# Predicting bacterial growth conditions from mRNA and protein abundances.

**Supplementary file**

Mehmet U. Caglar1, 2, 3, Claus O. Wilke3, 4, 5\*

1Department of Integrative Biology, The University of Texas at Austin, Austin, Texas, USA

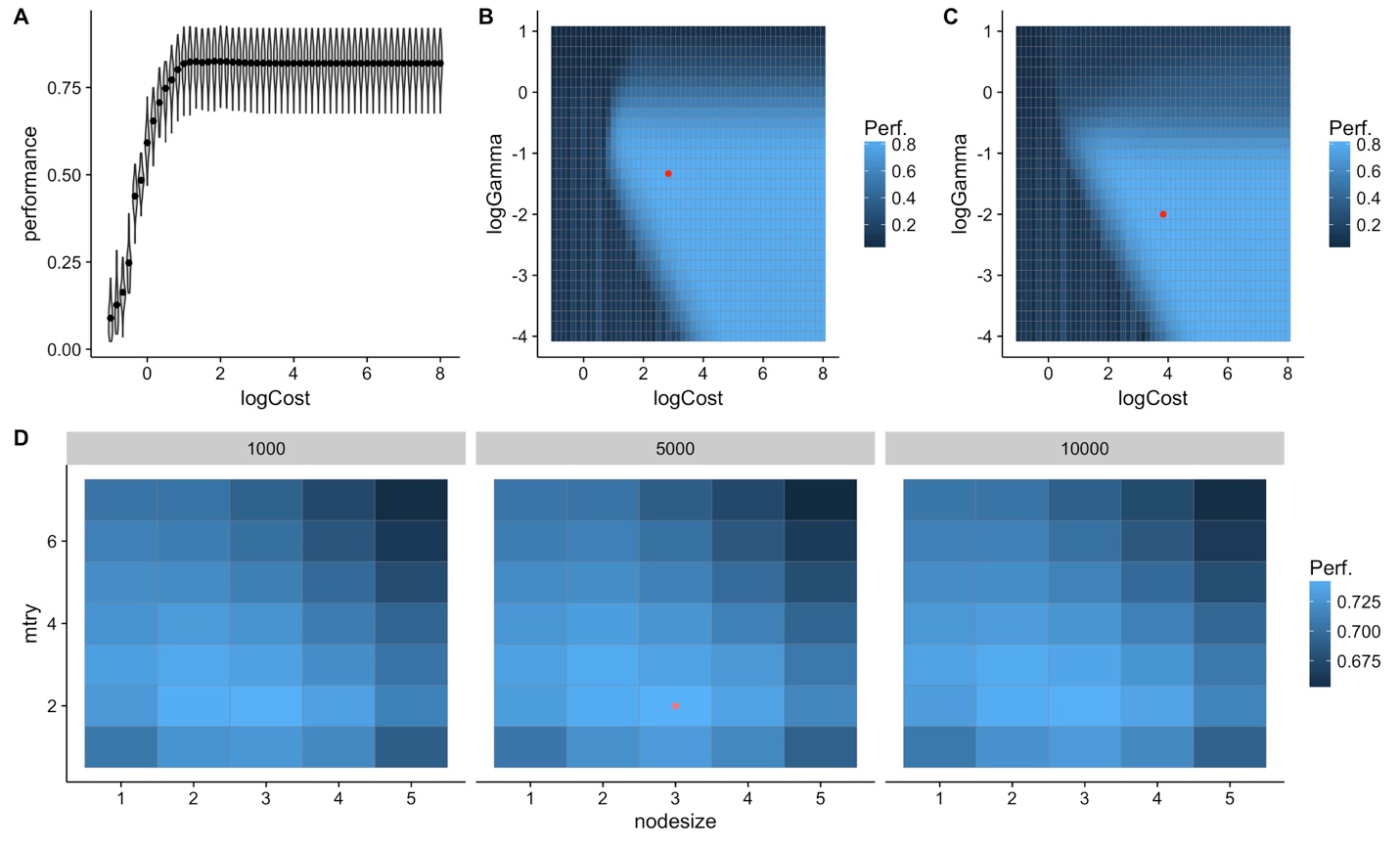
2Center for Computational Biology and Bioinformatics, The University of Texas at Austin, Austin, Texas, USA

