# Statistical Analysis Plan (SAP)

Calorie Restriction Intervention Induces Enterotype-associated BMI Loss in Nonobese Individuals


| | |
|---|---|
| **Principal Investigator** | Kaiye Cai, Master |
| | BGI-Shenzhen |
| | Shenzhen, Guangdong, China |
| **Protocol identification number** | BGI-20170204 (dated 2017-3-20) |
| **Author** | Hua Zou, Bachelor |
| | BGI Education center |
| | University of Chinese Academy of Sciences |
| | Shenzhen, Guangdong, China |
| **Version** | Draft 1.1 |

# SIGNATURE PAGE

| Kaiye Cai | Kaiye Cai | BGI - Shenzhen | 2017. 3. 20 |
|---|---|---|---|
| **Principal Investigator** | **Signature** | **Site name or ID number** | **Date** |

| Hua Zou | Hua Zou | BGI Education Center | 2017. 3. 20 |
|---|---|---|---|
| **Author** | **Signature** | **Site name or ID number** | **Date** |

## Abbreviations

| | |
|---|---|
| CR | Calorie restriction |
| BMI | Body mass index |
| ETB | Enterotype *Bacteroides* |
| ETP | Enterotype *Prevotella* |
| BH | Benjamini-Hochberg |
| LC-MS/MS | Liquid chromatography–mass spectrometry |
| Lasso | Least absolute shrinkage and selection operator |
| LOOCV | leave-one-out cross validation |
| GLM | generalized linear model |
| PERMANOVA | Permutational multivariate analysis of variance |

## Table of Contents

## 1. Introduction

The aim of this project is to simultaneously evaluate impacts of CR on BMI, gut microbiota and blood amino acids in nonobese individuals.

This statistical analysis plan (SAP) will give more detailed description of the endpoints in the study and the corresponding analyses

## 2. Study design

Volunteer-wanted posters were propagated at the China National Gene Bank in Shenzhen from March to April 2017. A non-obese healthy volunteer was considered if his/her BMI less than 28 $kg/m^2$.[1] In addition, recruited volunteers should meet all the following criteria: 1) without antibiotics in the recent 2 months; 2) without prebiotic or probiotic supplements in the recent 2 months; 3) without hypertension, diabetes mellitus, gastrointestinal disease and other severe auto-immune disease; 4) regular eating and lifestyle patterns; 5) no international travel in the recent 3 month; 6) without prebiotic or probiotic supplements in the recent 2 months. 50 individuals met all the criteria and were recruited in this study, and 41 individuals (24 females and 17 males aged 30 ± 6 years old) completed the whole intervention (**Table 1**).

**Table 1. Cohort Description**

| | **Cohort (Mean ± SD)** |
|---|---|
| **Number of subjects** | 41 |
| **Sex (female/male)** | 24/17 |
| **Age** | 30 ± 6 |
| **BMI ($kg/m^2$)** | 23.72 ± 2.81 |
| **Weight(kg)** | 64.84 ± 11.57 |

The study included a one-week run-in period (baseline) and a three-week CR dietary intervention period.

During the first week (run-in period), all healthy volunteers consumed their usual diet and were encouraged to avoid yogurt, high-fat foods and alcohol. The CR diet was comprised of ~50% calories of a normal-calorie diet (female, 1000kcal/day; male, 1200kcal/day). It was designed with carbohydrate, fat and protein as approximately 55%, 30% and 15% of the total energy intake respectively, according to the Dietary Guidelines for Chinese Residents (2016) and nutritionally balanced [2] and a recent large nutritional study in China [3]. Common foods in low-calorie diets such as rice, vegetables, eggs, pork and beef were prepared in this study center to control experimental variables introduced by different foods and calorie estimation errors. Traditional Chinese cooking style - boiled, stir-fried and stewed, were applied for foods. For each meal, digital scales were used to measure the nutritional and caloric values of different foods and total meal for male and female respectively. BMI, blood and fecal samples of each volunteer were collected at this study center at baseline and after the 3-week CR intervention (**Figure 1**).

