Reconfiguration of Protein Interaction Networks during Nematode Development

Florian Klimm[1], Charlotte M. Deane[1], Jonny Wray[2], and Mason A. Porter[3]

[1] University of Oxford Woodstock Rd, Oxford OX2 6GG
[2] e-Therapeutics Plc 17 Blenheim Office Park, Long Hanborough, OX2 98LN
[3] University of California at Los Angeles Box 951555 Los Angeles, CA 90095-1555

1 Introduction

Protein interaction networks (PINs) allow the representation and analysis of biological processes in cells. Because cells are dynamic and adaptive, these processes change over time. One example of adaptive regulation is the change of gene expression, which may occur at very different time scales [1]: responses to environmental signals take minutes [2], and developmental changes take days in *C. elegans* [3] and years in humans [4]. This change in protein expression results in an altered protein abundance in an organism.

Thus far, research has focused either on the static PIN analysis or the temporal nature of gene expression. By analysing temporal PINs using multilayer networks [5], we want to link these efforts. The construction and analysis of temporal PINs gives insights into how proteins, individually and in their entirety, change their biological functions. In our investigation, we find that modular structure in the roundworm *C. elegans*' PIN changes during development. Using gene ontology (GO) terms, we connect this structural change with a reorganisation of biological functions. To our knowledge, our results represent the first direct identification of dynamic modular structure in PINs, despite having been hypothesised more than a decade ago [6].

2 Data sets

We use the mutlilayer network constructed in [7]. It consists of a total of $N = 4,792$ proteins. Interactions between them are aggregated from BioGrid [8] and other protein interaction databases. This gives a monolayer network of all protein interactions, as shown in Fig. 1. Each layer is then constructed as a subnetwork consisting of all proteins expressed at that developmental stage and all interactions between them. The gene expression information for six developmental stages (blastula, gastrula, embryo, nematode, prime adult, and life cycle) is extracted from the Bgee repository [9]. The layers consist of a variable number of nodes, ranging from 2,848 in the gastrula stage to 4,755 in the nematode stage.

After the construction of the layers, we connect them with interlayer edges of different strength ω. In this abstract, we illustrate results only for $\omega = 0.1$, but we will present results for multiple values in the oral presentation. We exclusively connect two nodes in layers of successive developmental stages and if they represent the same protein.

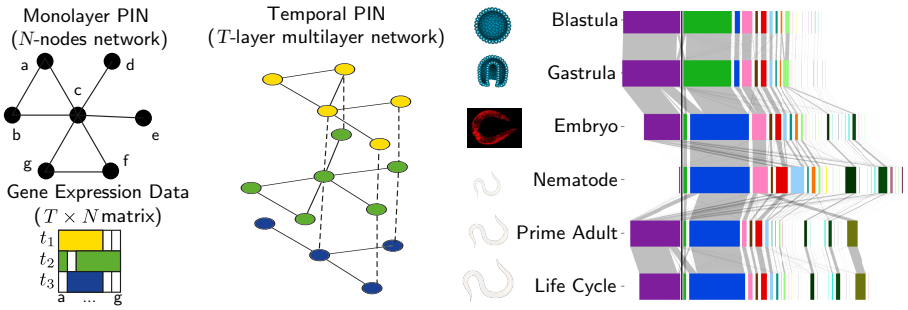

Fig. 1. Left: Temporal PIN links temporal gene expression with static PIN. Example given for $T = 3$ time points and $N = 7$ nodes. Right: Alluvial plots of the developmental modular structure of *C. elegans*' PIN . Rectangles represent modules of nodes with their width indicating the module's size. The most left rectangle in purple represent all proteins that are not expressed in a given developmental stage. The width of the gray lines indicate the overlap between modules in temporally adjacent layers and thus give a strength of transition from one developmental stage to the next.

3 Results

Because the functionality of biological processes change during the development, we suspect that modular structure also changes during development. To test this hypothesis, we detect and analyse the modular structure in the developmental PIN of *C. elegans*. We use GENLOUVAIN [10], a modularity optimisation method suited for multilayer networks, for the community detection.

The detected modular structure (see Fig. 1) consists of two facets: The network is organised in modules inside each layer and modules change over time (i.e., across layers). The modular structure inside each layer gives an indication of the functional organisation of the proteins at a given developmental stage. The modules vary in size and the number of modules in each layer ranges from eleven to more than twenty.

