miTALOS: Analyzing the tissue-specific regulation of signaling pathways by human and mouse microRNAs

Andreas Kowarsch, Martin Preusse, Carsten Marr, Fabian Theis
Institute for Bioinformatics and Systems Biology, Helmholtz Zentrum München, 85764 Neuherberg, Germany

Introduction

MicroRNAs regulate cellular signal transduction via controlling mRNA levels and thus tuning protein abundance[1]. The functional impact of microRNAs is usually assessed by identifying pathways with enriched targets. However, this measure ignores the topology of the underlying network and is biased towards large target numbers.

We propose a novel measure that is independent of pathway size and takes network structure into account. Our proximity score calculates average distances between targets of microRNAs in signaling networks and compares this values with a randomized null model. We applied the proximity score to a published data set of microRNA targets in mouse brain that is based on high-throughput sequencing of RNAs isolated by immunoprecipitation of crosslinked Ago-RNA complexes (HITS-CLIP)[2]. The proximity measure reveals associations with a smaller number of targets than the enrichment approach. This indicates that both methods detect alternative forms of microRNA control.

To make the new score available for the research community, we created the miTALOS webserver (http://mips.helmholtz-muenchen.de/mitalos/). Integrating five different microRNA target prediction tools and two different signaling pathway resources (KEGG and NCI), miTALOS computes microRNA-pathway associations with the proximity score and the standard enrichment method. As an additional feature, miTALOS considers the tissue-specific expression signatures of microRNAs and target transcripts to improve the analysis of microRNA regulation in biological pathways. Multiple microRNAs can be combined in a single analysis to highlight their combinatorial effects. A graphical visualization of microRNA targets is provided to illustrate their respective pathway context.

Methods

MicroRNA data

Human and mouse microRNAs were extracted from the miRBase database, which is a collection of published microRNA sequences and annotation. As there is strong evidence that microRNAs can act in concert with each other, miTALOS also provides a list of predefined microRNA clusters. MicroRNA clusters are defined as a set of microRNAs, where each member is having at least one other member of the same cluster within 5kb distance according to chromosomal locations. Chromosomal positions of all human and mouse microRNAs were obtained from the mirBase database.

MicroRNA target prediction

The miTALOS web resource uses several target prediction methods to infer microRNA target transcripts: TargetScanS, RNA22, PicTar, PiTa, TargetSpy. Sethupathy et al.[3] showed that the intersection of the prediction tools can yield improved specificity with only a marginal decrease in sensitivity relative to any individual algorithm. miTALOS can handle this issue by generating intersections from at least two prediction methods.

Tissue expression profiles

MicroRNA and their corresponding target transcripts show a highly tissue-specific expression pattern.

We used the tissue atlas provided by Su et al.[4] to filter potential microRNA targets in a specific tissue. The human and mouse data was downloaded from the NCBI Gene Expression Omnibus (GEO) and the processed data was used. We mapped the predicted microRNA target transcripts on the tissue atlas and considered a transcript as expressed in a specific tissue, if either one replicate has a present call or both show at least a marginal call, similar to the method used by[5].

Signaling pathways

For the functional analysis of microRNA-pathway associations, miTALOS offers two different resources of signaling pathway. All non-metabolic pathways for human and mouse were integrated from the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2008). The KEGG Pathway database is a collection of manually curated pathway maps for various genomes. For the analysis of human microRNA regulation, we also included signaling pathway information from the National Cancer Institute Pathway Interaction Database (NCI PID) (Schaefer et al., 2009). NCI PID is a manual collection of biomolecular interactions and key cellular processes assembled into signaling pathways. The database is curated by Nature Publishing Group editors and reviewed by experts in the field.

Enrichment score

The identification of microRNA-pathway associations by miTALOS is obtained using two different approaches. A first approach use the number of target genes in a specific pathway to calculate an enrichment score. Here, we assume that microRNAs target specific pathways to influence specific functions of the cell by the sheer number of target transcript. Calculating the enrichment of targets TP i in a pathway i with Pi proteins leads to an enrichment score E, which has been used in previous studies:

$$E = \frac{T_{Pi}/P_i}{T_P/P}$$

where TP_i is the number of targets in pathway i, P_i is the number of all proteins in pathway i, TP is the number of all targets in all pathways and P the number of all protein in the KEGG or NCI PID pathways. The significance is obtained by Fisher's exact test, corrected by the Benjamini-Hochberg procedure.

Proximity score

We assume that some microRNAs target signaling cascades in a proximal manner. To reveal pathways with proteins that function in a proximal manner and are targeted by the same microRNA, we introduce a proximity measure P. Let us consider a pathway i with Pi proteins and TP_i targets of a specific microRNA, we can determine the distances dxy between all $TP_i * T(P_i - 1)/2$ pairs of targets x, y ($x \neq y$) in the corresponding signaling pathway. To condense this set of distances for a microRNA pathway pair into a real number in [0, 1], we calculate the minimal distance for each target x. The proximity score P is then defined as the mean of all minimal distances $< d_{xy} >$ as the power of base α. The proximity score P is the defined as:

$$P = 1 - < \alpha^{-d_{xy}} >_{xy} .$$

The base α can be chosen appropriately to ensure a reasonable separation of the distances occurring in the network. Based on the observed distance scores, we chose α = 1.1. In order to obtain significant microRNA-pathway associations we perform random sampling. For each pathway i and specific number of targets, we randomly choose 10.000 times microRNA targets and calculate the corresponding proximity scores P. These samplings are then used to calculate the p-values by counting the number of proximity scores that are less than the original score divided by 10.000. Final p-values were then corrected by the Benjamini-Hochberg procedure.

Conclusion

Given the increasing amount of evidence that microRNAs have an important impact on signaling pathways[6,7], the proximity measure is a useful tool to infer systematical insights into microRNA-mediated regulation. miTALOS provides a substantial support to the research community by identifying tissue specific microRNA-pathway associations. More generally, we think that the concept of proximity can serve as a powerful tool to identify patterns in networks beyond microRNA regulation in signal transduction. Our tool might generate useful hypothesis beyond the commonly used enrichment method e.g. for drug targets in metabolic networks[8], disease genes in signaling pathways, or data sets of differentially expressed genes stem cell differentiation.

References

1. Baek, D. et al. The impact of microRNAs on protein output. *Nature* **455**, 64-71 (2008).

2. Chi, S.W., Zang, J.B., Mele, A. & Darnell, R.B. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* **460**, 479-86 (2009).

3. Sethupathy, P., Megraw, M. & Hatzigeorgiou, A.G. A guide through present computational approaches for the identification of mammalian microRNA targets. *Nature methods* **3**, 881-6 (2006).

4. Su, A.I. et al. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 6062-7 (2004).

5. McClintick, J.N. & Edenberg, H.J. Effects of filtering by Present call on analysis of microarray experiments. *BMC bioinformatics* **7**, 49 (2006).

6. Cui, Q., Yu, Z., Purisima, E.O. & Wang, E. Principles of microRNA regulation of a human cellular signaling network. *Molecular systems biology* **2**, 46 (2006).

7. Inui, M., Martello, G. & Piccolo, S. MicroRNA control of signal transduction. *Nature reviews. Molecular cell biology* **11**, 252-63 (2010).

8. Barabási, A.-L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nature Reviews Genetics* **12**, 56-68 (2011).