Article

MetaboVariation: Exploring Individual Variation in Metabolite Levels

Shubbham Gupta [1], Isobel Claire Gormley [2] and Lorraine Brennan [1,*]

1   School of Agriculture and Food Science, University College Dublin, Belfield, D04 V1W8 Dublin, Ireland
2   School of Mathematics and Statistics, University College Dublin, Belfield, D04 V1W8 Dublin, Ireland
*   Correspondence: lorraine.brennan@ucd.ie

Abstract: To date, most metabolomics biomarker research has focused on identifying disease biomarkers. However, there is a need for biomarkers of early metabolic dysfunction to identify individuals who would benefit from lifestyle interventions. Concomitantly, there is a need to develop strategies to analyse metabolomics data at an individual level. We propose "MetaboVariation", a method that models repeated measurements on individuals to explore fluctuations in metabolite levels at an individual level. MetaboVariation employs a Bayesian generalised linear model to flag individuals with intra-individual variations in their metabolite levels across multiple measurements. MetaboVariation models repeated metabolite levels as a function of explanatory variables while accounting for intra-individual variation. The posterior predictive distribution of metabolite levels at the individual level is available, and is used to flag individuals with observed metabolite levels outside the 95% highest posterior density prediction interval at a given time point. MetaboVariation was applied to a dataset containing metabolite levels for 20 metabolites, measured once every four months, in 164 individuals. A total of 28% of individuals with intra-individual variations in three or more metabolites were flagged. An R package for MetaboVariation was developed with an embedded R Shiny web application. To summarize, MetaboVariation has made considerable progress in developing strategies for analysing metabolomics data at the individual level, thus paving the way toward personalised healthcare.

Keywords: metabolite levels; intra-individual variation; Bayesian generalised linear model; personalised healthcare

1. Introduction

Nutrition research has made significant progress in recent years. Though it is widely accepted that early lifestyle interventions could reduce disease burden, it is difficult to identify individuals that would benefit most from targeted nutrition interventions. While metabolomic approaches have been highly successful in identifying disease biomarkers [1–3], there is a lack of biomarkers related to early metabolic disturbances. Successful areas to date include the study of acylcarnitines in various diseases—there is strong evidence showing alterations in acylcarnitine levels in metabolic disorders, cardiovascular diseases, diabetes, depression and neurologic disorders [4]. In the last 10 years, accumulating evidence has supported the role of branched-chain aminos acids as biomarkers for the development of type 2 diabetes. Recently, a systematic review and meta-analysis reported that plasma branched-chain amino acids were associated with type 2 diabetes incidence, irrespective of the time of follow-up [5]. Though branched-chain amino acids present an interesting proof of concept for biomarkers of early metabolic dysfunction, it is essential to explore alternative approaches. Concomitant with this is the need to embrace personalised approaches and analyse metabolomics data at the individual level [6]. This is driven partly by the emergence of personalised nutrition approaches, where dietary advice is tailored to the individual [7].

Recent research [8] has demonstrated that individual variation in blood metabolites may help discover potential indicators of human ageing or pertinent diseases. Metabolites

represent individual variations in human biological samples. Using metabolomics data from whole blood, a total of 48 metabolites showed moderate to high coefficients of variations (CVs); the authors [8–10] proposed these metabolites as potential biomarkers and put forward the concept that metabolites with high CVs could be considered personal markers. However, many factors such as age, BMI, nutrition, ethnicity, gender or lifestyle impact the metabolome, and there is a need to develop methods to examine metabolites at an individual level. Furthermore, considering these and other factors, such as diet, it is challenging to establish workable reference ranges for the metabolome. Hence, the concept of monitoring an individual and the variation within an individual has emerged. Using repeated measures from individuals can highlight deviations from one's typical profile and can indicate the onset of metabolic perturbations [11]. Applying this concept using omics technologies over a 14-month monitoring period [12] demonstrated the presence of dynamic alterations in profiles as a participant moved from a healthy to diseased state. In a follow-up study, the longitudinal signature of omics data was related to insulin resistance, with the data indicating that each individual was unique, calling for the need to analyse data at the level of the person [13]. However, there is a paucity of statistical approaches that identify individuals with small metabolic perturbations and that model metabolomic data at the individual level.

The objective of the present study was to develop a method called MetaboVariation that uses a Bayesian generalised linear model (BGLM) to flag individuals with intra-individual variations in their metabolite levels. To accomplish this, we modeled repeated measures of individuals' metabolite levels to infer typical fluctuations of metabolites in individuals. Given our objective, the Bayesian paradigm is particularly advantageous due to its natural provision of posterior predictive distributions; after inferring typical fluctuations, posterior predictive distributions are used to flag individuals with intra-individual variability in their metabolite levels. The novelty of the MetaboVariation approach is twofold: firstly, it models metabolomics data at an individual level and secondly, it flags individuals with intra-individual variation in their metabolite levels.

The statistical details underpinning MetaboVariation are described in Section 2. Section 3 assesses the performance of MetaboVariation and discusses the results using simulated and real-world data. Section 4 discusses the use of MetaboVariation and potential enhancements, and our conclusions are in Section 5. All reported results were created using the R package MetaboVariation (https://github.com/shubbham28/MetaboVariation, accessed on 20 December 2022), which is freely accessible. An R Shiny web application is available within the R package to aid user accessibility.

