, Johann Blakytny, Kira Hoffmann, Viktor Bonin



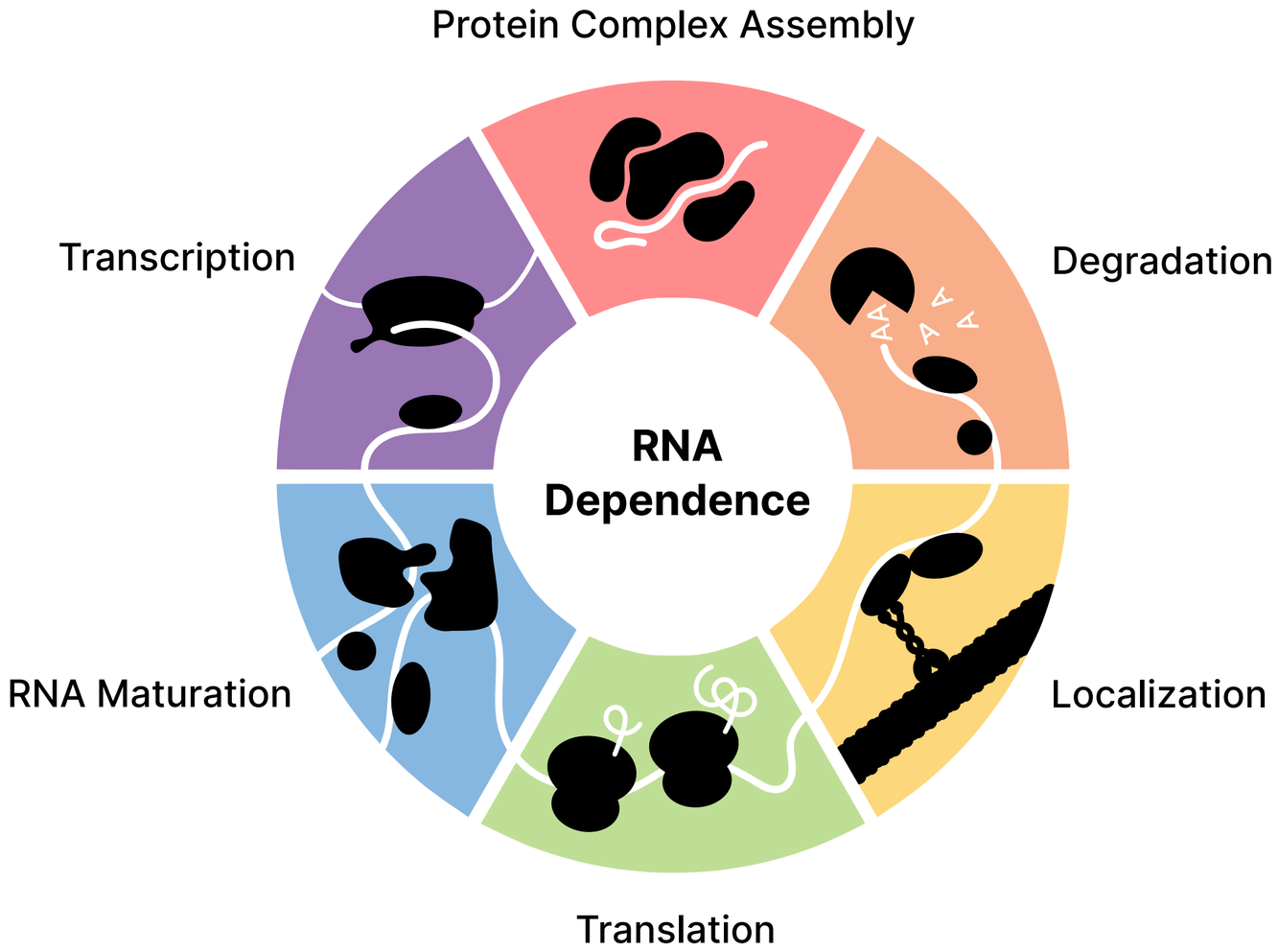
Data Analysis Project Summer Term 2023

Supervisor: Dr. Maïwen Caudron-Herger

Tutor: Fabio Rauscher

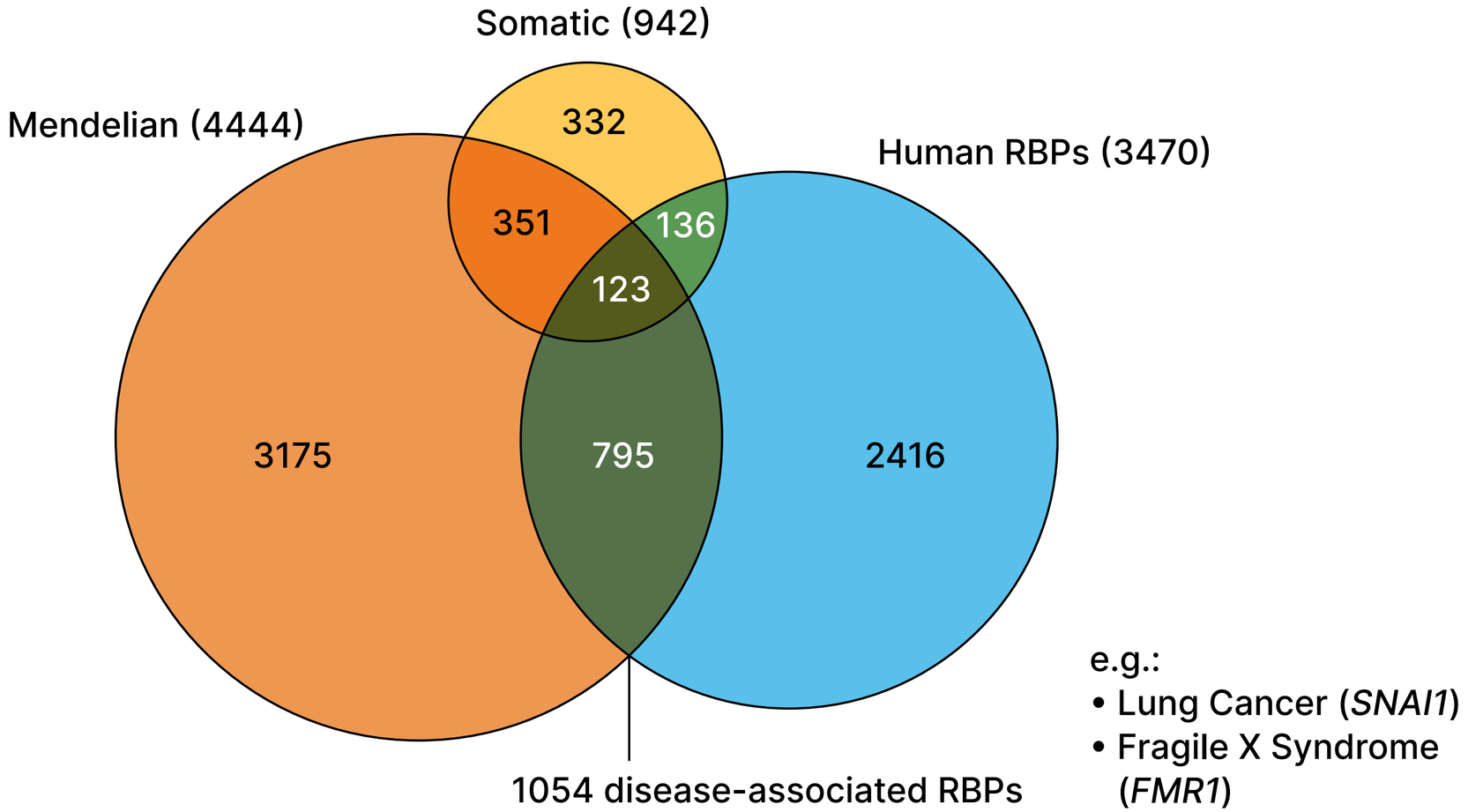
Wednesday, 17.05.2023