3Institute for Cellular and Molecular Biology, The University of Texas at Austin, Austin, Texas, USA

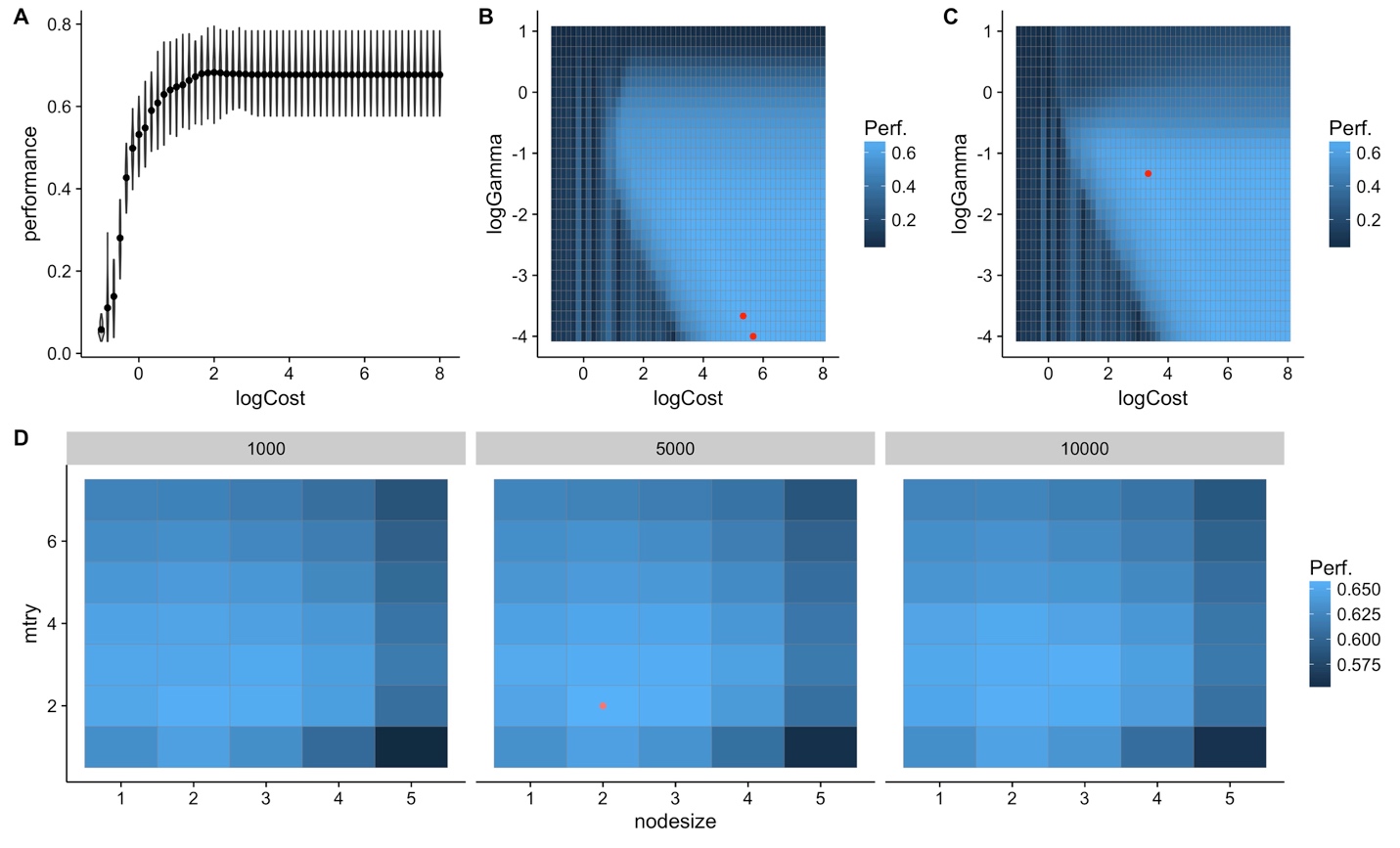
4Center for Systems and Synthetic Biology, The University of Texas at Austin, Austin, Texas, USA