2. Materials and Methods

In metabolomics, repeated measurements of individuals are frequent, but their analysis can be complex. Despite the availability of a range of specialised analytic tools, such as generalised linear models (GLMs) [14], there has not been widespread use of such tools to model repeated measures metabolomics data.

2.1. The MetaboVariation Method

In this study, we propose MetaboVariation, which employs a Bayesian generalised linear model (BGLM) to model repeated measures of metabolite levels of individuals. MetaboVariation is fitted using the MCMCglmm package [15], which uses Metropolis–Hastings updates [16]. In the following, we assume independence between the individuals in the cohort whose metabolite levels are being measured, and between the metabolites. The model considers metabolomics data that contain the levels of M metabolites for N individuals across T time points. We consider a single metabolite m, recorded in an $N \times T$ matrix Y^m. The covariates of individuals are stored in an $N \times L$ matrix X, where L denotes

the number of covariates. The level of metabolite *m* of a single individual *i* at time point *t* is modelled as

$$y_{it}^m = \beta_0 + \sum_{l=1}^{L} \beta_{1,l} x_{il} + s_{it} + \epsilon_{it} \tag{1}$$

where y_{it}^m is the level of metabolite *m* for individual *i* at time point *t* and x_{il} is covariate *l* for individual *i*. The mean intercept is denoted by β_0, whereas $\beta_{1,l}$ is the regression coefficient for covariate *l*. The random effect for the ith individual at time point *t* is denoted by s_{it}, where $s_{it} \sim N(0, \sigma^2)$, and ϵ_{it} is the random error for individual *i* at time point *t* where $\epsilon_{it} \sim N(0, \sigma_\epsilon^2)$. Here, *S* and ϵ respectively denote $N \times T$ matrices of random effects and random errors for *N* individuals at *T* time points. However, as metabolite levels among individuals will differ, the intercept should also differ for different individuals, i.e., $\beta_{0,i} \neq \beta_{0,j}$ for individuals *i* and *j*. To simplify the model, for every individual *i*, the difference $\beta_{0,i} - \beta_0$ is incorporated into the term s_{it}, thus not influencing the estimates of β_0 and β_1 where $\beta_1 = (\beta_{1,1}, \beta_{1,2}, ..., \beta_{1,L})$. As (1) considers only a single metabolite *m*, correlation among metabolites is not considered here.

The posterior distribution of the BGLM, given the covariates $\boldsymbol{X}$ and metabolite data $\boldsymbol{Y}^m$, is:

$$P(\beta_0, \beta_1, \sigma^2, \sigma_\epsilon^2 | \boldsymbol{Y}^m, \boldsymbol{X}) \propto P(\boldsymbol{Y}^m | \beta_0, \beta_1, \sigma^2, \sigma_\epsilon^2, \boldsymbol{X}) \times P(\beta_0, \beta_1, \sigma^2, \sigma_\epsilon^2) \tag{2}$$

where the first term on the right-hand side of (2) is the likelihood function of the data and the second term denotes the assumed prior distributions of the regression parameters β_0 and β_1 and variance parameters σ^2 and σ_ϵ^2. Given the model, the likelihood is Gaussian. Independent prior distributions are assumed for the regression and variance parameters: zero-centred Gaussian priors with large variance (1e + 10) are assumed for the regression parameters with inverse Wishart prior distributions assumed for the variance parameters, following the default settings in the MCMCglmm package [15].

Samples of $\beta_0, \beta_1, \sigma^2$ and σ_ϵ^2 are drawn from the posterior distribution using a Metropolis–Hastings algorithm and are used to predict individuals' metabolite levels. A posterior predictive distribution is the distribution of new data points conditional on the observed data [17]. To construct the posterior predictive distribution for an individual *i* at time point *t*, several linear models are constructed by drawing model parameters and latent variables from their joint posterior distribution. The number of models is equal to the number of chains multiplied by the difference between the number of iterations and burn-in iterations. Each model gives a predicted value, thus creating samples from the posterior predictive distribution of $\tilde{y}_{it}^m$, where $\tilde{y}_{it}^m$ denotes the predicted level of metabolite *m* for individual *i* at time point *t*.

After drawing samples from the posterior predictive distribution $p(\tilde{y}_{it}^m)$, a W% highest posterior density (HPD) prediction interval is formed, where typically *W* = 95. If the observed value y_{it}^m lies outside the HPD prediction interval, the individual *i* at time point *t* is flagged, suggesting that the individual *i* has intra-individual variability in their level of metabolite *m* at time point *t* compared with their metabolite levels at other time points.

In order for the posterior predictive distributions to be valid, the Metropolis–Hastings algorithm must have converged such that independent samples of $\beta_0, \beta_1, \sigma^2$ and σ_ϵ^2 have been drawn from the posterior distribution. Assessing the convergence of the Metropolis–Hastings algorithm and the independence of the samples drawn is therefore necessary. Though a range of diagnostic tools are available to assess convergence [17], for this study, we employed the potential scale reduction factor (PSRF). The PSRF relies on running multiple Metropolis–Hastings algorithms or chains for each model fitted, where the starting points for each chain are dispersed throughout the parameter space. Here, the starting values for the multiple chains follow the default settings in the MCMCglmm package [15]. The PSRF approach considers the estimated between- and within-chain variances for each model parameter, with approximate equality indicating possible convergence. For each model parameter, PSRF values in the range 0.9 to 1.05 are typically used to indicate convergence. Here, PSRF values are monitored for the $\beta_0, \beta_1, \sigma^2$ and σ_ϵ^2 parameters of the model.

