Learning Gene Network Structure from Time Laps Cell Imaging in RNAi Knock-Downs

Henrik Failmezger [1+] and Paurush Praveen [2+] and Achim Tresch [1] and Holger Fröhlich [2]

[1] *Computational Biology and Regulatory Networks, Max-Planck Institute for Plant Breeding Research, Carl-von-Linne-Weg 10, 50829 Cologne, Germany*

[2] *Algorithmic Bioinformatics, Bonn-Aachen International Center for IT, Rheinische Friedrich-Wilhelms-Universität Bonn, Dahlmannstr. 2, 53113 Bonn, Germany*

frohlich@bit.uni-bonn.de

[+] *These authors contributed equally.*

Abstract: As RNA interference is becoming a standard method for targeted gene perturbation, computational approaches to reverse engineer parts of biological networks based on measure-able effects of RNAi become increasingly relevant. The vast majority of these methods use gene expression data, but little attention has been paid so far to other data types. Here we present a method, which can infer gene networks from high-dimensional phenotypic perturbation effects on single cells recorded by time-lapse microscopy. We use data from the Mitocheck project to extract multiple shape, intensity and texture features at each frame. Features from different cells and movies are then aligned along the cell cycle time. Subsequently we employ Dynamic Nested Effects Models (dynoNEMs) to estimate parts of the network structure between perturbed genes via a Markov Chain Monte Carlo approach. Our simulation results indicate a high reconstruction quality of this method. A reconstruction based on a 22 gene knock-downs yielded a network, where all edges could be explained via the biological literature. The implementation of dynoNEMs is part of the Bioconductor R-package *nem*.

1 Introduction

The availability of large RNAi screens has raised the interest in computational approaches for reverse engineering parts of biological networks from measureable effects of targeted gene perturbations. Most existing methods make use of gene expression data. The few attempts to reverse engineer gene networks from phenotypic data include [BACP07], who rely on hierarchical clustering of

static images, and [KDZ+09] who use a probabilistic graphical model for only one binary phenotypic variable in static images. To our knowledge, there is yet no method for the inference of networks from time lapse microscopy based on large numbers of statistical image features.

Nested Effects Models (NEMs) are a class of probabilistic graphical models that have been introduced originally by [MBS05] and extended substantially later on by several other authors. In NEMs indirect, high-dimensional down-stream effects of multiple single-gene knock-downs are studied. NEMs allow for inferring the signaling flow between these perturbed genes on a transcriptional as well as non-transcriptional level based on the measured intervention effects. [ASJ+09] and [FPT11] extended the theory of NEMs to time series data, and applied it to infer parts of a transcriptional network involved in murine stem cell development. Originally NEMs assumed downstream effects to be measured via gene expression profiling, but here we use phenotypic image features from movies instead. Our movies were taken from the Mitocheck database [NWH+10], in which ~20,000 human genes were silenced via RNAi and subsequently screened for cell cycle defects. We use Dynamic Nested Effects Models [FPT11] to estimate the network between perturbed genes based on the dynamic response of the phenotype along the cell cycle. The inference is based on a Markov Chain Monte Carlo (MCMC) algorithm. The whole approach consisting of image feature extraction, estimation of perturbation effects and network estimation via dynoNEMs is called MovieNEM and described in detail in our paper [FPTF13]. A schematic overview about our method is shown in Figure 1.

2 Main Results

2.1 Simulations

In our paper we conducted extensive simulations on reconstruction of randomly selected sub-graphs of KEGG signaling pathways using simulated phenotypic features. In general we observed a very high accuracy of our method, which was notably dependent on the network size and network topology. Very densely connected networks with many loops appeared to be harder to learn than acyclic graphs. This can be explained by the fact that with many loops it is more unlikely to observe a time delayed nested effects structure, which is exploited by our method. We also investigated the influence of uninformative features on our method. Here we observed a highly robust behavior of our method, which underlines the success of the automated feature selection mechanism, which is inbuilt in MovieNEM.