5Department of Molecular Biosciences, The University of Texas at Austin, Austin, Texas, USA

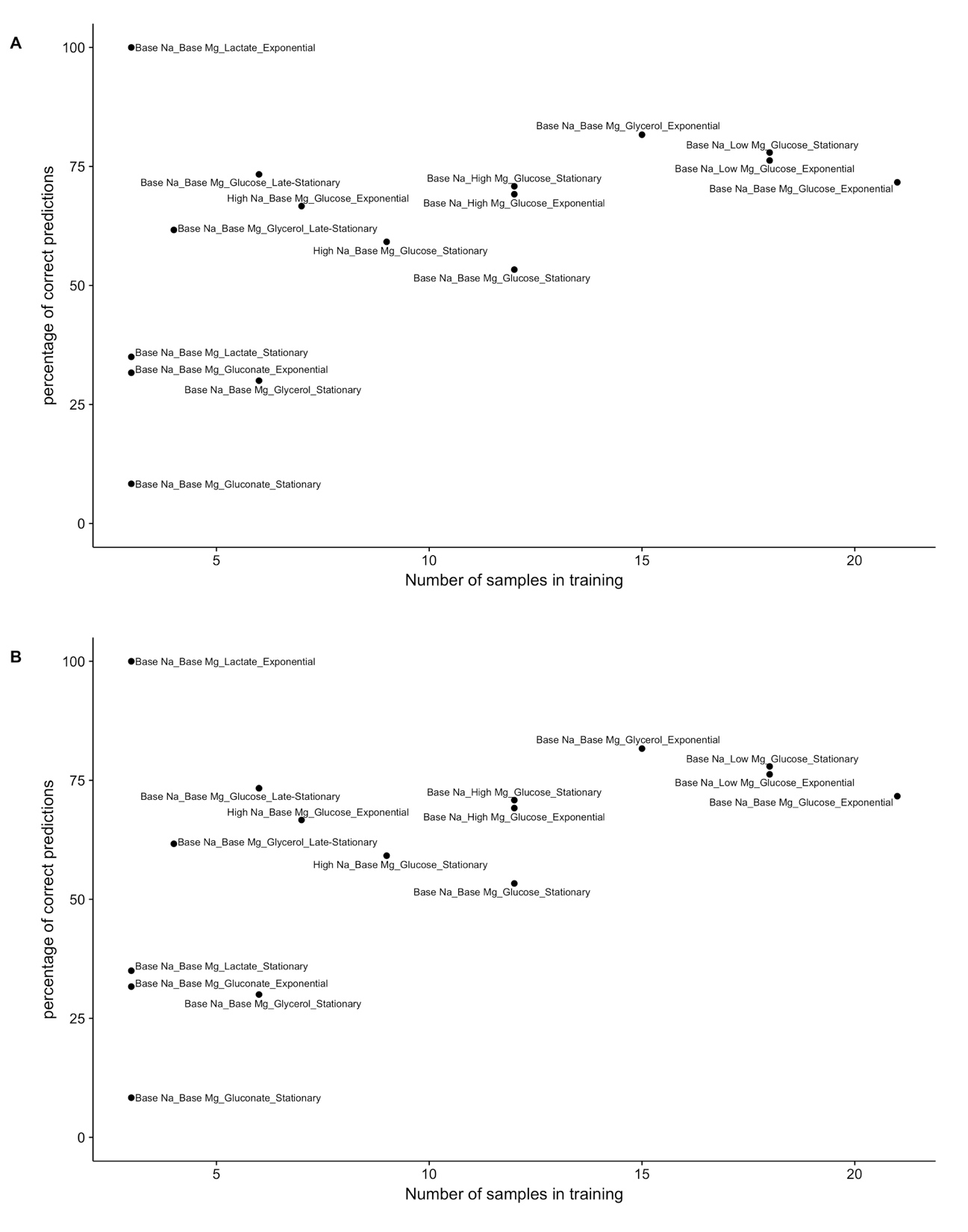
\*Corresponding author: [wilke@austin.utexas.edu](mailto:wilke@austin.utexas.edu) (COW)



