Identifying molecular pathways underlying synaptic function & disease

A.J. Pocklington*, S.G.N. Grant** & J.D. Armstrong*

*Institute for Adaptive & Neural Computation, School of Informatics, University of Edinburgh, UK
* * Genes to Cognition Programme, The Wellcome Trust Sanger Institute, Hinxton, UK

Abstract

A major challenge in understanding the molecular basis of function and disease lies in uncovering relevant pathways and identifying key molecules for functional genomic study. We have developed a heuristic approach that uses network structure to identify and prioritise candidate molecules given a potentially limited amount of data. This was applied to the interaction network of the NMDA receptor complex (NRC), a postsynaptic signalling complex, where it was used to identify molecules with a potential role in synaptic plasticity, behaviour and disease (see figure).

Model system - the NRC

The NRC is a major component of the synapse proteome, processing information encoded in neural activity and orchestrating biochemical changes to the postsynaptic neuron in response. The ability of the NRC to induce changes at the synapse (synaptic plasticity) underlies its role in behaviour and disease. Nearly a third of all NRC genes/proteins have been linked to synaptic plasticity and behavioural learning in rodents, and human psychiatric disorders [1].

Algorithm (sketch)

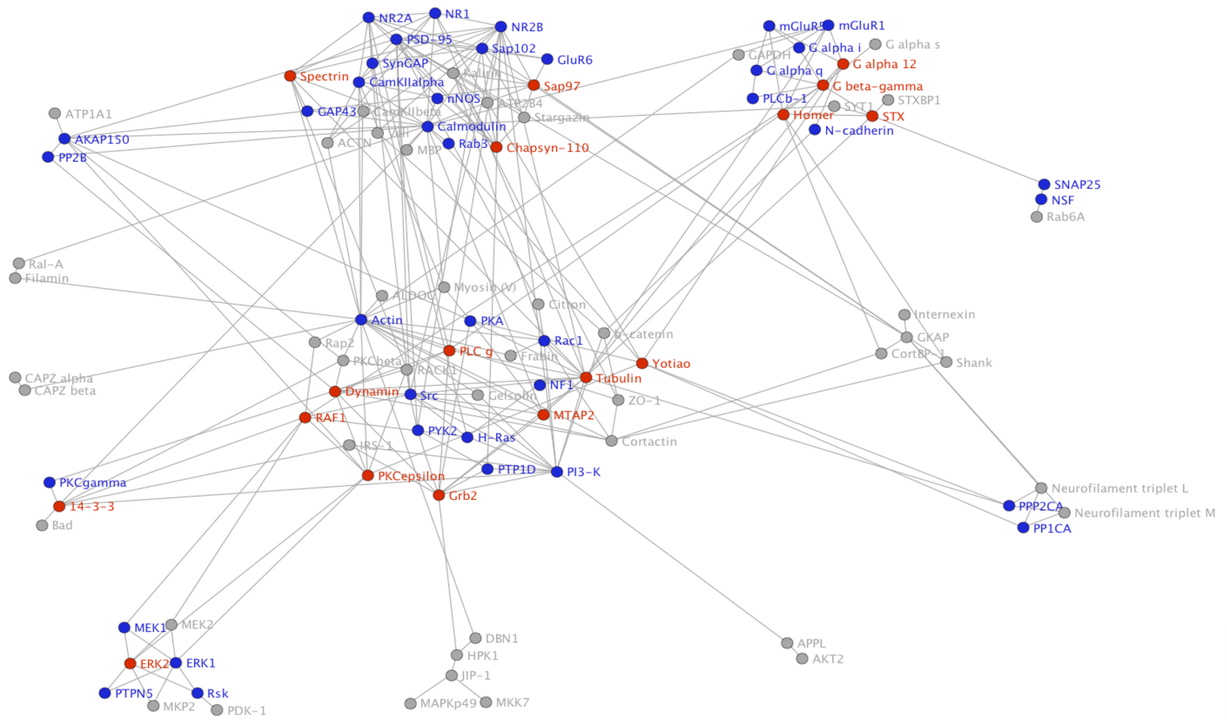
Our starting point is a molecular interaction network represented as an undirected graph. Within this network we wish to identify pathways involved in some biological process, for which we have (limited) experimental data.

a) Probability extrapolation

Given experimental data for molecule *i*, we can estimate *p(i)* - the probability that *i* is functionally relevant. In many cases, data will only be available for a subset of molecules, and we must extrapolate *p(j)* for each remaining molecule *j*. Using network topology to estimate the functional connection *f(i,j)* between *j* and each *i* for which we have experimental data, we defined *p(j)* in terms of a weighted sum over the *p(i)*:

$$p(j) \propto \Sigma f(i,j)\, p(i)$$

A simple measure of 'functional connection' is the (inverse) distance between *i* and *j*. However, as the average distance between molecules tends to be quite low in biological networks, this does not provide much discrimination. A more selective measure was obtained using vertex betweenness [2], which was found to give much better results.