Functions of RNA-dependent Proteins



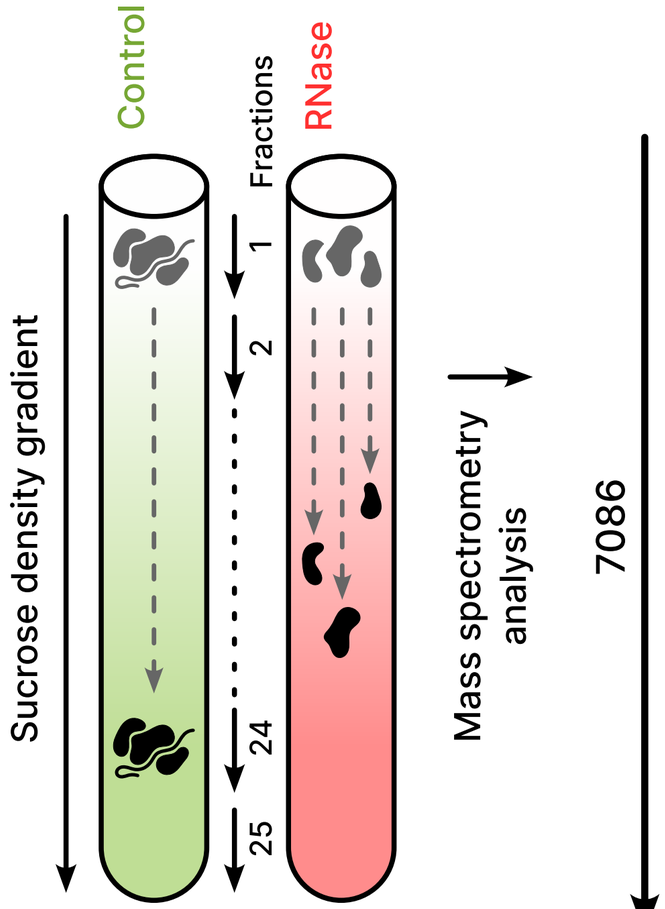
Adapted from: Gebauer, F., Schwarzl, T., Valcárcel, J., and Hentze, M.W. (2021). RNA-binding proteins in human genetic disease. Nat. Rev. Genet. *22*, 185-198.

