**A bioinformatics tool to identify network biomarkers in cancer**

**Henning Lund Hanssen, M.Sc. Candidate**

It is now becoming well established that the development and progression of cancer is dependent on the interplay of clusters of molecules in the network of the cell, rather than from the malfunction of one individual molecule. Furthermore it is the dynamic perturbations of a cancer cell’s network over time and the varying environmental conditions this network is exposed to, which directs the progression of the disease. The recent developments of increased availability of big-data in cancer research offered by the advent of new technologies such as Next Generation Sequencing (NGS) and Mass Spectrometry (MS) proteomics, allows us to now investigate the global network-perspective of the cell and the role of perturbations in that network in cancer progression. Thus, the identification of **network biomarkers** become critical to help cancer researchers and oncologists to monitor progression, predict development of cancer, and to design novel therapies to combat the disease.

Two advances in the field of bioinformatics allow us now to develop tools to identify network biomarkers in cancer. Firstly, bioinformatics methods have emerged using NGS data whereby single “driver genes” are identified as potential biomarkers using various algorithmic approaches, some of which are network-based. This advancement is known as **gene prioritization**, and is typically gene centric**.** Secondly, there have been great leaps forward in recent years to understand the topological organization of large-scale molecular networks and cluster these networks into groups, which are subnetworks corresponding to function. This advancement is known as **community/module detection in networks**. The goal of this project will be to integrate the progress offered by these two advancements and develop a bioinformatics tool which caters for the detection of network biomarkers from large gene lists, sourced from big data analysis, in cancer research. The tool will be optimally developed in the Java based Cytoscape network analysis environment, which has become the gold standard in the scientific community.

The supervisors will guide the optimal direction in each of the following main activities to achieve the above goal, which are summarized as follows:

1. Literature survey on single gene prioritization or driver gene methods, particularly the network based ones
2. Literature survey on network biology in medicine and the application of network clustering or network community algorithms.
3. Gain a brief overview of the current state-of-the-art with respect to biomarker identification in cancer. Familiarize as to what a biomarker is, and the current drawbacks.
4. Familiarize with the Cytoscape environment, and Cytoscape plugin development needs
   * The current tool space available in Cytoscape for activities 1 and 2 above. The advantages and disadvantages, and the benefit of integrating both in a unified framework.
5. Survey and select the optimal molecular network databases to be employed by the tool: STRING, GeneMania, iRef Index, *etc.*
6. Code and develop the Cytoscape plug-in:
   * Transforms a input large gene list into a network format based on the choice of activity 5 above
   * Clusters the network based on the optimal method(s) based on activity 2 above
   * Rank prioritize the network clusters (or potential network biomarkers) based on the optimal method(s) from activity 1 above
7. With the help and collaboration with the prostate cancer Movember project, to assess the ability of the tool to identify potential network biomarkers in prostate cancer.