**Supplementary figure 1.** The model parameter and error distribution for mRNA data. A. parameter and error distribution with SVM using linear kernel, B. parameter and error distribution with SVM using radial kernel, C. parameter and error distribution with SVM using sigmoidal kernel, D. parameter and error distribution with SVM using random forest. SVM with radial kernel gives the best results xx/60 of the independent training&tuning runs.



**Supplementary figure 2.** The model parameter and error distribution for protein data. A. parameter and error distribution with SVM using linear kernel, B. parameter and error distribution with SVM using radial kernel, C. parameter and error distribution with SVM using sigmoidal kernel, D. parameter and error distribution with SVM using random forest. SVM with sigmoidal kernel gives the best results xx/60 of the independent training&tuning runs.

**Supplementary figure 3**. The precision of individual conditions in tests for mRNA (A) and protein (B) datasets with respect to number of samples in training data.



**Supplementary figure 4.** The error count distribution for mRNA (A) and protein (B) confusion matrixes.



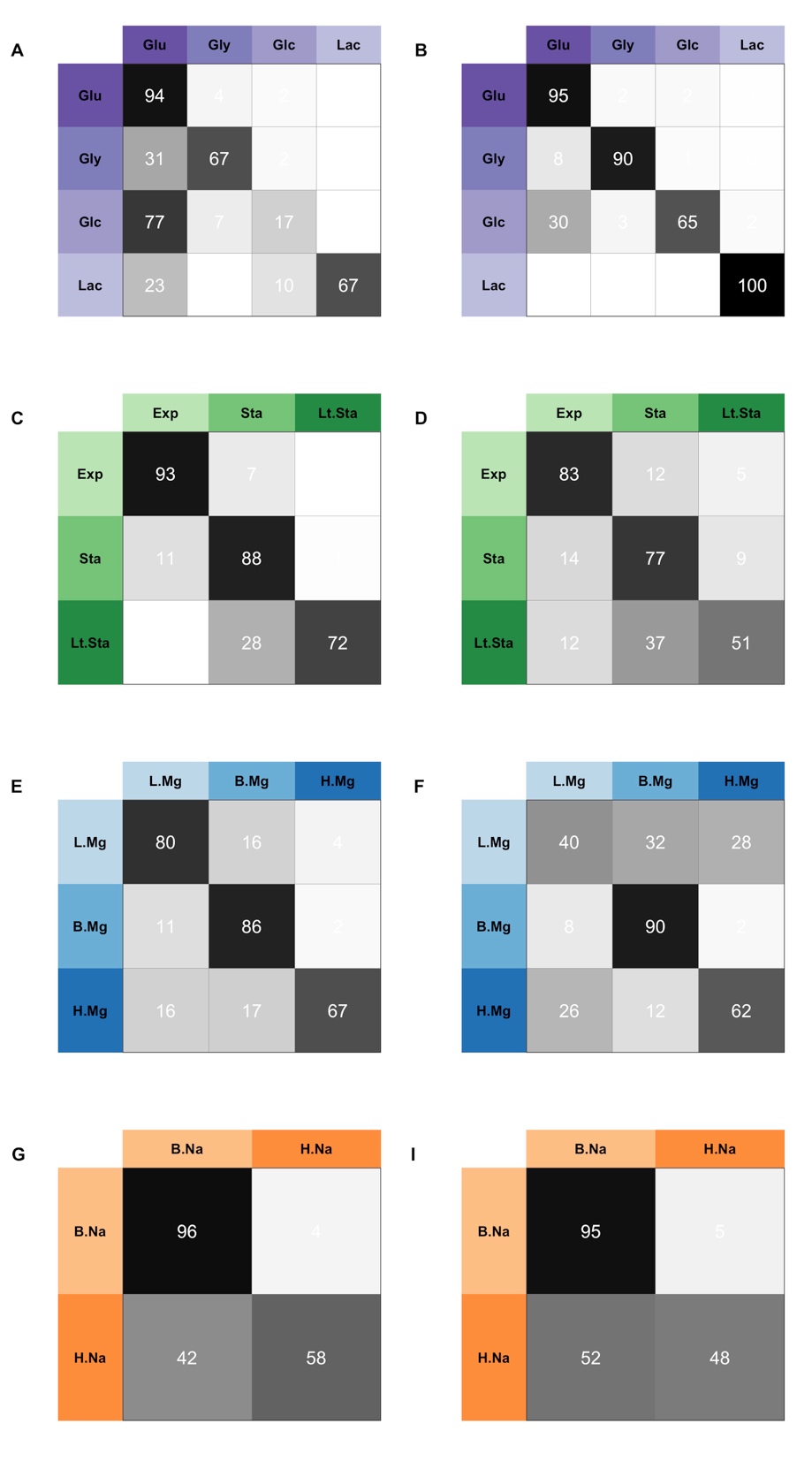
**Supplementary figure 5.** Complex predictions with intersection mRNA data. Corresponding multivariate F1 score is xx