RNA-binding Proteins in Human Diseases



Gebauer, F., Schwarzl, T., Valcárcel, J., and Hentze, M.W. (2021). RNA-binding proteins in human genetic disease. Nat. Rev. Genet. *22*, 185-198.

Dataset for HeLa Cells Synchronized in Interphase



7086

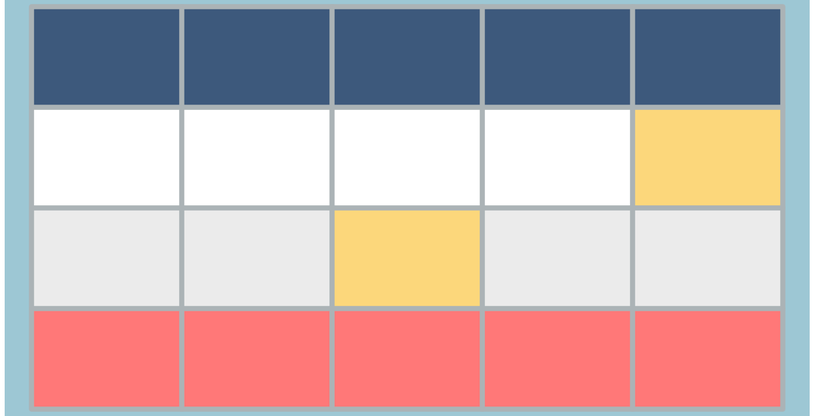
150 →

	Fraction1_ Ctrl_Rep1	Fraction1_ Ctrl_Rep2	Fraction1_ Ctrl_Rep3	Fraction1_ RNase_Rep1	...	Fraction25_ RNase_Rep3
1433B_HUMAN	134729.90	146356.70	192643.80	293592.10	...	0.00
1433E_HUMAN	120892.73	129849.82	195391.90	280333.80	...	0.00
1433F_HUMAN	73231.50	77305.20	105863.00	165178.00	...	0.00
1433G_HUMAN	91820.40	91078.10	123128.20	190751.20	...	0.00
1433S_HUMAN	152139.00	144703.30	186727.30	289461.00	...	0.00
1433T_HUMAN	116713.77	125549.61	175656.99	266160.90	...	0.00
1433Z_HUMAN	277512.69	267068.00	366251.21	535765.40	...	0.00
...	...	...	...	...	...	...

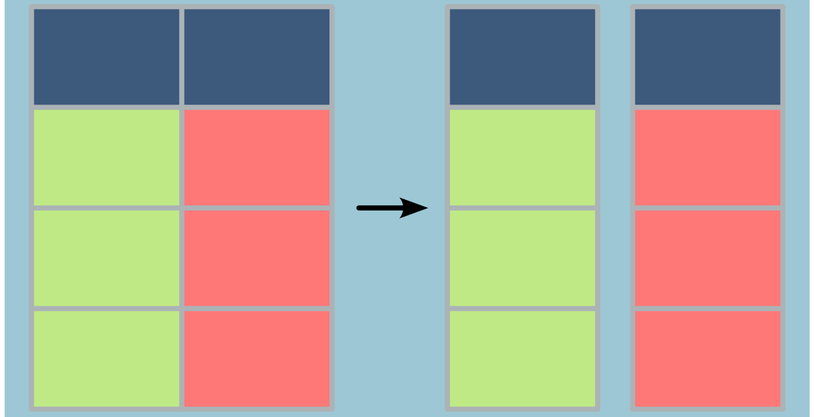
Caudron-Herger, M., *et al.* (2019). R-DeeP: Proteome-wide and Quantitative Identification of RNA-Dependent Proteins by Density Gradient Ultracentrifugation. Mol. Cell *75*, 184-199.

