****

**Noakhali Science and Technology University**

# **CSTE 4111: Thesis**

**Under The Supervision of**

## **Koushik Chandra Howladar**

**Assistant Professor**

**Department of Computer Science and Telecommunication Engineering,**

**Noakhali Science and Technology University**

**Thesis Proposal**

**Bioinformatics methodologies for obesity and its associated cancer.**

**Shuvo Shaha Roy**

**Roll No: ASH1601036M**

**Session: 2015-16**

**Date of Submission: September 3, 2019**

**Introduction:**

Overweight and obesity are increasing problems that lead to significant health and social difficulties for people. Commonly defined by a measurement of Body Mass Index (BMI - calculated by dividing body weight (kilograms) by height (meters) squared), the prevalence of overweight (adult BMI of between 25 and 29.9) and obesity (BMI of 30 or over) is increasing. Overweight and obesity are global problems. A statistic shows that almost 30% of people in the world are obese or overweight. Obesity is directly linked to a number of different illnesses including type 2 diabetes, hypertension, gallstones, and gastro-oesophageal reflux disease, as well as psychological and psychiatric morbidities. The Health and Social Care Information Centre reported that there were 10,660 inpatient admissions to hospitals in England with a primary diagnosis of obesity in 2017/2018, which is 3 times as many as 7 years earlier in 2011/2012. Around 3 in every 4 patients are female. Cancer is the second leading of death globally and is responsible for an estimated 9.6 billion deaths in 2018. Globally, about 1 in 6 deaths is due to cancer. Approximately 70% of deaths from cancer occur in low and middle-income countries. Around one-third of deaths from cancer are due to 5 leading behavioral and dietary risks: high body index, low fruit and vegetable intake, lack of physical activity, tobacco use and alcohol use.In our research, we will perform a detailed survey of gene expression data available in public repositories on obesity and less commonly considered comorbidity such as cancer. Then we will try to develop an innovative pipeline that will integrate gene expression, cell-type data and online resources (e.g. a list of comorbidities from the literature), using bioinformatics methods such as gene set enrichment analysis and semantic similarity. We will try to find a list of common differentially expressed genes, gene ontology terms, and pathways among Obesity and comorbidities and the closeness among the selected pathologies by means of disease ontology terms. Physicians and other researchers, such as molecular biologists, systems biologists and pharmacologists can use it to analyze pathology in detail, from differential expressed genes to ontologies, performing a comparison with the pathology comorbidities or with other diseases.

**Literature Review:**

L Khaodhiar, KC McCowen, and GL Blackburn studied Obesity conditions in the human body and showed its comorbid diseases are led by Obesity[1]. Dixon and John B also carried out research on the health conditions of Obesity and the risk to other comorbid diseases[2]. Vucenik, Ivana, and Joseph P. Stains showed how obesity is linked to a different variety of cancer[3]. Del Prete, Eugenio, Angelo Facchiano, and Pietro Liò has developed an innovative pipeline using bioinformatics methods such as gene set enrichment analysis and semantic similarity that integrates gene expression, cell-type data and online resources (e.g. a list of comorbidities from the literature) for Coeliac Diseases and its comorbidities[4].

**Objectives:**

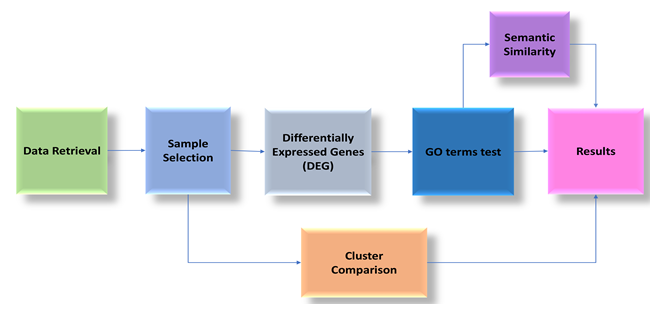
The objectives of the thesis study are-

* Investigating the relationships of obesity and its comorbidities
* Building an innovative pipeline for analyzing the gene expression and cell type data related to obesity
* By means of disease ontology terms, to find common differential expressed genes, gene ontology terms, and pathways among obesity and its comorbidities and calculating the closeness among the selected pathologies
* To develop a system structure so that the physicians and other medical researchers can use it to analyze pathology in detail from differential expressed genes to ontologies, performing a comparison with the pathology comorbidities or with other diseases

**Research Methodology:**  
The research will be carried out in four steps

* First of all, we will retrieve experimental data sets from the Gene Expression Omnibus (GEO) database available from the National Center for Biotechnology Information (NCBI), and the Array Express database available from the European Bioinformatics Institute
* The gene ontology (GO) project [39] provides comprehensive structured information of biological systems, from the molecular level to larger pathways, cellular- and organism-level systems. There are three GO domains: cellular component, molecular function and biological process (BP). Our approach is based on BP
* Then we will perform gene set enrichment analysis and semantic similarity among the different datasets
* Finally, we will approach to design the pipeline in R

**Workflow diagram for the proposed pipeline:**



**Figure:** A block diagram of the proposed pipeline

**Expected Output of the Research:**

The expected outcomes of the study will be:

* Statistics and GO terms tree for the selected Obesity studies
* Pathways that are common to different comorbid conditions
* Semantic similarity and KEGG enrichment for differential expressed genes

**Conclusion and Future Work**Bioinformatics statistical analysis of Obesity and its comorbidities is a thriving direction of medical research. It has two important benefits: the improved understanding of the Obesity that shapes the network of comorbidity diseases and the introduction of a new workflow of analyses and statistical estimators for complex diseases. We will develop a general methodology for analyzing Obesity and comorbidities. So this work can be extended and use in (1) other Obesity data sets, (2) other Obesity comorbidities and (3) other complex pathologies. The only constraint is the typology of data, downloaded by the GEO repository (or by Array Express repository), which consists of gene expression sets from sick and healthy patients, even GFD treated in Obesity case.

**References:**

[1] Khaodhiar, Lalita, Karen C. McCowen, and George L. Blackburn. "Obesity and its comorbid conditions." *Clinical cornerstone* 2.3 (1999): 17-31.

[2] Dixon, John B. "The effect of obesity on health outcomes." *Molecular and cellular endocrinology* 316.2 (2010): 104-108.

[3] Vucenik, Ivana, and Joseph P. Stains. "Obesity and cancer risk: evidence, mechanisms, and recommendations." *Annals of the New York Academy of Sciences* 1271.1 (2012): 37.

[4] Del Prete, Eugenio, Angelo Facchiano, and Pietro Liò. "Bioinformatics methodologies for coeliac disease and its comorbidities." *Briefings in bioinformatics* (2018).