Statistical plan for Fermentable carbohydrate and gut hormone release

Statistical analysis will be performed using Statistical Program for Social Sciences (SPSS) version 25 (International Business Machines, New York, USA). The average of the two baseline (fasted) measurements will be used as the reference baseline measurement for all parameters. Comparison between diets will be done using linear mixed models. Where possible, area under the curve will be calculated to compare the diets. Total area under the curve (tAUC) and incremental area under curve (iAUC) will be calculated using GraphPad Prism version 7 (GraphPad Software Inc, California, USA). If the baseline measurements were different between the dietary intervention groups, iAUC will be used used for comparison instead of tAUC.

The AUC and baseline measurements will be compared between the three dietary intervention groups using linear mixed models with diet as the fixed factor and participant as the random factor. If AUC calculations are not possible due to missing values, difference between diets will be compared using all time point data and linear mixed models with diet and time as the fixed factor and participant as the random factor. The model will us restricted maximum likelihood estimation. Fisher's least significant difference (LSD) test was used as the post hoc test for both tests. Significance is considered $P < 0.05$ for all statistical tests. P values lower than 0.0005 will be represented as $P<0.0005$.

Metabolomic statistical plan

One dimensional 1H NMR spectra were automatically corrected for phase and baseline distortions and referenced to the TSP singlet at δ 0.0. Spectra were digitized using an in-house MatLab script (2014a, MathWorks, Natick, U.S.A). Spectra were referenced according to the internal chemical shift reference for TSP at δ 0.0. Spectral regions corresponding to the internal standard (δ -0.5 to 0.5), water (δ 4.7 to 4.9) and noise (δ 9.5 to 11) were excluded. All spectra were normalised using the median spectrum and median fold change normalisation. These were imported into MatLab (2014a, MathWorks, Natick, U.S.A) to conduct multivariate statistical analysis. Data were centred and scaled to account for the repeated measures design and then modelled using partial-least-squares–discriminant analysis (PLS-DA) with Monte Carlo cross-validation (MCCV). The models were built comparing between diets (LF, I-HF and D-HF) and within a diet. The fit and predictability of the models obtained were determined and expressed as R2Y and Q2Y values, respectively. Using the variance and mean regression coefficient, a t-score for each variable and p-value were calculated. P-values were then adjusted for multiple testing by calculating the q-value using the Storey Tibshirani method (Storey and Tibshirani, 2003). The goodness of fit (R2Y) of the MCCV models was calculated across all models using the training data and the goodness of prediction (Q2Y) for the test data.

Metabolite Identification

A combination of data-driven strategies will be used as previously described (Garcia-Perez et al., 2020). These included STOCSY and STORM. In addition, selective 1D TOCSY sequence and 2D NMR

experiments such as J-Resolved spectroscopy (jresgpprqf), gradient 1H–1H COrrelation SpectroscopY 45° (cosygpprqf), 1H–1H Total Correlation Spectroscopy (mlevgpphprzf), 1H–13C Heteronuclear Single Quantum Coherence (hsqcedetgpsisp2.3) and 1H–13C Heteronuclear Multiple Bond Correlation (hmbcedetgpl3nd) are acquired to identify metabolites.

Contour plots showing the change of metabolites were generated using Matlab (2019a, MathWorks, Natick, U.S.A). The pairwise comparisons (within diet) will be carried out using Wilcoxon signed rank test. P-values were adjusted using Storey-Tibshirani method, that results in Q-values. Cut off values for the Q values were <0.1 or <0.05 for significance.