1 INTRODUCTION

One of the major aims in nutrition research is to elucidate the relationship between diet, lifestyle, genetics and disease, and large amounts of evidence have been obtained from nutritional intervention and epidemiologic studies in the past decades [1]. In these studies, biomarkers serve as an important and objective tool to measure the status or change of the biological system reflecting the characteristics of the system or its response to certain exposure and environment factors. They have been widely used to evaluate the nutritional status of the study subjects, dietary intake and its resulting biological consequences [2]. Although being extensively applied, many issues regarding the application of biomarkers in the nutrition field have emerged due to the lack of ontology and guidelines. This has also led to the difficulties in interpreting and understanding the measurements of biomarkers [3].

Meanwhile, biomarkers in the nutrition field are experiencing a considerable change due to the achievements in modern analytical technologies and the development of omics approaches, especially metabolomics. Metabolomics measures hundreds or thousands of metabolites at a time and provides an individual or group characterization of the subjects when they are exposed to certain environmental factors [4]. It allows the ‘global’ measurements of metabolites that, associated with the change of the biological system or process, therefore offers the potential for the discovery of new biomarkers. A large number of studies have been carried out using untargeted metabolomics as an approach to identify dietary biomarkers [5–7]. They face a common challenge that metabolomics generates huge amounts of measurements from different subjects at a time; however, the extraction of the relevant information from noisy data with large variability is rather difficult. Various multivariate analyses are adopted in such studies aimed at reducing the dimension of data and selecting important variables (potential biomarkers) related to the intervention/treatment. However, in many cases they ended up selecting a too small number of variables to maintain the high prediction or a too large number of variables to be inclusive. Variable selection is a crucial step as it decides the scale and cost of further validation and other downstream studies. Therefore, better strategies are still needed to produce results that are more precise.

Biomarkers related to dietary exposures play an important role in nutritional studies. Such biomarkers can be used to assess the accurate dietary intake, exploit the association between diet and disease, and monitor the compliance in dietary intervention studies. Onion has been used as a “food medicine” since ancient times for its anticarcinogenic, antibiotic and anti-inflammatory

effects [8], however, none of those effects have been proved in human studies and no cause and effect relationship has been established so far. To further document any potential beneficial effect of onion and to reveal the mechanism behind it, more biomarkers related to onion consumption are needed.

1.1 The aim of the thesis

The overall aim of the thesis is to enhance the use of biomarkers in nutrition science by adding necessary ontology, data analysis methodology and new markers. The strategy to reach this aim is to develop a classification scheme for the current dietary and health biomarkers applied in the nutrition field thereby improving their application, to optimize the data analysis strategy in untargeted metabolomics studies for more precise discovery of biomarkers, and to apply such strategy for the discovery of biomarkers, exemplified by search for biomarkers related to onion consumption. The specific aims of the projects within the thesis are therefore as follows:

- Develop a classification scheme for dietary and health biomarkers to provide a well-defined ontology to better understand and communicate the research on dietary biomarkers in the nutrition field.
- Compare bi- and tri-linear PLS models for variable selection in untargeted metabolomic studies with time series design and provide a better strategy for identifying important variables (biomarkers) discriminating different intervention groups.
- Identify biomarkers of food intake for onion in an LC-MS based untargeted metabolomics study using the optimal approach selected in the previous project and other popular multivariate analyses.