The protein interaction network of the NRC, as detailed in [1]. Molecules with known involvement in synaptic plasticity are highlighted in blue, novel predictions in red.

b) Subset identification

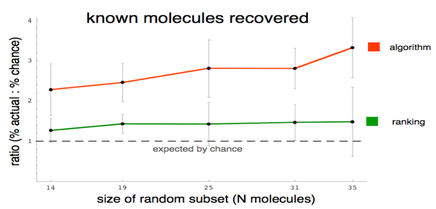
Simply ranking molecules by their *p(k)* does not account for their distribution within the network. We also wish to identify molecules that, despite a relatively low *p(k)*, play a significant role in mediating interactions between those with higher probabilities. Defining a topology-dependent score in terms of random walks between pairs of molecules, we used a simple search algorithm (similar to that of [3]) to identify high scoring sets.

Performance

The ability of the algorithm to make meaningful predictions was tested using the NRC network. From the set of NRC molecules with a known function/phenotype (e.g. involvement in synaptic plasticity), we took random subsets of various sizes. Using these as the basis for probability extrapolation (setting p(i) = 1 within the subset) and prediction, we evaluated the relevance and consistency of the results. The performance shown here (for synaptic plasticity) is representative of that for all phenotypes tested.

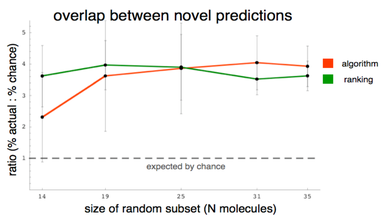
i) Identifies relevant molecules

Of the implicated molecules not used to make predictions, the algorithm consistently recovered over 50% (2-3 times more than expected by chance).



ii) Makes robust predictions

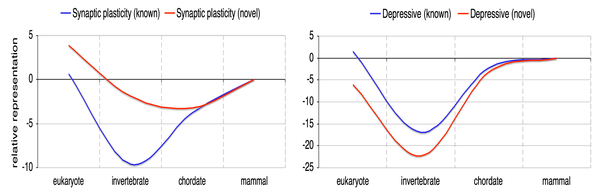
Novel candidates identified using different random subsets overlapped to a significant extent, typically by more than 60% (3-4 times that expected by chance).



Comparing the full algorithm (red lines in i and ii) to simple ranking based on extrapolated probabilities (green lines), it is clear that the subset identification step leads to significant enrichment of predictions with relevant molecules. It is also clear that only the full algorithm shows much improvement in performance as the size of the dataset increases.

Predictions mirror phylogenetic profile of implicated genes

In examining the evolutionary origins of the NRC, we have found that genes with certain phenotypes appear to be differentially represented at different stages of evolution (Emes et al, submitted). For example, genes linked to synaptic plasticity and depressive illness are under-represented amongst NRC genes present in invertebrates. Interestingly, novel predictions for both show similar general trends (see below).



Summary

The data presented suggests that the algorithm produces stable predictions and that it identifies relevant sets of molecules. Candidate molecules for various functions/phenotypes are now the subject of further analysis and evaluation.

References

[1] A.J. Pocklington, J.D. Armstrong, M. Cumiskey and S.G.N. Grant. Neurotransmitter receptor proteomes form modular networks with distributed functionality underlying plasticity and behaviour. *Molecular Systems Biology* doi:10.1038/msb4100041 (2006)

[2] M.E.J. Newman. A measure of betweenness centrality based on random walks. *cond-mat/0309045*

[3] T. Ideker, O. Ozier, B. Schwikowski and A.F. Siegel. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics* **18**, S233-S240 (2002)

AJP was supported by the Medical Research Council (UK), SGNG & JDA were supported by the Wellcome Trust Genes to Cognition programme.