Working title:
Exploring the plausible lean phenotype in a northern Swedish cohort.

Student: JERNEJA MISLEJ

University of Lund

bif15jmi@student.lu.se

Project supervisor: ALAITZ POVEDA

University of Lund

Alaitz.Poveda@med.lu.se

Assistant project supervisors: PAUL W. FRANKS, AZRA KURBASIC

University of Lund

Paul.Franks@med.lu.se

Azra.Kurbasic@med.lu.se

Abstract

The master thesis project objective is to design, implement and preform the extraction and analysis of the lean phenotype, potentially present in a large, northern Swedish cohort. An extensive collection of lifestyle and clinical longitudinal data is available within a population-based prospective cohort study, enabling exploration of lifestyle risk factors with respect to various cardio-metabolic diseases and obesity. Several lifestyle risk factors and their combination have been well established to lead to obesity. Subjects, who are exceptions and maintain to be lean, despite having established lifestyle risk factors for obesity, present potential candidates with a lean phenotype.

Successful extraction of subjects with a lean phenotype will enable further research of the lean phenotype phenomena. By switching the focus from subjects who are diagnosed, predisposed or at risk for obesity, to subjects with a lean phenotype, protective characteristic against obesity can be explored and established. Amongst other, clinical data available within the population-based prospective cohort study, includes genome-wide set of genetic variants. Purposed extraction of subjects with a lean phenotype could therefore lead to identification of genes associated with the lean phenotype or leastwise confirm and further establish genes, which have already been associated with the lean phenotype.

I Basic information

Within the masters program of Bioinformatics at Lund University, Department of Biology, I will carry out a 30 credit master thesis project under the code BINP30.

The project will take place at the department of Genetic and Molecular Epidemiology in the Lund University Diabetes Center at the Clinical Research Center in Malmo, under the supervision of Alaitz Poveda and co-supervision of Paul Franks and Azra Kurbasic.

The project will start in week 4 on 23th of January 2017 and will finish in week 26, on the 26th of June 2017.

II Project

I Introduction

The project will be concerned with the extensive collection of lifestyle and clinical longitudinal data, available within a population-based prospective cohort study[1]. The data consists of approximately 600 variables, with partially obtained values for approximately 25000 subjects from the northern Swedish population of the province Västerbotten. Values were obtained sequentially, at baseline, between the years 1990 and 1999, and at a follow-up, between the years 1995 and 2008, resulting in an approximate time difference of 10 years. Lifestyle variables cover environmental exposures such as diet, physical activity, smoking, alcohol and education. Clinical variables cover a large range of medical and anthropometric characteristics, including plasma glucose, serum total cholesterol, triglycerides concentrations and for a smaller subset of 5000 subjects, the genomic DNA samples. DNA samples were used to genotype approximately 150 gene variants, which were mostly selected based on the published GWAS meta-analyses of cardiometabolic traits.

II Project aims

Project goal is to extract and analyze subjects with a lean phenotype.

The first step will include familiarization with the data and topic, followed by data exploration and preparation. In the second step, script(s) for extraction of certain specific subjects will be designed, implemented and tested. Subjects in the extracted subset will have to be persistently exposed to lifestyle factors established to lead to obesity. At the same time, the subjects will have to possess characteristics of a lean and healthy person,

with no confounding medical condition or environmental factor that could explain the persistent leanness despite the persistent unhealthy lifestyle. When having extracted a lean phenotype subset, the following steps would be to analyze characteristics that were not included in the extraction, in order to further research the lean phenotype[3][4][5]. Particularly focusing on assessment of lean phenotype heritability and on potential extraction and analysis of associated gene variants.

III   Methods

Several lifestyle factors, such as diet, physical activity, smoking and education have been established to lead to obesity, with a high long-range predictive accuracy[2]. These predictions were based on lifestyle risk scores at the baseline. Methods, serving for the lean phenotype subset extraction will attempt to combine these same lifestyle risk factors at baseline and at follow up to ensure persistence. Similar procedures are to be used to further extract those subjects who are persistently lean and healthy, based on the complement of the obese subset, used to calculate the predictive accuracy of lifestyle risk scores associated with obesity[2]. Additional methods are to be adapted and developed to ensure the final subset, presenting the lean phenotype, will not include subjects who are lean due to confounding medical conditions or environmental factors.

The remaining characteristics, which were not used in the lean phenotype subset extraction procedure, will be further explored by appropriate statistical analyses, primarily focusing on heritability assessment. The subset of subjects, who were genotyped for a large set of gene variants associated with cardiometabolic traits, might or might not also include a sufficient sample size of subjects with the lean phenotype. If the resulting sample size will be sufficient, genome-wide association study can be carried out with an attempt to associate certain gene variants to the lean phenotype.

IV   Timeline

- 23.1.2017 - 10.2.2017: Project starts, the student gets acquainted with the topic and available data and prepares the data for processing.

- 13.2.2017 - 24.3.2017: Design, implementation and testing of script(s) for lean phenotype subset extraction.

- 27.3.2017 - 10.4.2017: Analysis of the lean phenotype subset, focusing on heritability assessment.

- 13.4.2017 - 26.5.2017: Analysis of the lean phenotype subset, focusing on extraction and analysis of associated gene variants.

- 29.5.2017 - 9.6.2017: Completion and closure.

- 12.5.2017 - 26.6.2017: Master thesis report and presentation preparation.

References

[1] A. Kurbasic *et al.*, Gene-Lifestyle Interactions in Complex Diseases: Design and Description of the GLACIER and VIKING Studies. *Curr Nutr Rep*, (2014)

[2] A. Poveda *et al.*, Innate biology versus lifestyle behaviour in the aetiology of obesity and type 2 diabetes: the GLACIER Study. *Diabetologia*, (2016)

[3] I. Harosh , Rare Genetic Diseases with Human Lean and/or Starvation Phenotype Open New Avenues for Obesity and Type II Diabetes Treatment. *Current Pharmaceutical Biotechnology*, (2013)

[4] I. Harosh *et al.*, Enteropeptidase: A Gene Associated with a Starvation Human Phenotype and a Novel Target for Obesity Treatment. *PLoS ONE* 7(11): e49612. doi:10.1371/journal.pone.0049612 (2012)

[5] Jolivet G, Braud S, DaSilva B, Passet B, Harscoet E. *et al.*, Induction of Body Weight Loss through RNAi-Knockdown of APOBEC1 Gene Expression in Transgenic Rabbits. *PLoS ONE* 9(9): e106655. doi:10.1371/journal.pone.0106655 (2014)