To ensure independence of the sampled model parameters, informed by autocorrelation plots, the chains were thinned every second iteration.

To prevent overfitting, the total number of observations for each metabolite (i.e., $N \times T$ metabolite levels) was divided into twenty-five sub-datasets (i.e., 4% of observations in each sub-dataset). A BGLM was fitted to all observations with one sub-dataset held out and the posterior predictive distributions were constructed for the held-out sub-dataset. The process of fitting the BGLM while holding out one sub-dataset was repeated for all sub-datasets, thus constructing posterior predictive distributions for all $N \times T$ metabolite levels.

The data used in the study was taken from the A-Diet Confirm study [18,19], for which ethical approval was granted by the University College Dublin Sciences Human Research Ethics Committee (LS-16-91-Gibbons-Brennan). Briefly, the A-Diet Confirm study was designed to examine the habitual dietary intake of participants during a four month period with collection of biological samples and dietary data once a month. For the present study, the amino acid data was used from the previously reported dataset [18]. The data was acquired using the Biocrates AbsoluteIDQ p180 kit on a Sciex QTRAP 6500+ mass spectrometer coupled to a UHPLC column. The amino acids were quantified using isotopically labelled internal standards and seven-point calibration curves in AB Sciex Analyst version 1.7.2 software. In the quality control analysis, all amino acids had CV% less than 20% in pooled samples [18].

2.2. Simulation Study Design

Different simulation scenarios were constructed to explore the performance of MetaboVariation. Every scenario considered different settings of the parameters α and ν, where a proportion α of individuals was selected, and for each selected individual, the variation in one of their randomly selected repeated time points was inflated by the factor ν. For each scenario, 30 datasets were simulated to assess robustness of the method.

The simulated data were constructed such that they were similar to the real-life data of fasting plasma samples from the A-Diet Confirm study [18,19]. Each simulated dataset contained four repeated measurements of a single metabolite for 150 individuals with three covariates. Two covariates were simulated from Gaussian distributions, with means and standard deviations estimated from the age and BMI covariates in the real data of fasting plasma samples; similarly, the third covariate was simulated from a binomial distribution with a similar male to female ratio as in the real data of fasting plasma samples. A BGLM was fitted to alanine metabolite levels from the real data using the three simulated covariates, age, sex and BMI. The estimates of the regression coefficients $\hat{\beta}_0$ and $\hat{\beta}_1$, error variance ($\hat{\sigma}_\epsilon^2$) and variance at an individual level ($\hat{\sigma}^2$) were used along with the simulated covariates to simulate metabolite levels across four repeated measurements using (1), where $s_{it} \sim N(0, \hat{\sigma}^2)$ and $\epsilon_{it} \sim N(0, \hat{\sigma}_\epsilon^2)$.

Once the metabolite levels were simulated, a metabolite level was generated from $N(0, \hat{\sigma}^2 \nu)$, where ν was the inflation factor. This metabolite level was added to the existing metabolite level at one randomly selected time point for each of the randomly selected proportion α of individuals. We ensured that the inflated metabolite levels lay outside the $[1, 99\%]$ interval of the non-inflated metabolite levels. After forming the simulated dataset, MetaboVariation was fitted to the data and the number of flagged individuals with inflated variance was assessed. This process was repeated over the 30 simulated datasets.

3. Results

MetaboVariation analyses repeated metabolite measurements and flags individuals with different levels in at least one of these repeated measurements. Before detailing our results, we provide an overview of the MetaboVariation method. Using repeated measures from individuals, MetaboVariation uses a Bayesian generalised linear model (BGLM) that models metabolite levels as a function of covariates while accounting for intra-individual variation. If no covariates are provided, the model will only consider the repeated measurements of metabolite levels. The BGLM generates a posterior predictive distribution

for each individual at each time point. When an individual's observed metabolite level falls outside a specified percentage, e.g., the 95% highest posterior density (HPD) prediction interval, at a given time point, the model flags that particular individual as having intra-individual variation in their metabolite levels. Figure 1 provides an overview of the MetaboVariation method.
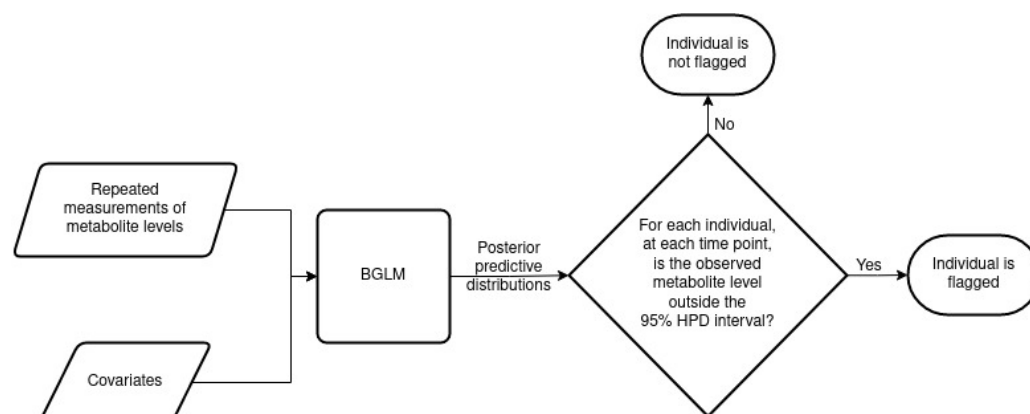
Figure 1. An overview of the MetaboVariation method.

