Supporting material for:
In silico characterization and prediction of global protein-mRNA interactions in yeast
Vera Pancaldi1, Jürg Bähler1

In this document:

Additional files:

Datasets:

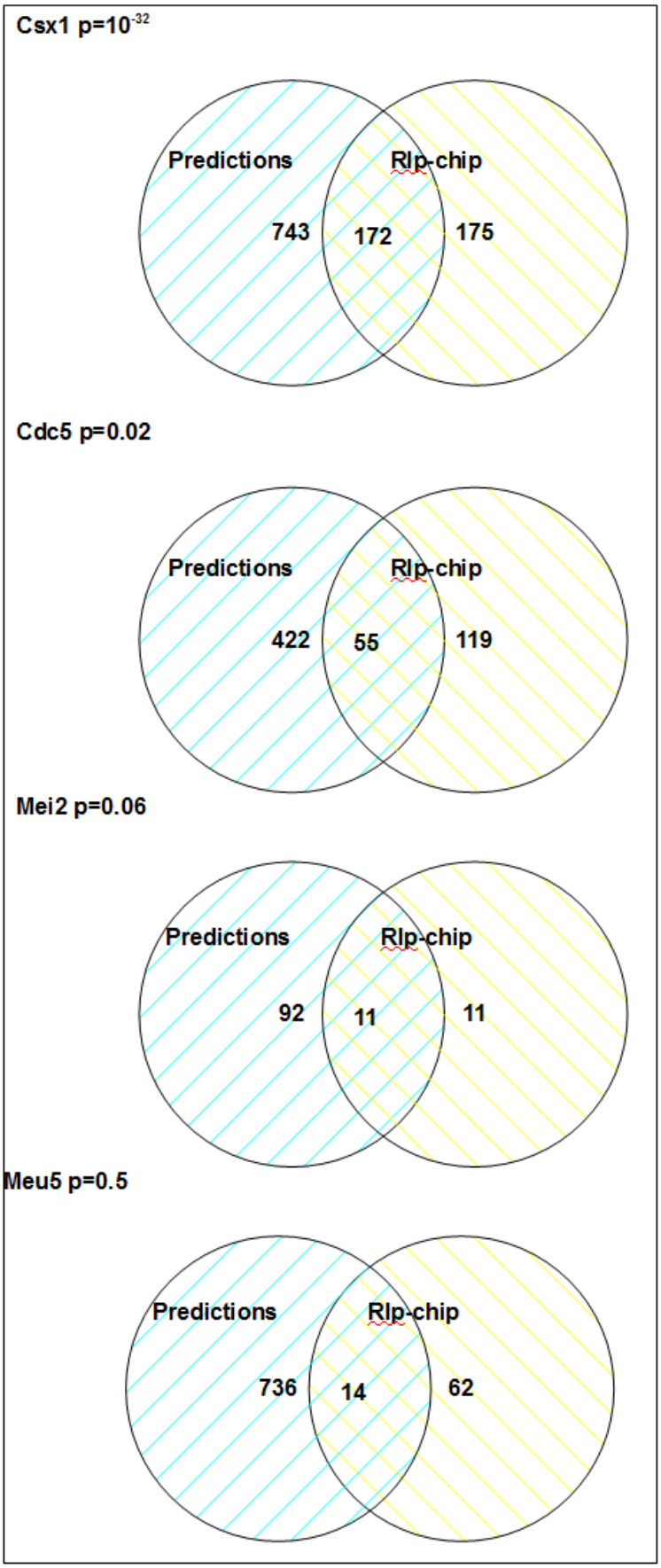
correlation_data.xls, workbook contains Datasets S1, S2, S2b, S4, S5, S6:
Dataset S1: Correlations across genome, source for Figure 1A, B
Dataset S2: Difference in feature values between RBPs and other proteins
Dataset S2b: Difference in feature values between Hogan proteins and RBPs.
Dataset S4: Correlations between protein features and features of their targets in the positive set, with corresponding p-values.
Dataset S5: Correlations between protein features and features of their targets averaged over 100 negative sets, with corresponding p-values
Dataset S6: Correlations between protein and mRNA features only significant in the positive set
Dataset S7: Workbook: Correlation data for human protein-mRNA pairs
Dataset S3: Positive training set
Dataset S8: Negative training set for predictions

prediction_data.xls, workbook contains Dataset S9, S10, S11:
Dataset S9: Cross validation results
Dataset S10: RF feature importance
Dataset S11: Leave-protein-out results and description of RBPs in Hogan et al. 2008.

