Ranklust - A bioinformatics solution to identify network biomarkers in cancer

8th May 2015

Contents

Part I

Introduction

To this day we still struggle with cancer. Even with all our modern equipment and knowledge we have still not been able to tame this horrible disease. My thesis is about making a tool which provides cancer researchers with an easier way of identifying network biomarkers in cancer.

Chapter 1

Single gene prioritization

Chapter 2

Network generation

Chapter 3

Clustering

Chapter 4

Cytoscape

Cytoscape is the open-source software platform the Ranklust app will be developed on. Its main purpose is to visualize molecular interaction networks and biological pathways. It has an easy way to integrate ***Apps***, which may be combined by other apps again to build big and complex applications which may solve problems in a bigger picture. The goal of Ranklust is to cluster the networks we get from single gene prioritization and rank them in order to identify network biomarkers. Apps taking care of making the networks already exists, but there still has to be made a decision about whether or not they could be modified in order to better support the clustering.

Which databases to use has to be considered. The reason to use databases is because they have information on how protein and genes form a network based on how they interact with eachother. The initial database candidates in Ranklust are iRefIndex [1], GeneMania [2] and STRING [3]. These databases all have in common that it exists Cytoscape apps made to use these databases. STRING however, does not have any repository available through the Cytoscape app store, so interacting with the database through a new app in Cytoscape without making new plugins may be difficult. On the other hand, both iRefIndex and GeneMania have their repositories easily available to the public together with decent documentation. However, the difference between them is what they contain information about. iRefIndex contains data about protein-protein interaction (PPI), while GeneMania contains data about genes. Since proteins come from genes, GeneMania can also give us some information about proteins. Differences between the two databases will be discussed in greater depth at a later stage.

The open-source plugins in Cytoscape to communicate with the databases are iRefScape [4] for iRefIndex and GeneMANIA [5] for GeneMania.

Cytoscape is not j

Part II

Background

Introduction

We have our body, and inside our body we have our organs. These organs are made up of tissue, and tissue is made up of cells. Our cells performs two type of functions, to execute chemical reactions needed to stay alive, and to pass information for maintaining life onto the next generation.

Background

DNA, RNA and protein in general

The name of the process when DNA goes from DNA to RNA to protein is called the **Central Dogma**. I will focus on the cell and how we perceive it and the interaction inside it. The cell can be seen as a network community of interacting protein molecules. The protein comes from our DNA.

In our DNA, there are areas that contain codes for making protein. These areas are called genes. Also, all of the possible interactions between the cells are specified by the proteins in complex social networks. The reason for calling these networks for "social networks" is because it is the interactions that we look at as the connections between the cells. It is our protein that performs chemical reactions in our body. DNA on the other hand stores and passes information about how our body is built up. RNA is the intermediate stage between DNA and protein.

Protein

Amino acids

The protein in our body is made up from amino acids. The amount of amino acids in a protein may vary from 20 to 5000. But on average there is about 350 amino acids in a protein in our body. The way we identify amino acids is through using the alphabet. We use almost every letter in the alphabet, and it is organized in a chronological order when we show them in a list, but we miss some characters. They can also be identified by three letters, or the whole name of the amino acid.

Amino acids have some very basic attributes like volume and mass. But beeing acids we also have information about their polarity and their basicity/acidity. So the amino acids are either polar or non-polar, combined with beeing neutral, acidic or basic. And of course we are able to see to what degree they are acidic/basic.

Amino acids consist of an amino group, carboxyl group and a R group. Very often there is also a central carbon that combines all of these groups together. We have about 20 different amino acids and they can be classified into 4 types. The positively charged, the negatively, the polar and the non-polar. The positive amino acids are basic and the negative ones

are acidic. The four types just mentioned are not totally exclusively to eachother, though an amino acid cannot be both basic and acidic at the same time. At the same time an amino acid cannot be both polar and non-polar. But any combination of acidity/basicity together with polarity can occur.

Genome

Mutation

When the genome changes suddenly and unexpected, we have a mutation. A mutation may happen when several things occur:

- Insertion
  - A part of a chromosome gets inserted into another one
- Deletion
  - A part of a chromosome gets deleted
- Duplication
  - A part of a chromosome gets duplicated
- Inversion
  - A part of a chromosome gets inverted, but yet it stays in place
- Translocation
  - Two chromosomes exchanges parts and they become what is called a derivative

Cytoscape![6]

Bibliography

[1] URL: http://irefindex.org/wiki/index.php?title=iRefIndex (visited on 07/05/2015).

[2] URL: http://www.genemania.org/ (visited on 07/05/2015).

[3] URL: http://string-db.org/ (visited on 07/05/2015).

[4] URL: http://apps.cytoscape.org/apps/irefscape (visited on 07/05/2015).

[5] URL: http://apps.cytoscape.org/apps/genemania (visited on 07/05/2015).

[6] me. 'ifi'. In: *uio* (2015).