3.1. Simulation Study

The objective of the simulation study was to assess the robustness and performance of the MetaboVariation method. Thirty simulated datasets were created, each having four repeated measures of one metabolite from 150 individuals with three covariates (see Section 2.2 for specific details). A proportion α of individuals was selected, and for each selected individual, the variation in one of their randomly selected repeated measures was inflated by the factor ν. Varying the values of α and ν allows assessment of the robustness and performance of MetaboVariation; $\alpha \in \{10\%, 20\%\}$ and $\nu \in \{5, 10, 25\}$ were considered here. The performance of MetaboVariation was assessed through examination of the proportion of individuals with known intra-individual variations that were correctly flagged across the 30 simulated datasets.

At $\alpha = 10\%$, when ν was increased from 5 to 10 to 25, the mean proportion (standard deviation in parentheses) of individuals flagged correctly was 0.97 (0.05), 0.96 (0.04) and 0.95 (0.05), respectively. These results suggest that across different variance inflation factors, when 10% of the individuals have intra-individual variation in their metabolite levels, MetaboVariation is able to correctly flag individuals that have intra-individual variability in their metabolite levels.

At $\alpha = 20\%$, the mean proportion (standard deviation in parentheses) of individuals flagged correctly was 0.9 (0.06), 0.89 (0.05) and 0.87 (0.05) for a ν of 5, 10 and 25, respectively. These results imply that as the proportion of individuals that have intra-individual variation in their metabolite levels increases, it is more difficult for MetaboVariation to discriminate between increased and minor intra-individual variation.

For every setting of α and ν for each of the 30 simulated datasets, convergence was assessed by running four chains and calculating the PSRF for each of the four parameters, $\beta_0, \beta_1, \sigma^2$ and σ_ϵ^2. Across all cases, the average PSRF was 1.0003 (standard deviation 0.0004).

3.2. Case Study: Fasting Plasma Samples from A-Diet Confirm Study

Fasting plasma samples from the A-Diet Confirm study [18,19] were collected monthly over four consecutive months in a free-living population. For this study, twenty metabolites were used from the original data and all 164 participants had at least two repeated measures. The participants had an average BMI of 24.1 ± 3.07 kg/m^2 and average age of 35 ± 12.6 years. Data from a total of 53 men and 111 women were analysed.

When MetaboVariation was applied to the data, 28% of individuals were flagged as having intra-individual variation in at least three of the twenty metabolites. A circos plot for the metabolite alanine is shown in Figure 2.
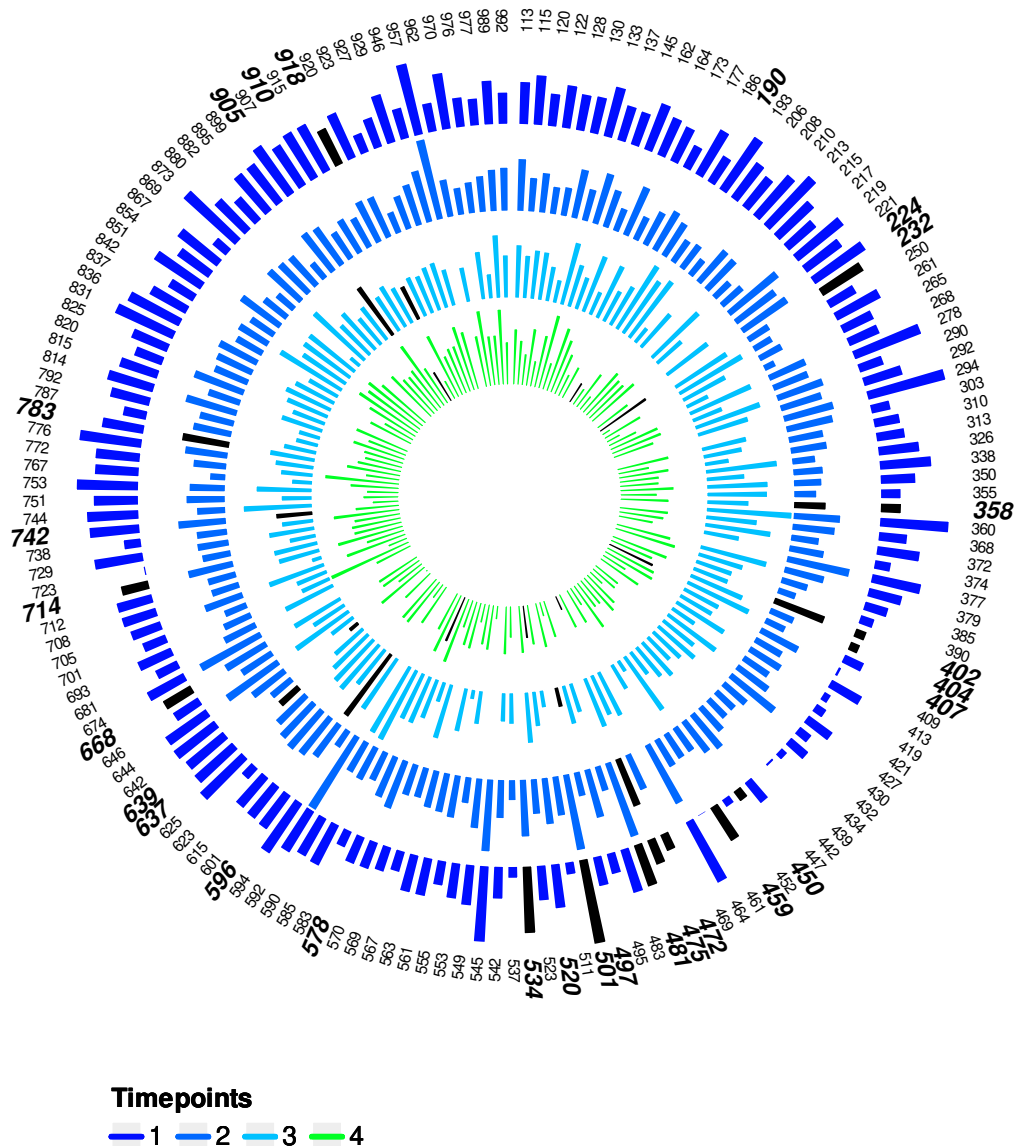
Figure 2. Circos plot for alanine. Each individual's label is shown in the outer circle, with bold font indicating those who have been flagged. The length of a bar within a time point has been scaled to represent the width of the central 95% posterior predictive interval. Flagged individuals have a black bar at the flagged time point. No bar indicates that the individual had no data for that time point.