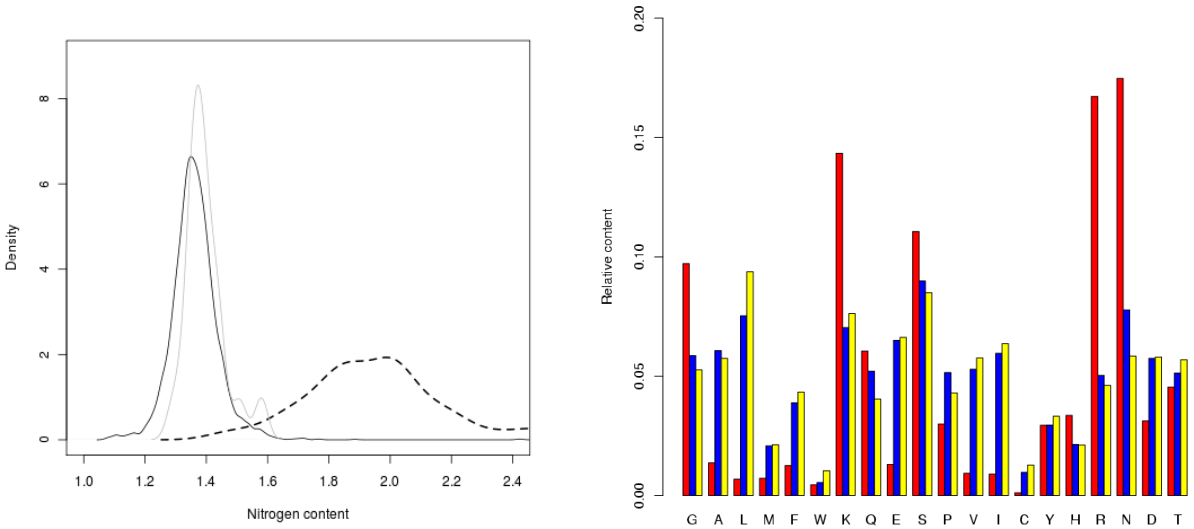
Dataset S12: She2p predicted target analysis
Dataset S13: Predictions of targets bound by all fission yeast RBPs considered.

Text S1:
We report the predicted targets of a few RBPs that have been experimentally studied in *S. pombe*. Only a few of the features could be included in the SVM without obtaining no predictions at all for RBP targets (strand, start, stop, length, IP, G,A, L, M, F, W, K, stems, UTR properties). The SVM trained on *S. cerevisiae* protein-mRNA binding data predicted 750 mRNA targets for Meu5, of which 14 overlap with the targets validated by RIp-chip (Amorim, Cotobal et al. 2010). Moreover, 915 mRNAs were predicted to be bound by Csx1, which was reported to stabilize *atf1* mRNA(Rodriguez-Gabriel, Burns et al. 2003) , but we only predict this interaction with a probability of 43%. We find that 181 of our 915 predictions overlap targets of Csx1 identified in preliminary RIp-chip experiments (Juan Mata, personal communication). 211 mRNAs are predicted to be bound only by Csx1. We also predict 477 targets for Cdc5, 55 overlapping RIp-chip experiments (Juan Mata, personal communication) of which 81 are uniquely predicted for Cdc5, 603 targets forMei2, 11 of which are also in preliminary RIp-chip experiments (Juan Mata, personal communication) and 11 that are predicted uniquely for Mei2 (Supplementary Figure S1). A total of 53 mRNAs are predicted to be bound by all the proteins investigated (Dataset S12). This could either suggest bias in our method, which predicts these mRNAs to be always bound, or it could actually reflect a subset of transcripts which are all bound by these proteins. More data and further investigation are needed to identify the correct explanation. Further investigations could also shed light on why some proteins are predicted to have more targets.

