Motif enhanced ensemble network inference prediction

## Problem introduction

With the era of high-throughput data, the amount of gene expression, miRNA targeting, and methylation data increased while the understanding of these mechanisms lacked behind. Currently it is known that transcription factors and miRNA targeting affect and regulate gene expression. Regulatory epigenetic layers as methylation or histone modification appear to have a silencing or stimulating effect on gene expression. Histone modifications come in many variants, which induce different effects based on what histone modifications are at play, further dysregulating genetic expression leading to altered phenotypes such as cancer. Little detail is known how these regulating layers work together, however, using network inference techniques and motif analysis these could be clarified and used to treat or avoid alternate phenotypes.[1,2]

## Aim

In this proposal the aim is to estimate the improvement of network inference prediction using network motifs and ensemble network inference methods.

## Study design

Data is obtained from a dedicated network inference challenge, the HPN-DREAM breast cancer network inference challenge (https://www.synapse.org/HPN\_DREAM\_Network\_Challenge). This dataset consists out of *in sillico* simulations of a model described in Chen et al.[3] Therefore, the network inference can be determined and compared to the correct model. The simulated model data will be analyzed in R using the following network inference techniques; Bayesian networks, correlation, regression and mutual information.[1] These methods can be used together to achieve ensemble knowledge, increasing predictive power.[4] By comparing the resultant network with the original network, the overlap can be determined which indicates the performance of the ensemble inference methods. In addition, statistically significant overrepresented topological patterns, network motifs, will be used to compare if certain motifs are conserved between networks.[2,5] Using this information of conserved motifs, it is possible to increase the initial prediction of the inferred network.

## Context within the course

In this course several aspects were discussed, such as network inference and network motifs. Network inference methods can be used for many applications; however, databases are often used to explain inference based on literature. This information should be used to enhance the inference prediction based on obtained datasets, possibly giving new insight to a well known network. Network motifs are statistically significant overrepresented topological patterns; thus, these should be statistically significant for a (biological) reason. The causality of this significance is believed to be the underlying mechanism of the data, which could be used to enforce the network inference prediction.

[1] Albert R. (2007). Network inference, analysis, and modeling in systems biology. The Plant cell, 19(11), 3327–3338. doi:10.1105/tpc.107.054700

[2] Ahnert, S. E., & Fink, T. M. (2016). Form and function in gene regulatory networks: the structure of network motifs determines fundamental properties of their dynamical state space. Journal of the Royal Society, Interface, 13(120), 20160179. doi:10.1098/rsif.2016.0179

[3] W. W. Chen, B. Schoeberl, P. J. Jasper, M. Niepel, U. B. Nielsen, D. A. Lauffenburger, and P. K. Sorger, “Input-output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data.,” Mol. Syst. Biol., vol. 5, no. 239, p. 239, Jan. 2009.

[4] Guo, S., Jiang, Q., Chen, L., & Guo, D. (2016). Gene regulatory network inference using PLS-based methods. BMC bioinformatics, 17(1), 545. doi:10.1186/s12859-016-1398-6

[5] Wong, E., Baur, B., Quader, S., & Huang, C. H. (2012). Biological network motif detection: principles and practice. Briefings in bioinformatics, 13(2), 202–215. doi:10.1093/bib/bbr033