There was a significant relationship between several metabolites and sex, whereas some metabolites had a significant relationship with age and BMI. Table 1 shows the estimates of the regression coefficients for sex, age and BMI for each metabolite, with 95% credible intervals. To assess convergence, four chains were run for each of the twenty metabolites, with an average PSRF of 1.0003 (standard deviation 0.0003) for each of the parameters (β_0, β_1, σ^2 and σ_ϵ^2).

Table 1. Estimates of all regression coefficients for each metabolite, with 95% credible intervals in parentheses.

Metabolite	Sex	Age	BMI
Ala	−4.41 (−26.17, 17.41)	−0.57 (−1.35, 0.2)	2.65 (−0.69, 5.95)
Arg	**−5.49** (−10.35, −0.65)	**0.19** (0.01, 0.37)	−0.64 (−1.41, 0.14)
Asn	−1.32 (−4.8, 2.17)	**−0.16** (−0.28, −0.03)	−0.27 (−0.82, 0.29)
Cit	**−3.82** (−5.81, −1.77)	**0.13** (0.05, 0.2)	−0.3 (−0.61, 0.02)
Gln	**−38.53** (−61.06, −16.57)	**1.57** (0.79, 2.36)	−3.03 (−6.46, 0.46)
Glu	**−8.38** (−13.07, −3.73)	**0.26** (0.1, 0.43)	**1.35** (0.63, 2.08)
Gly	15.45 (−9.12, 40.1)	**0.95** (0.06, 1.85)	−2.61 (−6.51, 1.34)
His	-0.89 (−4.26, 2.56)	−0.06 (−0.18, 0.06)	0.37 (−0.17, 0.89)
Ile	**−19.65** (−23.48, −15.77)	**−0.26** (−0.4, −0.12)	**0.79** (0.19, 1.37)
Leu	**−40.57** (−47.84, −33.33)	**−0.27** (−0.53, −0.02)	**1.71** (0.57, 2.87)
Lys	-8.88 (−17.9, 0.12)	**0.36** (0.03, 0.69)	−0.54 (−1.92, 0.86)
Met	**−2.92** (−4.14, −1.7)	−0.04 (−0.09, 0)	0.01 (−0.18, 0.2)
Orn	**−9.88** (−13.94, −5.88)	**0.29** (0.15, 0.44)	**−0.83** (−1.45, −0.21)
Phe	**−6.19** (−9.23, −3.1)	**−0.11** (−0.22, 0)	0.35 (−0.13, 0.81)
Pro	**−37.61** (−53.97, −21.57)	−0.37 (−0.95, 0.2)	2.37 (−0.22, 4.91)
Ser	**7.51** (0.17, 14.86)	−0.15 (−0.41, 0.11)	−0.87 (−1.99, 0.25)
Thr	4.06 (−3.8, 11.91)	**−0.34** (−0.62, −0.06)	−0.38 (−1.6, 0.8)
Trp	**−8.06** (−11.12, −4.95)	**−0.17** (−0.29, −0.06)	0.33 (−0.17, 0.81)
Tyr	**−5.33** (−9.31, −1.4)	0.07 (−0.08, 0.21)	**0.95** (0.33, 1.57)
Val	**−34.47** (−43.8, −25.42)	−0.06 (−0.41, 0.27)	1.21 (−0.25, 2.64)

Bold font indicates regression coefficients whose 95% credible interval does not include zero.

