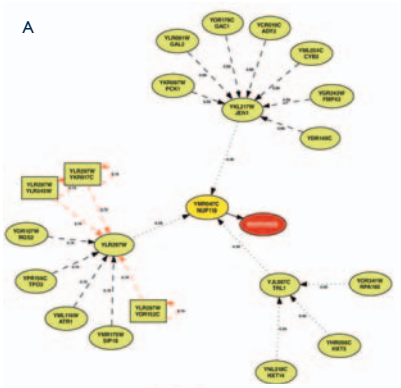
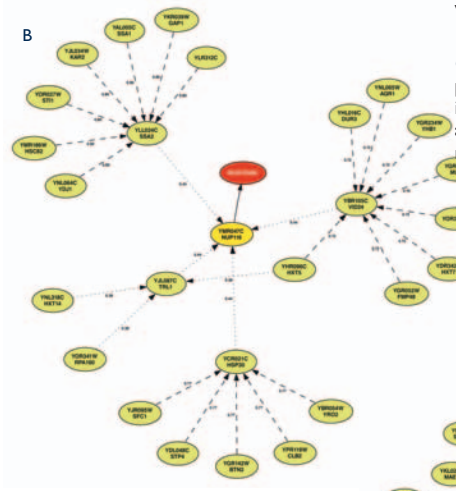
# Enhancing predicted gene interaction networks from microarray data using additional genomics data

## Integration of data with NetRaVE for enhanced microarray data analysis.

As microarray data is noisy and often sparse in nature, it would be beneficial to consider the wealth of genomics, bioinformatics and systems biological data available in public databases, and be able to add this during the analysis of array data. We used NetRaVE to combine array and proteomics data in constructing gene association networks.

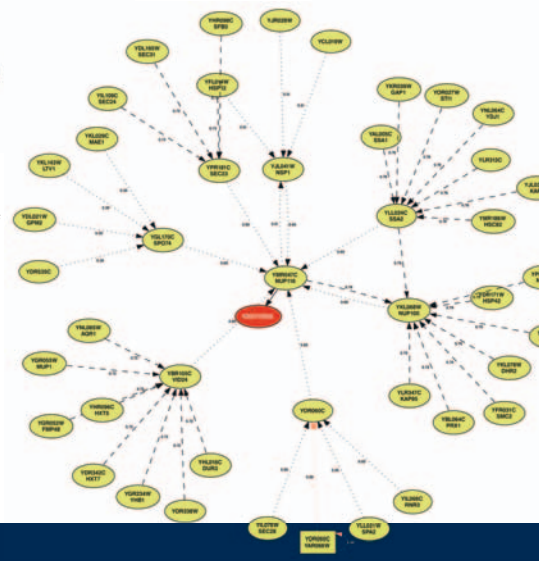



**networks - networks -networks**

*In the case of NUP116 we constructed networks by joining nodes to each other using either decision trees or linear regression. The treebased networks are shown here.*