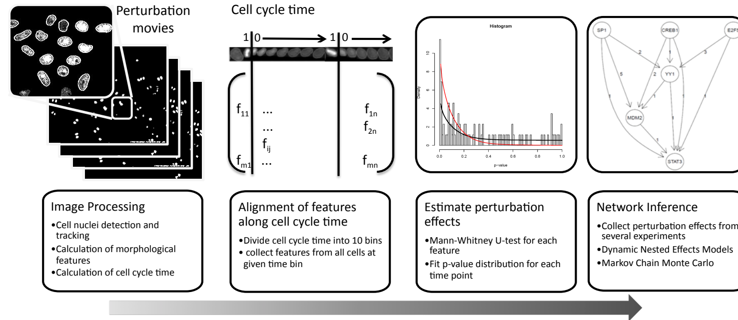

Figure 1: Overview about MovieNEM: Individual movies are first fed into an image processing pipeline consisting of four steps: (1) cell nuclei detection in the individual movie frames; (2) tracking of the nuclei over time; (3) calculation of morphological features and (4) calculation of cell cycle time. After image processing features are grouped according to the binned cell cycle time. This allows for estimating time-wise perturbation effects. Several movies, each showing one perturbation, are processed in this way and the perturbation likelihoods collected along the binned cell cycle time axis. This allows for applying Dynamic Nested Effects Models to infer the network between perturbed genes via Markov Chain Monte Carlo.

2.2 Application to Movie Data

In our paper we applied MovieNEM to infer a network between 22 genes with significant phenotype. These 22 genes are mainly involved into cell cycle, transcriptional regulation and cell differentiation. We looked at the network of edge-wise posterior expectations, which scored better than 1000 random S-gene permutations. All 122 edges could be mapped to known literature pathways. On the other hand and not very surprisingly the literature mentions some additional interactions, which could not be observed in our estimated network. This can have two reasons: Either the additional literature known interactions exist in reality, but MovieNEM could not infer them or they do not exist in HeLa cells and are hence not inferred. Notably, we can only infer interactions between genes that show a clear phenotypic knock-down effect.

3 Conclusion

In our paper we have shown that it is possible to learn pathway structures from phenotypic perturbation effects recorded in time-lapse movies. At the heart of

the method lies the extraction of morphological features yielding measurable differences in cell phenotypes. We have developed a method to quantify these differences such that an extended version of the dynoNEM method [FPT11] is applicable. We have developed a novel Markov Chain Monte Carlo sampler for network structure learning in order estimate the posterior likelihood of each interaction. Our method allows for the inclusion of prior knowledge in a Bayesian fashion. In summary MovieNEM offers an approach to exploit the rich information that is present in phenotypic RNAi screening data using image based techniques.

References

[ASJ+09] Benedict Anchang, Mohammad J Sadeh, Juby Jacob, Achim Tresch, Marcel O Vlad, Peter J Oefner, and Rainer Spang. Modeling the temporal interplay of molecular signaling and gene expression by using dynamic nested effects models. *Proc Natl Acad Sci U S A*, 106(16):6447–6452, Apr 2009.

[BACP07] Chris Bakal, John Aach, George Church, and Norbert Perrimon. Quantitative morphological signatures define local signaling networks regulating cell morphology. *Science*, 316(5832):1753–1756, Jun 2007.

[FPT11] H. Fröhlich, P. Praveen, and A. Tresch. Fast and Efficient Dynamic Nested Effects Models. *Bioinformatics*, 27(2):238–244, 2011.

[FPTF13] Henrik Failmezger, Paurush Praveen, Achim Tresch, and Holger Fröhlich. Learning gene network structure from time laps cell imaging in RNAi Knock downs. *Bioinformatics*, pages 1534 – 1540, May 2013.

[KDZ+09] Lars Kaderali, Eva Dazert, Ulf Zeuge, Michael Frese, and Ralf Bartenschlager. Reconstructing signaling pathways from RNAi data using probabilistic Boolean threshold networks. *Bioinformatics*, 25(17):2229–2235, Sep 2009.

[MBS05] Florian Markowetz, Jacques Bloch, and Rainer Spang. Non-transcriptional pathway features reconstructed from secondary effects of RNA interference. *Bioinformatics*, 21(21):4026–4032, Nov 2005.

[NWH+10] Beate Neumann, Thomas Walter, Jean-Karim Heriche, Jutta Bulkescher, Holger Erfle, Christian Conrad, Phill Rogers, Ina Poser, Michael Held, Urban Liebel, Cihan Cetin, Frank Sieckmann, Gregoire Pau, Rolf Kabbe, Annelie Wünsche, Venkata Satagopam, Michael H A. Schmitz, Catherine Chapuis, Daniel W. Gerlich, Reinhard Schneider, Roland Eils, Wolfgang Huber, Jan-Michael Peters, Anthony A. Hyman, Richard Durbin, Rainer Pepperkok, and Jan Ellenberg. Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature*, 464(7289):721–727, Apr 2010.