Figure 3 depicts the number of individuals flagged in each metabolite at each time point. The results indicate that some metabolites had many individuals with intra-individual variation in their metabolite levels. When individuals are flagged, they are counted only once, even if they are flagged at more than one time point for a metabolite. The metabolites leucine and ornithine had the highest number of individuals flagged across the four time points, whereas the metabolite asparagine had the lowest number of flagged individuals.
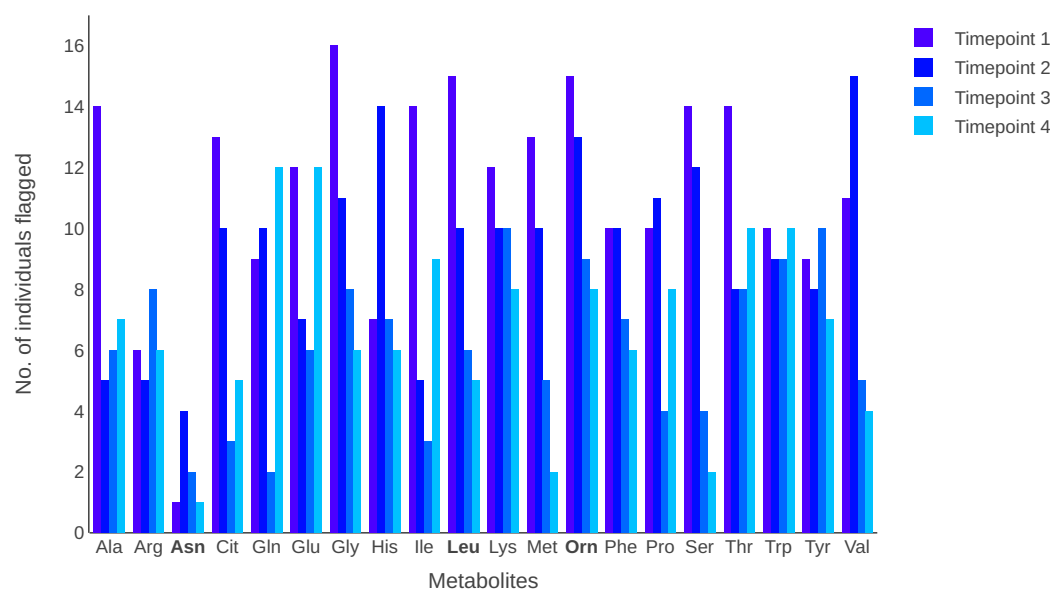
Figure 3. The number of individuals flagged for each metabolite. For each metabolite, a bar indicates a different time point. Bold font indicates metabolites with the highest or lowest number of individuals flagged across the four time points.

MetaboVariation was used on a Dell Latitude 5511 machine, running on Ubuntu 20.04 with an Intel 10th Gen i7 core and 16 GB memory. Applying MetaboVariation on a

single metabolite, running four different chains with 5000 iterations each and 2500 burn-in iterations, took approximately 2 minutes. Execution time varies depending on the system, number of chains, number of iterations and the number of individuals present in the data.

4. Discussion

There has been an increased interest in personalised healthcare, where an individual receives dietary advice tailored to their characteristics [20–24]. Omics technologies, such as metabolomics, metagenomics, proteomics and transcriptomics, have recently become highly relevant in pursuing these goals [25]. For metabolomics data, we developed Metabo-Variation to flag individuals with intra-individual variations in their metabolite levels. To address this problem, we present a conceptual method that employs repeated measurements of individual metabolite levels to understand intra-individual variances, and flag those with significant variations in their metabolite levels. An advantage of MetaboVariation is that it does not rely on generalised metabolite ranges and models the data at an individual level while also considering covariates. Further, an R package "MetaboVariation" was developed in conjunction with an R shiny application to facilitate widespread use of the method.

With a focus on personalised medicine, a team of researchers [26] developed the individual reference interval (IRI), which provides an interval for a particular variable based on an individual's biological characteristics. Population reference intervals (PRI) [27], which are derived from a healthy reference population, are crucial for interpreting clinical laboratory tests. Clinical choices today are typically binary: if the observation falls within the PRI, it is deemed normal. The IRI range adds a personal context, extending this interpretation to the individual level. Individual reference intervals describe test findings that, in the case of a healthy individual, would be anticipated to have a probability of about 95%. The key benefit of our study is the absence of a general range underpinning the prediction of whether or not an individual had intra-individual variations in their metabolite levels. This work has potential for the development of personalised healthcare. Continuous monitoring is becoming an embedded aspect of personalised healthcare, with the development of sensors such as continuous glucose monitors. With respect to metabolomics data, repeated measures from individuals over time offer an opportunity to identify the onset of metabolic perturbations and signal for the need of lifestyle interventions. The ability to do this at an individual level, without the need for a population reference value, indicates an important step forward for the development of personalised approaches.

Using the plasma data from 164 individuals [18], the two metabolites leucine and ornithine had the highest number of individuals flagged by MetaboVariation, whereas the metabolite asparagine had the lowest number of individuals flagged. Previous work [28] reported an excellent ICC for asparagine across two measurements. This agrees well with the low frequency of selection of individuals for asparagine by the MetaboVariation method. With respect to the two metabolites with the highest number of individuals flagged, the ICCs previously reported were good, indicating that our method is capable of picking up individual variation that may be missed with an ICC approach [28,29].

Though the present work represents a significant step forward, some limitations do exist. For example, MetaboVariation fits a single metabolite at a time, thus neglecting the correlations between different metabolites. Future work should aim to incorporate these correlations into the model. Further, MetaboVariation assumes that metabolite levels are Gaussian-distributed at each time point, and this assumption may be invalid in some cases. Future work could consider more flexible versions of the MetaboVariation method where a variety of distributions for metabolite levels are permitted. Though the computational costs of fitting the MetaboVariation method to a single metabolite is of the order of minutes, fitting it to larger datasets with high numbers of metabolites could become computationally prohibitive; approximate Bayesian inference approaches could assist in easing this burden.

5. Conclusions

In this study, we developed MetaboVariation, a method that uses the repeated measures of metabolite levels in individuals to understand intra-individual variations and flag individuals with significant variations in their metabolite levels. We examined the performance of MetaboVariation through a simulation study and, on applying MetaboVariation to real data, individuals with intra-individual variations in their metabolite levels were flagged. These findings make substantial progress in developing strategies for analysing metabolomics data at an individual level, thus paving the way toward personalised healthcare. Implementation of MetaboVariation by the wider community is made possible through the provision of open-source software.