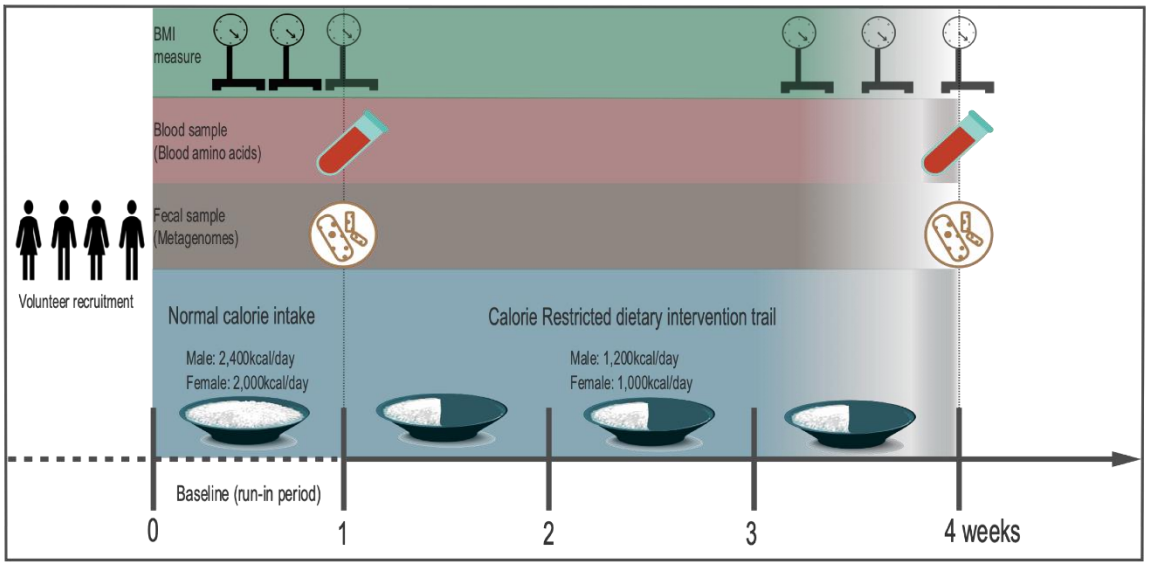


*Figure 1 Overview of the experimental design. Illustration of experimental design, including a 1-week run-in period (baseline) and a 3-week calorie restriction (CR) dietary intervention trial with 50% energy deficit diet (male, ~1200Kcal/day; female, ~1000Kcal/day). BMI, fasting blood samples and fecal samples of 41 enrolled healthy subjects were collected before and after the CR intervention to assess its effects on BMI, blood amino acids and gut microbiome in two enterotype groups.*

### 2.1 Sample size calculation

The primary outcome measure for the power calculation is the change of BMI between baseline and after the intervention. The sample size estimation is within the framework of Wilcoxon signed-rank test. Power analysis was conducted in G*Power to determine a sufficient sample size using an alpha of 0.05, a power of 0.80, a medium effect size (dz = 0.5), and two tails [4]. Based on the aforementioned assumptions, the desired sample size is 35. We have conservatively assumed that the overall dropout rate will be 30% during the 4-week study period. To account for loss to follow-up, the power will be 80% if 50 study subjects are recruited.

## 3. Aims and objectives

The purpose of this intervention study is to evaluate the effect of calorie restriction on BMI loss, amino acid, and gut microbiota in healthy volunteers of two different enterotypes and provide useful insights for potential application of gut microbiome stratification in personalized nutrition intervention.

## 4. Outcomes

This section will present the outcomes investigated to answer the study aims and objectives. The analyses are described in section 3 Analyses.

### 4.1 Primary outcome

**BMI**. It will be measured at baseline, CR intervention period.

### 4.2 Secondary outcomes

**Blood amino acids**

Blood samples of each volunteer will be collected at the investigators' study center at baseline and after the 3-week CR intervention. The concentrations of 31 blood amino acids and derivatives in the fasting serum samples will be measured by LC-MS/MS.