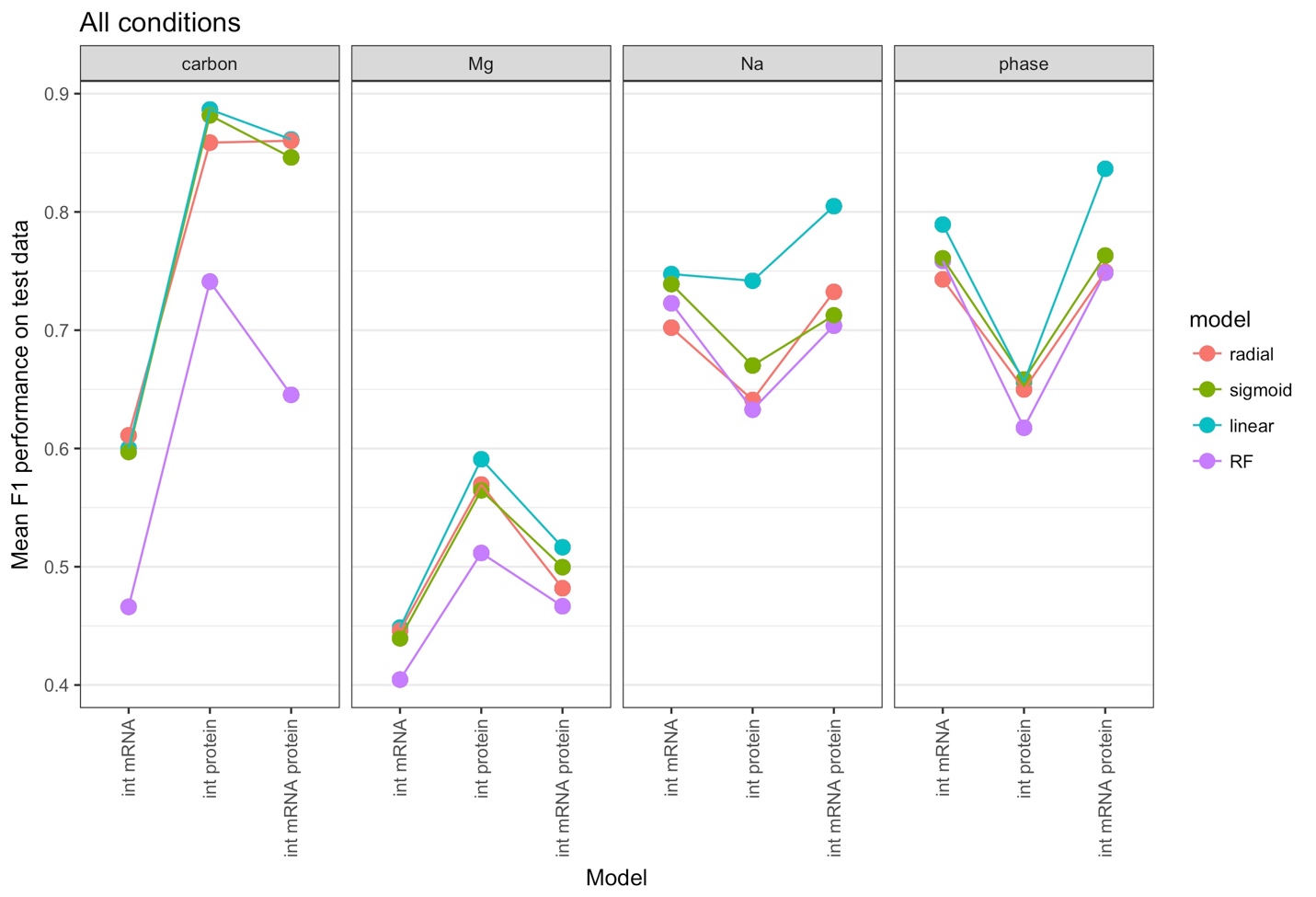
**Supplementary figure 6.** Complex predictions with intersection protein data. Corresponding multivariate F1 score is xx