Author Contributions: Conceptualization, L.B. and I.C.G.; Methodology, S.G., L.B. and I.C.G.; Validation, S.G., L.B. and I.C.G.; Formal Analysis, S.G. and I.C.G.; Writing—Original Draft Preparation, S.G.; Writing—Review & Editing, L.B. and I.C.G.; Visualization, S.G.; Supervision, L.B. and I.C.G.; Project Administration, L.B. and I.C.G.; Funding Acquisition L.B. and I.C.G. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: The data used in the study was taken from the A-Diet Confirm study by Yin et al. (2022) for which ethical approval was granted by University College Dublin Sciences Human Research Ethics Committee (LS-16-91-Gibbons-Brennan).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Restrictions apply to the availability of these data. They were obtained from the A-Diet Confirm study [18,19] and can be accessible through request from the corresponding author L.B.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Du, Y.; Chen, L.; Li, X.S.; Li, X.L.; Xu, X.D.; Tai, S.B.; Yang, G.L.; Tang, Q.; Liu, H.; Liu, S.H.; et al. Metabolomic Identification of Exosome-Derived Biomarkers for Schizophrenia: A Large Multicenter Study. *Schizophr. Bull.* **2021**, *47*, 615–623. [CrossRef] [PubMed]
2. Eroglu, E.C.; Tunug, S.; Geckil, O.F.; Gulec, U.K.; Vardar, M.A.; Paydas, S. Discovery of metabolomic biomarkers for discriminating platinum-sensitive and platinum-resistant ovarian cancer by using GC-MS. *Eur. J. Mass Spectrom.* **2021**, *27*, 235–248. [CrossRef] [PubMed]
3. Kalhan, S.C.; Guo, L.; Edmison, J.; Dasarathy, S.; McCullough, A.J.; Hanson, R.W.; Milburn, M. Plasma metabolomic profile in nonalcoholic fatty liver disease. *Metab. Clin. Exp.* **2011**, *60*, 404–413. [CrossRef] [PubMed]
4. Dambrova, M.; Makrecka-Kuka, M.; Kuka, J.; Vilskersts, R.; Nordberg, D.; Attwood, D.; Attwood, M.M.; Smesny, S.; Sen, Z.D.; Guo, A.C.; et al. Acylcarnitines: Nomenclature, Biomarkers, Therapeutic Potential, Drug Targets, and Clinical Trials. *Am. Soc. Pharmacol. Exp. Ther.* **2022**, *74*, 506–551. [CrossRef]
5. Ramzan, I.; Ardavani, A.; Vanweert, F.; Mellett, A.; Atherton, P.J.; Idris, I. The Association between Circulating Branched Chain Amino Acids and the Temporal Risk of Developing Type 2 Diabetes Mellitus: A Systematic Review and Meta-Analysis. *Nutrients* **2022**, *14*, 4411. [CrossRef]
6. Merino, J. Precision nutrition in diabetes: When population-based dietary advice gets personal. *Diabetologia* **2022**, *65*, 1839–1848. [CrossRef]
7. Jardon, K.M.; Canfora, E.E.; Goossens, G.H.; Blaak, E.E. Dietary macronutrients and the gut microbiome: A precision nutrition approach to improve cardiometabolic health. *Gut* **2022**, *71*, 1214–1226. [CrossRef]
8. Chaleckis, R.; Murakami, I.; Takada, J.; Kondoh, H.; Yanagida, M. Individual variability in human blood Metabolites identifies age-related differences. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 4252–4259. [CrossRef]
9. Guida, F.; Tan, V.Y.; Corbin, L.J.; Smith-Byrne, K.; Alcala, K.; Langenberg, C.; Stewart, I.D.; Butterworth, A.S.; Surendran, P.; Achaintre, D.; et al. The blood metabolome of incident kidney cancer: A case-control study nested within the MetKid consortium. *PLoS Med.* **2021**, *18*, e1003786. [CrossRef]

10. Imamura, F.; Fretts, A.; Marklund, M.; Ardisson Korat, A.V.; Yang, W.S.; Lankinen, M.; Qureshi, W.; Helmer, C.; Chen, T.A.; Wong, K.; et al. Fatty acid biomarkers of dairy fat consumption and incidence of type 2 diabetes: A pooled analysis of prospective cohort studies. *PLoS Med.* **2018**, *15*, e1002670. [CrossRef]

11. Jansen, J.J.; Szymańska, E.; Hoefsloot, H.C.J.; Smilde, A.K. Individual differences in metabolomics: Individualised responses and between-metabolite relationships. *Metabolomics* **2012**, *8*, 94–104. [CrossRef]

12. Chen, R.; Mias, G.I.; Li-Pook-Than, J.; Jiang, L.; Lam, H.Y.; Chen, R.; Miriami, E.; Karczewski, K.J.; Hariharan, M.; Dewey, F.E.; et al. Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell* **2012**, *148*, 1293–1307. [CrossRef]

13. Piening, B.D.; Zhou, W.; Contrepois, K.; Röst, H.; Gu Urban, G.J.; Mishra, T.; Hanson, B.M.; Bautista, E.J.; Leopold, S.; Yeh, C.Y.; et al. Integrative Personal Omics Profiles during Periods of Weight Gain and Loss. *Cell Syst.* **2018**, *10*, 2405–4712. [CrossRef]

14. McFarquhar, M. Modeling Group-Level Repeated Measurements of Neuroimaging Data Using the Univariate General Linear Model. *Front. Neurosci.* **2019**, *13*, 352. [CrossRef]

15. Hadfield, J.D. MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *J. Stat. Softw.* **2010**, *33*, 1–22. [CrossRef]

16. Hastings, W.K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **1970**, *57*, 97–109. [CrossRef]

17. Gelman, A.; Carlin, J.B.; Stern, H.S.; Dunson, D.B.; Vehtari, A.; Rubin, D.B. *Bayesian Data Analysis Third Edition (With Errors Fixed as of 13 February 2020)*; CRC Press: Boca Raton, FL, USA, 2013.