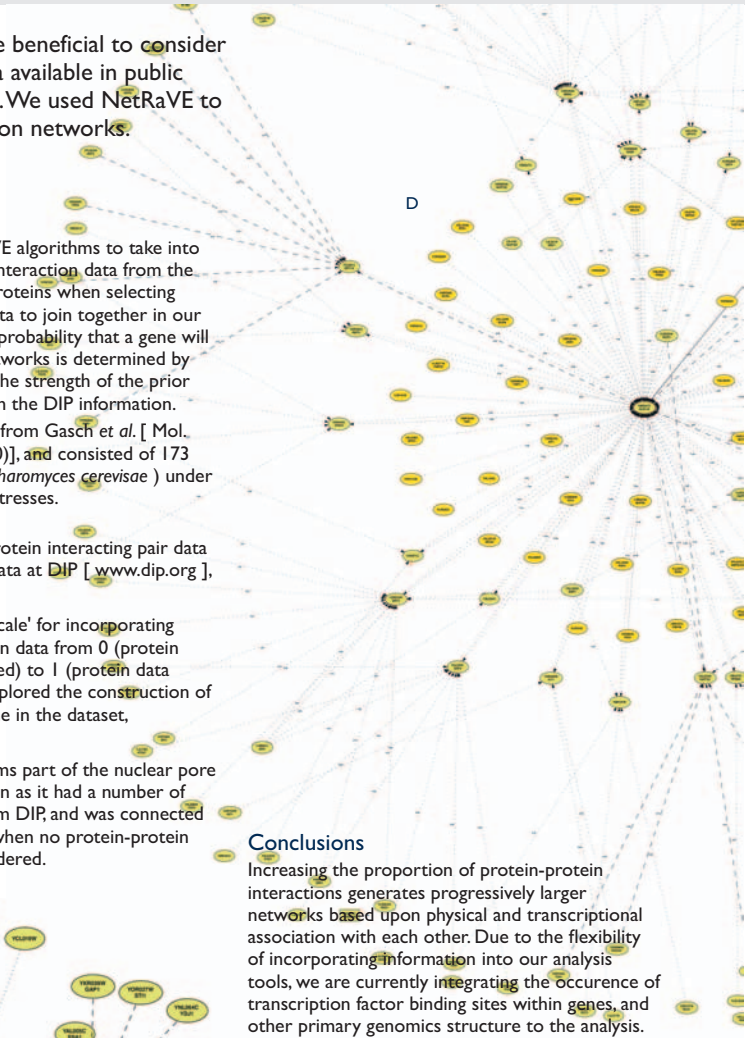
*In network **A**, we observe connections to several genes when there is no consideration of the protein-protein interaction data (ip=0). In network **B** (ip=0.2) we see the replacement of the subnets with other stronger correlated subnets, with several nodes representing proteins physically associated with the nuclear pore complex. This continues in network **C** (ip=0.4), and by network **D** (ip=0.8) we have over 300 nodes interconnected.*




### Methodology

We modified our NetRaVE algorithms to take into account protein-protein interaction data from the Database of Interacting Proteins when selecting genes from microarray data to join together in our association network. The probability that a gene will be excluded from our networks is determined by the array data, as well as the strength of the prior information obtained from the DIP information.

- Array data were taken from Gasch *et al.* [ Mol. Biol. Cell 11, 4241 (2000)], and consisted of 173 arrays from yeast (*Saccharomyces cerevisae* ) under various environmental stresses.
- Yeast binary protein-protein interacting pair data was extracted from the data at DIP [ www.dip.org ],
- By adjusting a 'sliding scale' for incorporating protein-protein interaction data from 0 (protein interactions not considered) to 1 (protein data always considered) we explored the construction of networks around one gene in the dataset,
- The gene NUP116 forms part of the nuclear pore complex. It was choosen as it had a number of interacting partners from DIP, and was connected to a network of genes when no protein-protein interactions were considered.




### Conclusions

Increasing the proportion of protein-protein interactions generates progressively larger networks based upon physical and transcriptional association with each other. Due to the flexibility of incorporating information into our analysis tools, we are currently integrating the occurence of transcription factor binding sites within genes, and other primary genomics structure to the analysis.

We have found that generating hypothesis from the networks and related literature (PubMed searches) to be very productive, and are currently seeking to further this research in model organisms with biological collaborators.

This application of NetRaVE is ment to add structure to the usually unordered gene lists obtained in looking for differentially regulated gene expression, as well as being an exploratory tool for array data. By incorporating different kinds of data we add value to the reults. Of course any hypothesis generated by our approach needs to be explored in the laboratory


For further information about Constructing Gene Networks from Microarray Data

| | | |
|---|---|---|
| Contact | Bill Wilson, PhD | Harri Kiiveri, PhD |
| Phone | +61 2 9325 3153 | +61 8 9333 6315 |
| Email | Bill.Wilson@csiro.au | Harri.Kiiveri@csiro.au |
| Web | www.cmis.csiro.au | |