****

**Supplementary figure 7.** Complex predictions with intersection combined mRNA and protein data. Corresponding multivariate F1 score is xx



**Supplementary figure 8.** Four distinctIndividual tests for mRNA and protein data. A. Carbon source for mRNA, B. carbon source for protein, C. growth phase for mRNA, D. growth phase for protein, E. Mg+2 levels for mRNA, F. Mg+2 levels for proteins, G. Na+1 levels for mRNA, H. Na+1 levels for proteins.



**Supplementary figure 9.** The performance change between intersection mRNA, intersection protein and combined datasets for individual tests.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Test** |  | **Predictions** | | | |
| **A** |  |  | **Na level** | **Mg Level** | **Carbon source** | **Phase** |
| **First approach** | **Glucose** | | base Na | high Mg | Glucose | Exponential |
| **Glycerol** | | base Na | high Mg | Glucose\* | Exponential |
| **Na 50 Stress** | | base Na\* | high Mg | Glucose | Exponential |
| **Stationary** | | base Na | base Mg | Glucose | Stationary\* |
| **Late stationary** | | base Na | base Mg | Glucose | Stationary\* |
|  | |  | |  |  |  |
| **B** |  |  |  |  |  |  |
| **Second approach** | **Glucose** | | base Na | base Mg | Gluconate | Exponential |
| **Glycerol** | | base Na | base Mg | Gluconate\* | Exponential |
| **Na 50 Stress** | | High Na\* | base Mg | Glucose | Exponential |
| **Stationary** | | base Na | base Mg | Glucose | Stationary\* |
| **Late stationary** | | base Na | base Mg | Glucose | Stationary\* |

**Supplementary table 1.** The performance of the protein model on external data. A) The gaps on test data is filled by the median values of the training data. B) The training data is trained with the subset of proteins that match with test data. Black cells indicate a correct prediction. The orange cells indicate a prediction in between. The red cells indicate a wrong prediction. “\*” indicates a true not-base value.