Step 1: Data Cleanup and Normalization

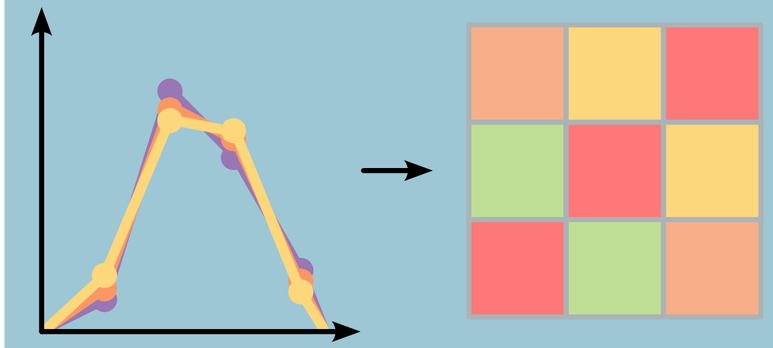
1.1 NA Values and Zero Rows

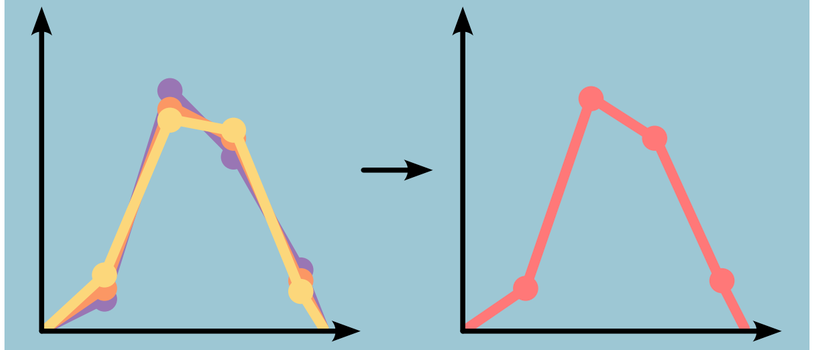


1.2 Split Dataset into Control and RNase



1.3 Pearson and Mean of Replicates

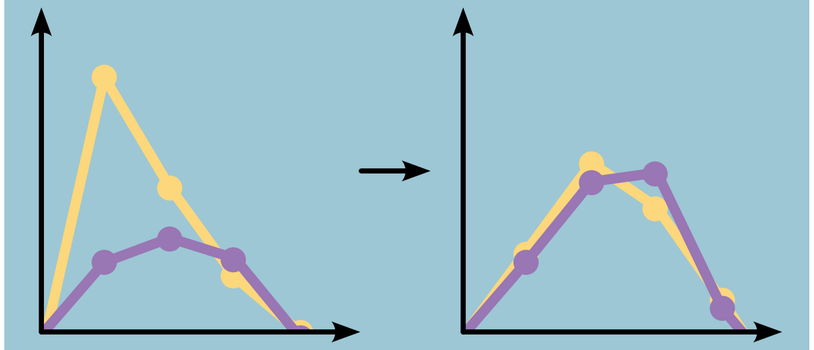




1.4 Normalization

