



# Master Thesis

Tu Hu

## Investigation of Whole Grain Wheat Intake Biomarkers

Supervisor: Gözde Gürdenz

Jun, 2019

# Contents

<b>1</b>	<b>Preface</b>	<b>4</b>
<b>2</b>	<b>Introduction</b>	<b>4</b>
2.1	R, Tidy Format and Data Wrangling . . . . .	4
<b>3</b>	<b>Materials and methods</b>	<b>4</b>
3.1	Software . . . . .	4
3.2	Data-preprocessing . . . . .	4
3.3	Data conversion from MATLAB to R . . . . .	4

## **Acronyms**

. 3

# 1 Preface

## 2 Introduction

### 2.1 R, Tidy Format and Data Wrangling

R is an open-source data analysis language.

R is perfect for bioinformatics analysis.

Tidy format is a term used in R describing a philosophy of storing data. Tidy format has the characteristics:

- each variable is in its own column
- each observation, or case, is in its own row

If data is stored in tidy form. Data scientists spend less time fighting with the tools and more time working on your analysis.

## 3 Materials and methods

### 3.1 Software

MATLAB

MZmine 2.31, an open source data processing software for LC-MS and GC-MS.

MassLynx

DataBridge, an LC-MS data file conversion program built-in MassLynx developed by Waters.

### 3.2 Data-preprocessing

Data-preprocessing consists x steps.

First, data format was converted by DataBridge from '.raw' to '.cdf'. '.raw' was the format directly generated by Waters analytical platform. In order to be readable by MZmine, data was converted<sup>1</sup>.

Then, the data was preprocessed by MZmine (2.31) following the steps: peak detection, deisotoping, alignment and gap filling.

Positive mode and negative were separately processed because of different noise level and in-source reaction. Blank samples were also excluded in pre-processing.

In the end, the detected features, including information of mass to charge ratio (m/z), retention time (rt) and intensities were output as '.csv' files for further investigation.

### 3.3 Data conversion from MATLAB to R

'm2r.R'<sup>2</sup> can convert a matlab dataset into R tidy form.

---

<sup>1</sup>N.B. Although in MZmine manual, '.raw' file is described as a compatible format, in practice some weird errors were generated when '.raw' format was input into MZmine.

<sup>2</sup>codes were reposited on GitHub: <https://github.com/tuhulab/bfi-wholegrain/blob/master/r/m2r.R>