**Amino acids**

- 1-Methyl_histidine
- 2-aminoadipic acid
- 3-Methyl_histidine

- Alanine
- Arginine
- Argininosuccinic
- Asparagine
- Aspartate
- Citrulline
- Cysteine
- Ethanolamine
- Glutamate
- Glutamine
- Glycine
- Histidine
- Hydroxyproline
- Isoleucine
- Leucine
- Lysine
- Methionine
- Ornithine
- Phenylalanine
- Proline
- Serine
- Taurine
- Threonine
- Tryptophan
- Tyrosine
- Valine
- α-aminoisobutyric acid
- β-Alanine

**Gut microbiota**

Fecal samples of each volunteer will be collected at the investigators' study center at

baseline and after the 3-week CR intervention. The gut microbial composition will be determined using shotgun metagenomics sequencing of fecal DNA.

## 5. Populations to be analyzed

41 participates will be analyzed in this study.

## 6. Analyses

All outcomes will be presented using descriptive statistics: Wilcoxon rank-sum test is used to detect the significant differences on phenotypes, the concentrations of blood amino acids and the relative abundances of genera and species between enterotypes. Wilcoxon signed-rank test is used to detect the significant differences on phenotypes, the concentrations of blood amino acids and the relative abundances of genera and species in paired samples before and after the intervention. Permutational multivariate analysis of variance (PERMANOVA) is used to detect the association between enterotypes and the overall blood amino acid profile at baseline by using PERMANOVA with 9,999 permutations on enterotypes. Principal coordinate analysis is used to visual overall gut microbial based on the relative abundance profiles between baseline and after the intervention. Principal component analysis is used to visual overall amino acid composition based on the blood amino acid profiles between baseline and after the intervention. To investigate whether we could predict BMI loss ratio using omics features, we will perform a Lasso (Least absolute shrinkage and selection operator) regression analysis between baseline relative abundances of gut common species and the concentrations of blood amino acids (independent variables), and BMI loss ratio (dependent variables). We first normalize values of both independent and dependent variables (R, scale function). We then use the R function cv.glmnet to choose the most appropriate value for $\lambda$ in the Lasso model (R glmnet package, alpha =1, family="gaussian", nfolds=10, alpha=1, nlambda=100). Here, $\lambda$ is the tuning parameter ($\lambda > 0$) which controls the strength of the shrinkage of the variables. We then apply the Lasso feature selection process by shrinking the Lasso regression coefficients of non-informative variables to zero and selecting the variables of non-zero coefficients. To reduce overfitting with a limited sample size (n=41), we apply leave-one-out cross

validation (LOOCV) to estimate the prediction performance of BMI loss ratio using a generalized linear model (GLM) of the selected features (creatFolds function in R caret package and the glm function in R base package). Likewise, we also use baseline BMI values for LOOCV to estimate its prediction performance for CR-associated BMI loss ratio. Spearman's rho values will be calculated between actual BMI loss ratios and the predicted values. P-value adjustment is applied for multiple hypothesis testing on the concentrations of blood amino acids, the relative abundances of gut microbial genera and species use Benjamini-Hochberg (BH) method. R 3.5.0 will be used for all statistical analysis.

## 7. Missing data

In this study, data of 5 participates' blood amino acids was missing at baseline and after the intervention. For a specific analysis, the study participates with missing data on the blood amino acids will be excluded from the analysis.

**Reference**


1. Potts M: **China's one child policy**. *BMJ* 2006, **333**(7564):361-362.
2. Wang SS, Lay S, Yu HN, Shen SR: **Dietary Guidelines for Chinese Residents (2016): comments and comparisons**. *J Zhejiang Univ Sci B* 2016, **17**(9):649-656.
3. Wan Y, Wang F, Yuan J, Li D: **Optimal dietary macronutrient distribution in China (ODMDC): a randomised controlled-feeding trial protocol**. *Asia Pac J Clin Nutr* 2017, **26**(5):972-980.
4. Faul F, Erdfelder E, Buchner A, Lang A: **G* Power Version 3.1. 7 [computer software]**. *Uiversität Kiel, Germany* 2013.