Motif enhanced ensemble network inference prediction

## Problem introduction

With the era of high-throughput data, the amount of gene expression, microRNA expression, and methylation data has increased tremendously while the understanding of these regulatory mechanisms is lagging behind. (Epi)genetic dysregulation can lead to altered phenotypes such as cancer, therefore, understanding the inference of these regulatory mechanisms is essential to treat or cure altered phenotypes. Little is known on how these regulating layers work together in detail [1,2]. Methods to infer regulatory mechanisms from experimental and modelled datasets have been studied greatly, and many methods have been developed. However, each method has its own strengths and weaknesses. By combining these methods in an ensemble fashion, the strengths can be combined which creates a better initial prediction of the inference in the network. To further improve the prediction, statistically significant overrepresented topological patterns (network motifs) are used as an additional filter to extract biologically-derived significant patterns found in the inferred networks.

## Aim

This project aims to estimate the improvement of network inference of regulatory networks by using network motifs.

## 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 was generated using *in sillico* simulations of a model described in Chen et al. [3]. This means that the inferred networks can always be compared to the correct, original model. The simulated model data will be analyzed in R using the following network inference techniques; Bayesian networks, correlation, regression and mutual information [1]. Each inferred network will be analyzed on network motifs using the ‘iGraph’ package in R. The resultant motifs possibly indicate which part of the inferred network is more likely to be coherent to the true model network, this information will be used to increase network inference prediction [2,5].By combining the resulted motif-filtered inference networks using an ensemble approach, predictive power can be increased [4]. By comparing the ensemble network with the original network, the overlap and mismatch can be determined which indicates the performance of prediction.

## Context within the course

In this course several aspects were discussed, such as network inference and network motifs. Network inference methods are widely used to improve understanding of complex regulatory networks. Often literature-based information is used to create the initial networks, such as pathways or protein-protein interaction networks. By combining network inference and literature-based information the inferred networks possibly give new insight to the understanding of regulatory networks. By incorporating network motifs, topological significance is included in the analysis. This results in a combined effort of information driven (all genes in model are known), network inference (inference based on *in sillico* data) and topological information (network motifs) to predict the true inference of the model.

[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