# Network Inference in Eukaryotes

**Stefan Altmann, Ralf Zimmer and Robert Küffner**
**Department of Informatics, Ludwig-Maximilians University, Munich, Germany**

## Introduction

Gene regulatory networks comprise transcription factor:target gene (TF:TG) interactions as basic building blocks. Such interactions can be discovered by different experimental techniques such as chromatin immunoprecipitation (ChIP). A range of computational methods have been proposed to complement experimentally derived interactions. Most inference methods depend on mutual information between the expression of TFs and their TGs (e.g. CLR [1], ARACNE [2]). These simple models can only be applied to prokaryotes as the precondition of mutual information is not satisfied for eukaryotes ([4]). Instead, a supervised machine learning (ML) setting utilizing known interactions is required, as expression dependencies between TFs and TGs are far more subtle than in prokaryotes.

## Methods

The inference of TF:TG interactions is transformed into a set of binary classification problems. For each TF, a dedicated local model (sirene approach, [3]) is trained to distinguish TGs of that TF from non-target genes. In contrast to most inference methods, we propose to apply a range of different machine learning methods (Decision Tree, SVM, Lasso, Elastic Net and Random Forest) to a range of different datasets (more than 3000 yeast microarrays) in a three-fold cross-validation setting. Performance is evaluated against three different yeast gold standards. Predictions derived from different methods and datasets are combined to improve interaction predictions.

## Results

We find that yeast interactions can be detected reliably by the sirene approach ([3]), which so far has only been applied to prokaryotes. Even if only one ML method and one dataset is used, the AUC of predictions is substantially above 0.5 (Fig. 1). Prediction performance steadily increases the more ML methods and expression datasets are integrated. We will also present results on why certain TFs (Fig. 2) or interactions are predicted better or worse by certain ML methods and on condition-specific regulation. Here, we examine how the subset of TGs that is regulated by a given TF depends on the experimental condition.

# References

[1] Jeremiah J. Faith, Boris Hayete, Joshua T. Thaden, Ilaria Mogno, Jamey Wierzbowski, Guillaume Cottarel, Simon Kasif, James J. Collins, and Timothy S. Gardner. Large-scale mapping and validation of escherichia coli transcriptional regulation from a compendium of expression profiles. *PLoS Biol*, 5(1):e8, Jan 2007.

[2] Adam A. Margolin, Ilya Nemenman, Katia Basso, Chris Wiggins, Gustavo Stolovitzky, Riccardo Dalla Favera, and Andrea Califano. Aracne: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics*, 7 Suppl 1:S7, 2006.

[3] Fantine Mordelet and Jean-Philippe Vert. Sirene: supervised inference of regulatory networks. *Bioinformatics*, 24(16):i76–i82, Aug 2008.

[4] Varun Narendra, Nikita I. Lytkin, Constantin F. Aliferis, and Alexander Statnikov. A comprehensive assessment of methods for de-novo reverse-engineering of genome-scale regulatory networks. *Genomics*, 97(1):7–18, Jan 2011.

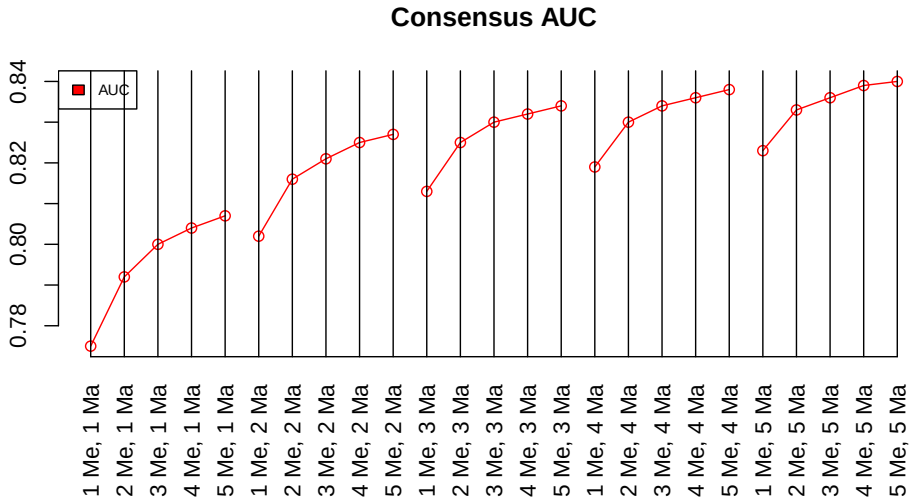


Figure 1: Prediction performance (AUC, area under the sensitivity specificity curve) for all combinations of methods (Me) and matrices (Ma).

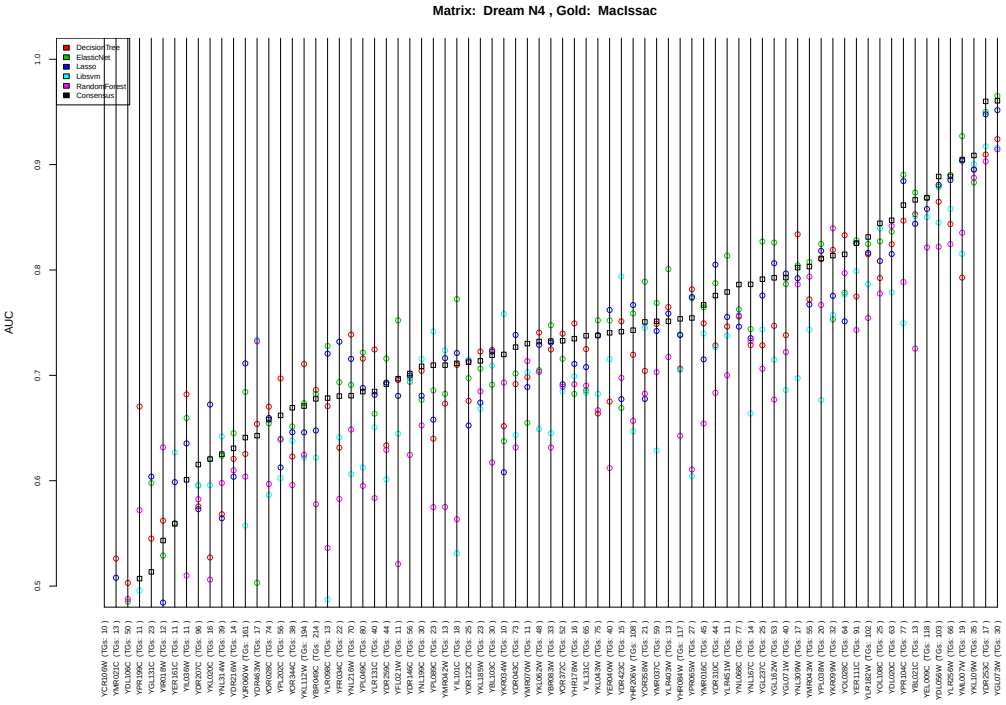


Figure 2: AUC values per transcription factor. For each transcription factor, consensus AUC values and the AUC values achieved by each of the five methods is plotted. In brackets the number of targets genes of the transcription factor is given.