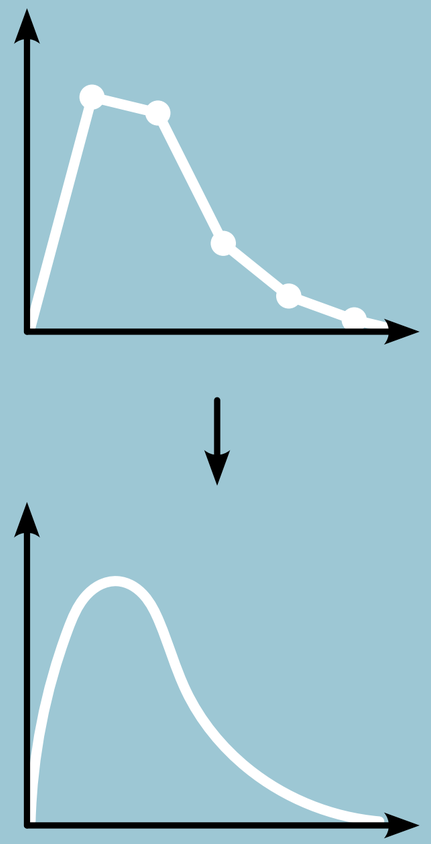
				Σ
				100
				100
				100

1.5 Z-Scaling

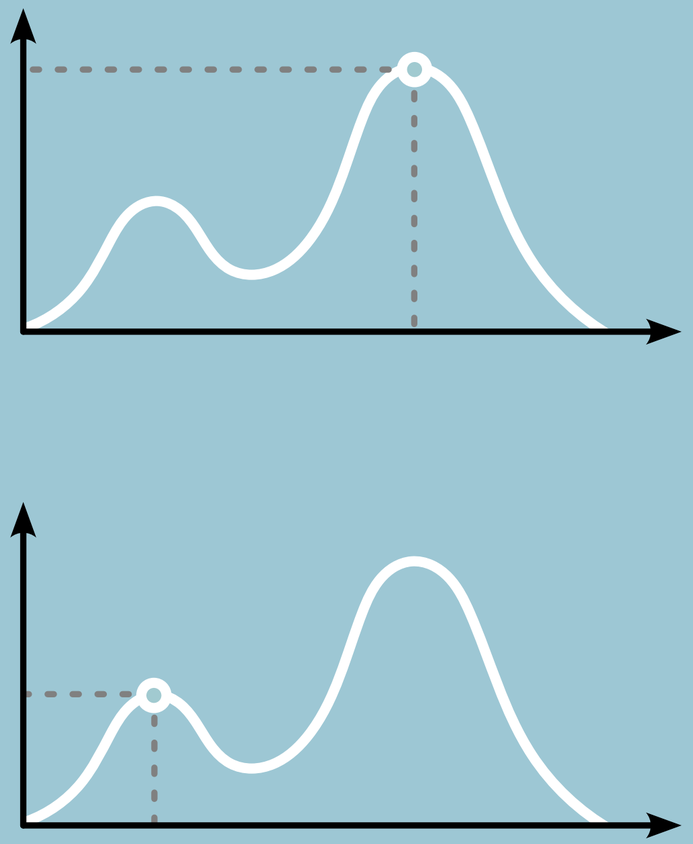


Step 2: Finding Maxima and Gaussian Fitting

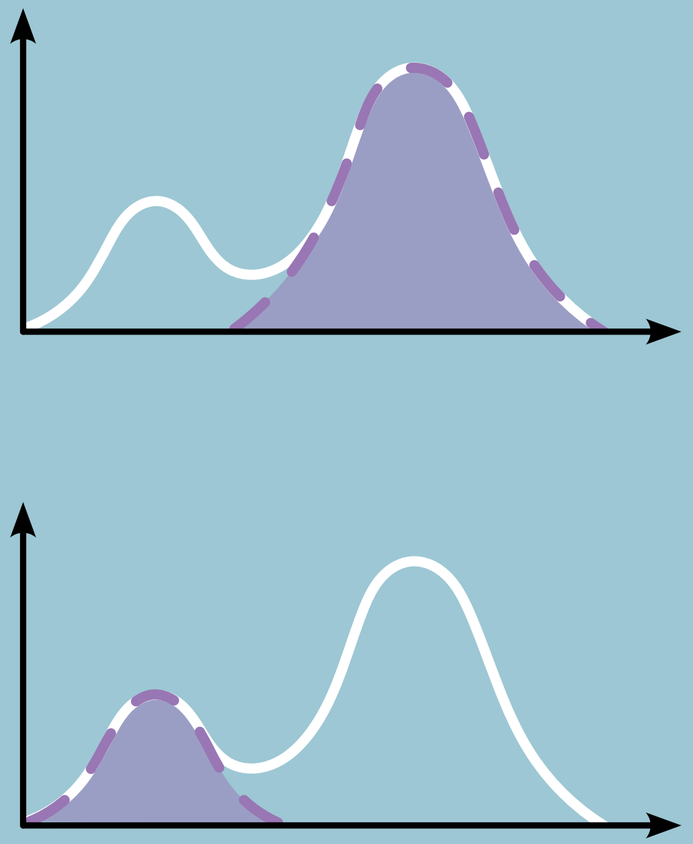
2.1 Smoothing



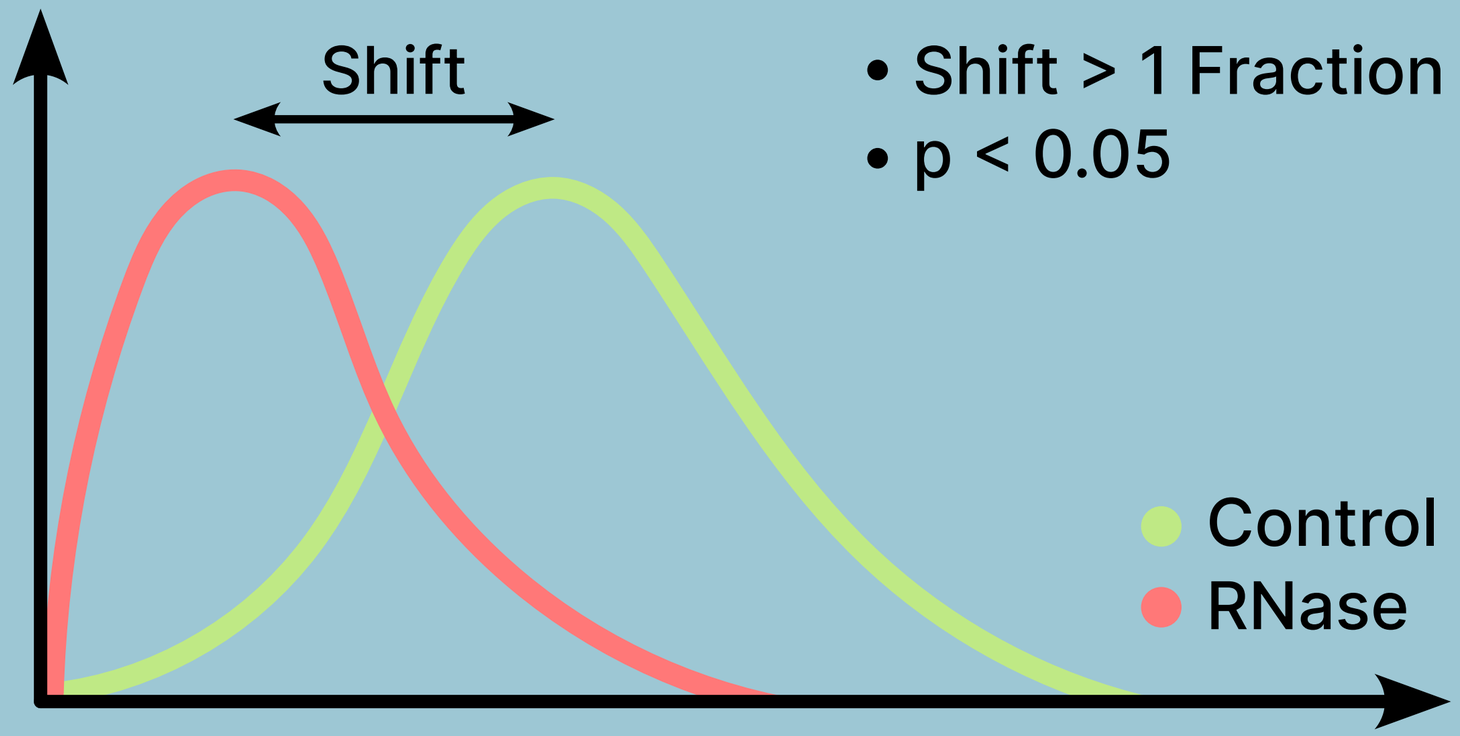