18. Yin, X.; Prendiville, O.; McNamara, A.; Brennan, L. Targeted Metabolomic Approach to Assess the Reproducibility of Plasma Metabolites over a Four Month Period in a Free-Living Population. *J. Proteome Res.* **2022**, *21*, 683–690. [CrossRef]

19. Prendiville, O.; Walton, J.; Flynn, A.; Nugent, A.P.; McNulty, B.A.; Brennan, L. Classifying Individuals Into a Dietary Pattern Based on Metabolomic Data. *Mol. Nutr. Food Res.* **2021**, *65*, e2001183. [CrossRef]

20. J Stewart–Knox, B.; Rankin, A.; P Bunting, B.; J Frewer, L.; Celis-Morales, C.; M Livingstone, K.; Fischer, A.R.H.; Poínhos, R.; Kuznesof, S.; J Gibney, M.; et al. Self-efficacy, habit strength, health locus of control and response to the personalised nutrition Food4Me intervention study. *Br. Food J.* **2022**, *124*, 314–330. [CrossRef]

21. Dello Russo, M.; Russo, P.; Rufián-Henares, J.Á.; Hinojosa-Nogueira, D.; Pérez-Burillo, S.; de la Cueva, S.P.; Rohn, S.; Fatouros, A.; Douros, K.; González-Vigil, V.; et al. The Stance4Health Project: Evaluating a Smart Personalised Nutrition Service for Gut Microbiota Modulation in Normal- and Overweight Adults and Children with Obesity, Gluten-Related Disorders or Allergy/Intolerance to Cow's Milk. *Foods* **2022**, *11*, 1480. [CrossRef]

22. Wilson-Barnes, S.; Hart, K.; Gymnopoulos, L.P.; Dimitropoulos, K.; Solachidis, V.; Rouskas, K.; Russell, D.; Oikonomidis, I.; Hadjidimitriou, S.; Botana, J.M.; et al. Personalised Nutrition for Healthy Living: The PROTEIN Project. *Nutr. Bull.* **2021**, *46*, 77–87. [CrossRef]

23. San-Cristobal, R.; Navas-Carretero, S.; Celis-Morales, C.; Livingstone, K.M.; Stewart-Knox, B.; Rankin, A.; Macready, A.L.; Fallaize, R.; O'Donovan, C.B.; Forster, H.; et al. Capturing health and eating status through a nutritional perception screening questionnaire (NPSQ9) in a randomised internet-based personalised nutrition intervention: The Food4Me study. *Int. J. Behav. Nutr. Phys. Act.* **2017**, *14*, 168. [CrossRef]

24. Livingstone, K.M.; Celis-Morales, C.; Navas-Carretero, S.; San-Cristobal, R.; Forster, H.; Woolhead, C.; O'Donovan, C.B.; Moschonis, G.; Manios, Y.; Traczyk, I.; et al. Characteristics of participants who benefit most from personalised nutrition: Findings from the pan-European Food4Me randomised controlled trial. *Br. J. Nutr.* **2020**, *123*, 1396–1405. [CrossRef] [PubMed]

25. Brennan, L.; De Roos, B. Nutrigenomics: Lessons learned and future perspectives. *Am. J. Clin. Nutr.* **2021**, *113*, 503–516. [CrossRef] [PubMed]

26. Pusparum, M.; Ertaylan, G.; Thas, O. Individual reference intervals for personalised interpretation of clinical and metabolomics measurements. *J. Biomed. Inform.* **2022**, *131*, 104111. [CrossRef]

27. Pusparum, M.; Ertaylan, G.; Thas, O. From Population to Subject-Specific Reference Intervals. In *Proceedings of the Computational Science—ICCS 2020, Amsterdam, The Netherlands, 3–5 June 2020*; Krzhizhanovskaya, V.V., Závodszky, G., Lees, M.H., Dongarra, J.J., Sloot, P.M.A., Brissos, S., Teixeira, J., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 468–482.

28. Carayol, M.; Licaj, I.; Achaintre, D.; Sacerdote, C.; Vineis, P.; Key, T.J.; Onland Moret, N.C.; Scalbert, A.; Rinaldi, S.; Ferrari, P. Reliability of Serum Metabolites over a Two-Year Period: A Targeted Metabolomic Approach in Fasting and Non-Fasting Samples from EPIC. *PLoS ONE* **2015**, *10*, e0135437. [CrossRef]

29. Floegel, A.; Drogan, D.; Wang-Sattler, R.; Prehn, C.; Illig, T.; Adamski, J.; Joost, H.G.; Boeing, H.; Pischon, T. Reliability of serum metabolite concentrations over a 4-month period using a targeted metabolomic approach. *PLoS ONE* **2011**, *6*, e21103. [CrossRef]