We use GO enrichment to test whether the detected structural modules consist of proteins with a mutual function. We use a significance level of 0.05 and use Bonferroni correction to take into account the problem of multiple comparisons, because we test the enrichment of more than 2×10^6 GO terms. The large modules tend to show enrichment for fairly broad terms, such as 'protein binding'; and the smaller modules, show more specific terms, such as 'embryo development' and 'proteasome complex'. This is consistent with earlier results, which show that, different GO terms tend to be enriched at different module sizes [11].

The modules are often enrichmed for many different GO terms at the same time. For example, module 9 (marked in light green) of blastula stage is enriched for approximately 50 terms. Amongst them are many terms that reflect different developmental processes like 'embryo development', 'larval development', 'hermaphrodite genitalia development', and 'reproduction of symbiont in host'.

We find that some parts of modular structure stay similar during the developmental, whereas others undergo considerable change. The reconfiguration of modular structure over time can give additional insights into the adaptive function of a cell. To give one example, we focus on module 5 (marked in red) of the nematode stage. GO enrich-

ment indicates that its dominant function is 'embryo development'. Its members are in three different modules at the next stage 'prime adult'. This suggests that this function may adapt and is now distributed across those three modules. To investigate this further, we separately examine the GO enrichment for each of these three groups of nodes that change their module. We find that all three of them have enriched 'embryo development', but one of them has a much stronger enrichment for 'ubiquitin-independent protein catabolic process via the multivesicular body sorting pathway'. This suggests that multivesicular body sorting pathway is connected to the embryo development.

Summary. We represent the PIN of the nematode *C. elegans* during development as a mutlilayer network. By investigating modular structure in this network, we detect a partial reconfiguration of its communities with development. Further comparison of the modular structure with gene ontology annotations hints at biological functions of the modules of proteins. By examing the structural change of modules from one developmental stage to the next we are able to detect modules that break apart or combine. This hints at functional change such as a strengthening of subfunctions.

References

1. Luis López-Maury, Samuel Marguerat, and Jürg Bähler. Tuning gene expression to changing environments: From rapid responses to evolutionary adaptation. *Nature Reviews Genetics*, 9(8):583, 2008.
2. Nir Yosef and Aviv Regev. Impulse control: Temporal dynamics in gene transcription. *Cell*, 144(6):886–896, 2011.
3. George L Sutphin and Matt Kaeberlein. Measuring *Caenorhabditis elegans* life span on solid media. *Journal of Visualized Experiments*, (27), 2009.
4. Urie Bronfenbrenner. The ecology of human development: Experiments by nature and design. *American Psychologist*, 32:513–531, 1979.
5. Mikko Kivelä, Alex Arenas, Marc Barthelemy, James P Gleeson, Yamir Moreno, and Mason A Porter. Multilayer networks. *Journal of Complex Networks*, 2(3):203–271, 2014.
6. Jing-Dong J Han, Nicolas Bertin, Hao Tong, Debra S Goldberg, et al. Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature*, 430(6995):88, 2004.
7. Pramod Shinde and Sarika Jalan. A multilayer protein–protein interaction network analysis of different life stages in *Caenorhabditis elegans*. *EPL (Europhysics Letters)*, 112(5):58001, 2015.
8. Chris Stark, Bobby-Joe Breitkreutz, Teresa Reguly, Lorrie Boucher, Ashton Breitkreutz, and Mike Tyers. BioGRID: A general repository for interaction datasets. *Nucleic Acids Research*, 34(suppl_1):D535–D539, 2006.
9. Frederic Bastian, Gilles Parmentier, Julien Roux, Sebastien Moretti, Vincent Laudet, and Marc Robinson-Rechavi. BGee: Integrating and comparing heterogeneous transcriptome data among species. In *Data Integration in the Life Sciences*, pages 124–131. Springer, 2008.
10. Lucas G. S. Jeub, Marya Bazzi, Inderjit S. Jutla, and Peter J. Mucha. A generalized louvain method for community detection implemented in MATLAB, version 2.1. http://netwiki.amath.unc.edu/GenLouvain, 2011-2014.
11. Anna CF Lewis, Nick S Jones, Mason A Porter, and Charlotte M Deane. The function of communities in protein interaction networks at multiple scales. *BMC Systems Biology*, 4(1):100, 2010.