2.2 Global and Local Maxima



2.3 Gaussian Fitting



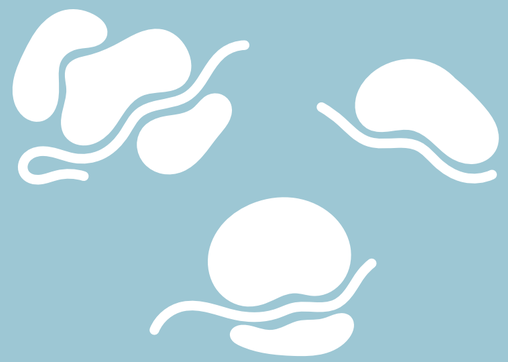
Step 3: Defining Selection Criteria



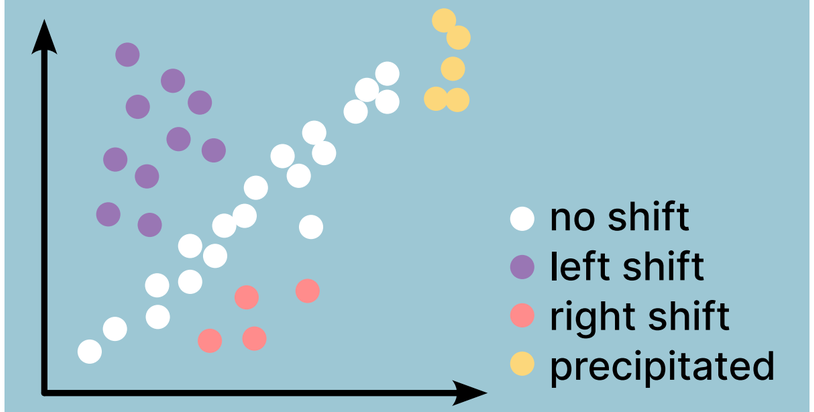
- Shift > 1 Fraction
- $p < 0.05$

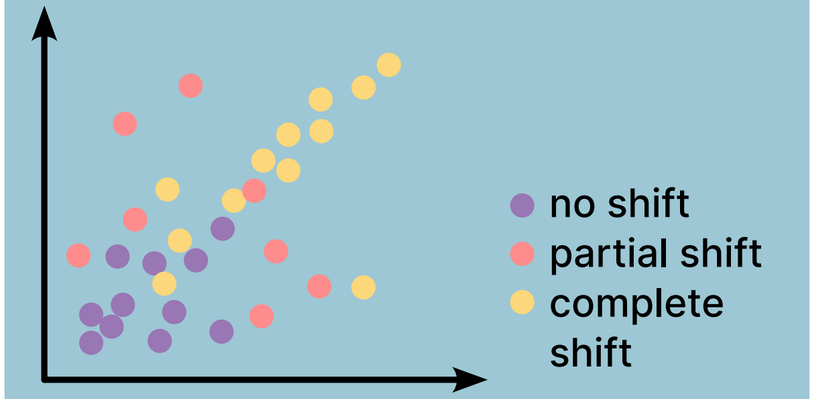
Step 4: Identification of RNA-dependent Proteins