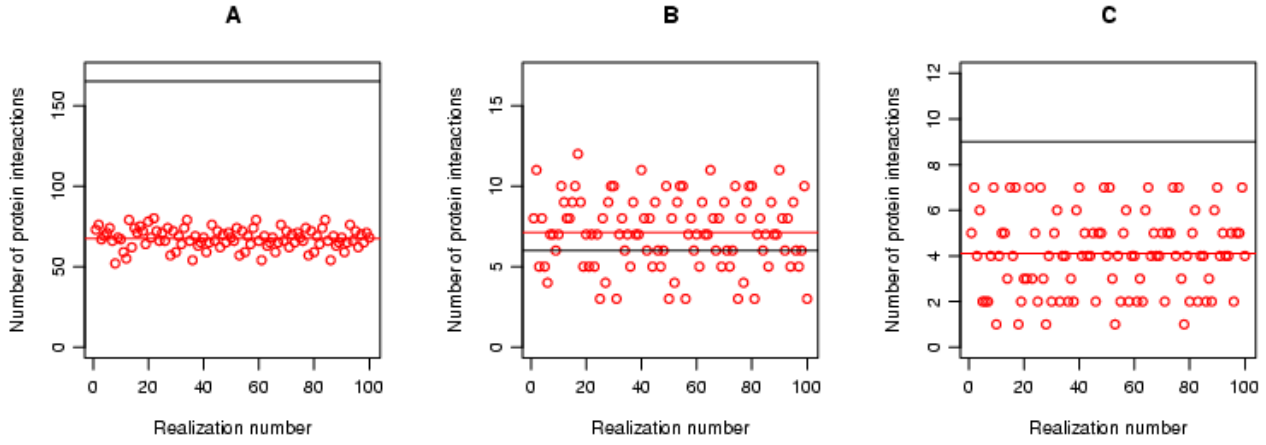
References:
Amorim, M. J., C. Cotobal, et al. (2010). "Global coordination of transcriptional control and mRNA decay during cellular differentiation." Mol Syst Biol **6**.
Rodriguez-Gabriel, M. A., G. Burns, et al. (2003). "RNA-binding protein Csx1 mediates global control of gene expression in response to oxidative stress." Embo J **22**(23): 6256-6266.



Supporting Figure S1: Comparison of predictions of fission yeast targets with preliminary experimental data.



Supporting Figure S2: Comparison of distributions of amino acids and nitrogen content.
Left: comparison of nitrogen content in all budding yeast proteins (black solid line), in the proteins analysed in the Hogan et al. study (grey solid line) and in the predicted RNA interfaces of the Hogan proteins (dashed line). The mean nitrogen content is 1.36 for all budding yeast proteins, 1.39 for the proteins studied in the Hogan data and 1.95 in the RNA interfaces of the Hogan proteins, showing a significant enrichment for amino acids rich in nitrogen in the part of the protein that mediates the interaction with the RNA ($p<10^{-15}$). This is consistent with the observed enrichment of Arginine, Histidine and other nitrogen rich aminoacids in the protein-RNA interfaces.
Right: Comparison of amino acid relative abundance in all budding yeast proteins (blue), Hogan proteins (yellow) and Hogan protein interfaces (red). Clear enrichment for nitrogen rich amino acids can be seen in the protein-RNA interfaces.



Supporting Figure S3: Comparison of physical interactions between RBPs and their targets in the positive set and 100 randomized negative sets. The data is shown using three different datasets for protein interactions: A) All complex interactions for budding yeast from BioGRID. B) Complex interactions from BioGRID with at least two lines of evidence. C) The Benschop consensus of complexes in budding yeast (Benschop, J.J. et al. (2010) A Consensus of Core Protein Complex Compositions for Saccharomyces cerevisiae. *Mol Cell*, **38**, 916-928).