Incorporating Interaction Networks into the Determination of Functionally Related Hit Genes in Genomic Experiments with Markov Random Fields

Sean Robinson*†[1,2], Jaakko Nevalainen , Guillaume Pinna , Anna Campalans , J. Pablo Radicella , and Laurent Guyon‡[1]

[1]Universite Grenoble Alpes/Biosciences and Biotechnology Institute of Grenoble, CEA – Université Grenoble Alpes, CEA – France
[2]The University of Turku – Finland

Abstract

High-throughput genomic experiments score thousands of genes and subsequent analysis identifies 'hits' (genes with significant scores) that are the most promising candidates for further investigation. However, well known issues can lead to both false positive and false negative gene hit identification, such as for example in RNAi screens with 'potential siRNA efficiency' and 'off-target effects' among others. A network based approach to address these problems assumes the 'guilt by association' principle in the corresponding protein-protein interaction (PPI) network as a means to determine network based gene hits.
Inspired by their widespread use for digital image segmentation, we present a network based gene scoring method using a Markov random field (MRF). We compare and contrast our MRF method against a number of previously proposed network based scoring methods such as Knode (PLOS Computational Biology, 2014), BioNet (Bioinformatics, 2010), NePhe (BMC Genomics, 2009) and NEST (Genome Biology, 2015). The two major advantages of our MRF method are that it allows for multivariate scores on the genes (e.g. multiple siRNA) as well as multiple hit classes beyond binary 'hit'/'non-hit' (e.g. positive and negative phenotypes). Using our MRF method we found improved hit identification in independent simulated experiments as well as new insights in a previously analysed lymphoma dataset comprising both differential expression and survival data. In our motivating RNAi screening data we found pertinent network based hits that were otherwise unidentified, allowing for the formation of further relevant hypotheses. In the investigation of data from genomic experiments with an associated PPI network, our MRF method gives a more beneficial way to find network based hit genes as against previously proposed methods.

*Speaker
†Corresponding author: sean.j.robinson@utu.fi
‡Corresponding author: laurent.guyon@cea.fr