4.1 Application of Our Selection Criteria

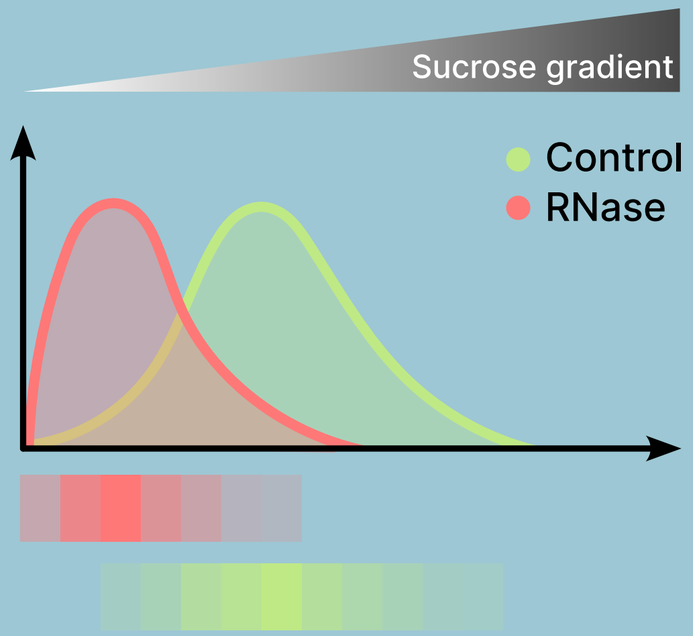


4.2 k-Means Clustering



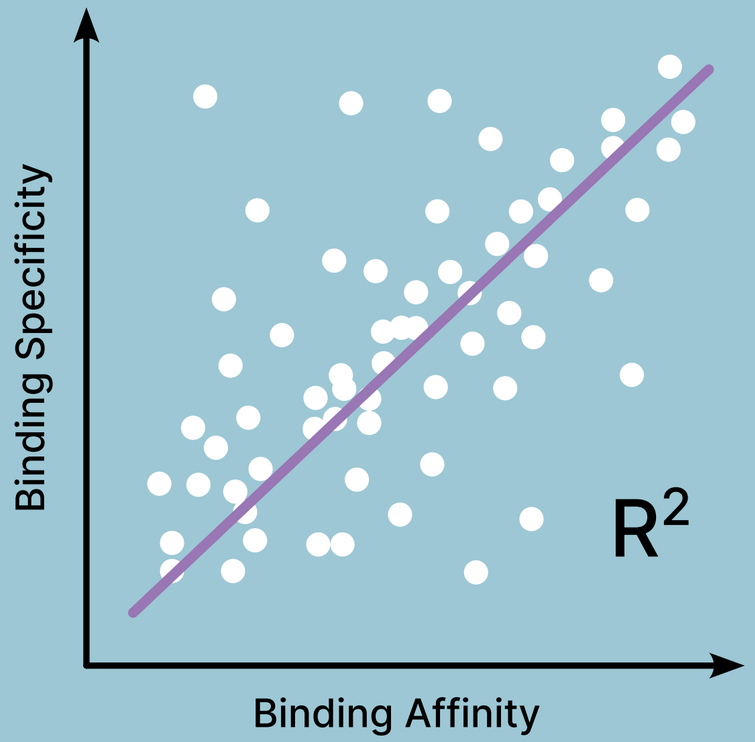


4.3 Graphical Representation

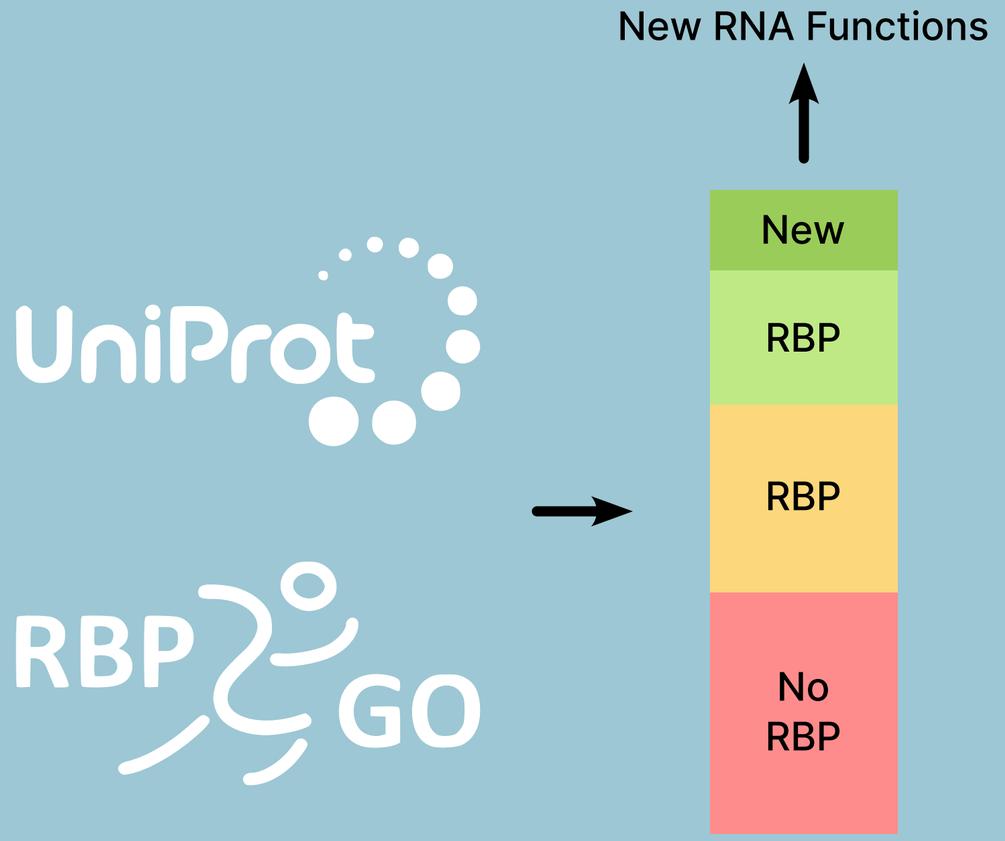


Step 5: Linear Regression and Further Analysis

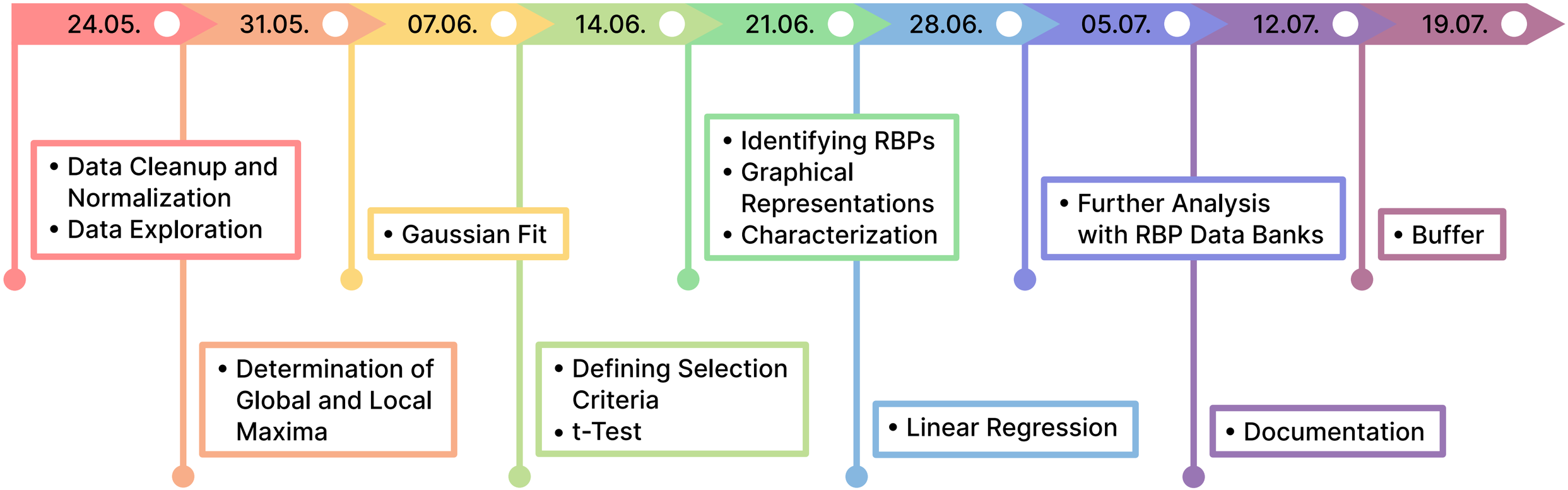
5.1 Linear Regression



5.2 Further Analysis



Timeline


24.05.
• Data Cleanup and Normalization
• Data Exploration
31.05.
• Determination of Global and Local Maxima
07.06.
• Gaussian Fit
14.06.
• Defining Selection Criteria
• t-Test
21.06.
• Identifying RBPs
• Graphical Representations
• Characterization
28.06.
• Linear Regression
05.07.
• Further Analysis with RBP Data Banks
12.07.
• Documentation
19.07.
• Buffer