Intersubject and intrasubject variability of potential plasma and urine metabolite and protein biomarkers in healthy human volunteers

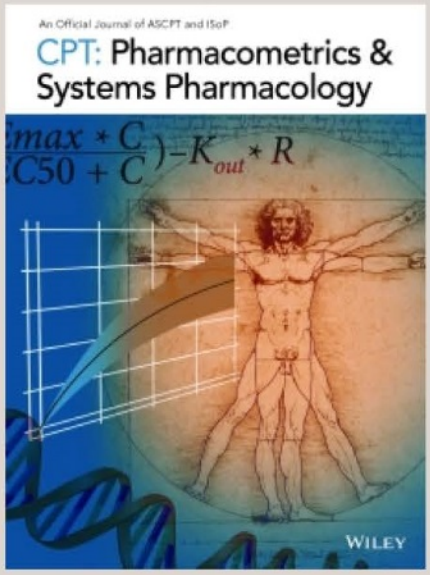
Duisters, K.L.W.; Ogino, S.; Andou, T.; Ito, K.; Akabane, T.; Harms, A.C.; ... ; Hankemeier, T.

CPT: PSP 1st Impact Factor - Coming soon

Why publish open access in *CPT: Pharmacometrics & Systems Pharmacology (PSP)*?

- **Open Access** official journal of ASCPT
- **Author support** with ASCPT tools to amplify research
- **Cross- disciplinary** publishing advances in quantitative methods
- **Easy submissions** with direct access to editorial office support

Submit online at
https://psp.msubmit.net/cgi-bin/main.plex



ASCPT

WILEY

Intersubject and Intrasubject Variability of Potential Plasma and Urine Metabolite and Protein Biomarkers in Healthy Human Volunteers

Kevin Duisters[1,*], Shinji Ogino[2,†], Tomohiro Andou[3,†], Kazumi Ito[4,†], Takafumi Akabane[2], Amy Harms[5], Matthijs Moerland[6], Yuka Hashimoto[2], Ayumi Ando[3], Yoshiaki Ohtsu[2], Naoya Wada[7], Hideo Yukinaga[4], Jacqueline Meulman[1,8], Hiroyuki Kobayashi[3], Nobuhiro Kobayashi[4], Kenichi Suzumura[2] and Thomas Hankemeier[5]

A limited understanding of intersubject and intrasubject variability hampers effective biomarker translation from *in vitro/in vivo* studies to clinical trials and clinical decision support. Specifically, variability of biomolecule concentration can play an important role in interpretation, power analysis, and sampling time designation. In the present study, a wide range of 749 plasma metabolites, 62 urine biogenic amines, and 1,263 plasma proteins were analyzed in 10 healthy male volunteers measured repeatedly during 12 hours under tightly controlled conditions. Three variability components in relative concentration data are determined using linear mixed models: *between* (intersubject), *time* (intrasubject), and *noise* (intrasubject). Biomolecules such as 3-carboxy-4-methyl-5-propyl-2-furanpropanoate, platelet-derived growth factor C, and cathepsin D with low *noise* potentially detect changing conditions *within* a person. If also the *between* component is low, biomolecules can easier differentiate conditions *between* persons, for example cathepsin D, CD27 antigen, and prolylglycine. Variability over *time* does not necessarily inhibit translatability, but requires choosing sampling times carefully.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

☑ When biomolecule variability of plasma or urine metabolites is segmented into *between* subject, *time*, and *noise* components, *noise* tends to be relatively large. This variability can jeopardize translation from *in vitro/in vivo* situations to clinical trials or clinical decision support. Prediction of success on possible biomarker translation from *in vitro/in vivo* to humans is highly desirable.

WHAT QUESTION DID THIS STUDY ADDRESS?

☑ How can an integrative approach using *absolute* variance components on individual biomolecules inform clinical trial design and translatability from *in vitro/in vivo* to humans?

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

☑ This work is the first to study *absolute* variance components of a wide array of 2,074 biomolecules (metabolites and proteins) in different sample compartments (plasma and urine) simultaneously. New tools supporting translation of potential biomarkers are presented based on individual biomolecule variability of healthy volunteers.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?

☑ Our contributions include important implications for designation of meal/sampling times, statistical power analysis, and selection of biomarkers with potential for translation to humans.

In many recently published studies, metabolomics and proteomics have been used to find explorative biomarkers in plasma or urine for disease diagnosis, disease progression, or treatment outcome prediction.[1–3] Moreover, biomarkers are increasingly applied in clinical and preclinical studies to support drug discovery and drug development. Finally, metabolomics and

[1]Mathematical Institute, Leiden University, Leiden, The Netherlands; [2]Astellas Pharma Inc., Tokyo, Japan; [3]Takeda Pharmaceutical Company Limited, Tokyo, Japan; [4]Daiichi Sankyo Co., LTD, Tokyo, Japan; [5]Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands; [6]Centre for Human Drug Research, Leiden, The Netherlands; [7]Daiichi Sankyo RD Novare Co., LTD, Tokyo, Japan; [8]Department of Statistics, Stanford University, Stanford, California, USA. *Correspondence: Kevin Duisters (k.l.w.duisters@math.leidenuniv.nl)

[†]These authors contributed equally to this work.

Received April 22, 2019; accepted July 20, 2019. doi:10.1002/cpt.1606

proteomics have shown potential to predict drug efficiency or side effects, and could, therefore, be used in principle for patient stratification. Ideally, biomarkers for target engagement, disease progression, or predicting drug efficacy/toxicity discovered and optimized in *in vivo* or *in vitro* models are translated to clinical studies and trials.

Identified biomarkers can only be utilized clinically if their measurement is scalable, reliable, and affordable, which is increasingly ensured by modern analytical techniques, such as mass spectrometry, nuclear magnetic resonance, and aptamer/antibody-based sensor arrays. However, the practical usability of biomarkers discovered in scientific studies has often been disappointing. Although metabolic and proteomic biomarkers have been used in drug research for decision support, so far only a few biomarkers discovered by metabolomics or proteomics were translated to clinical applications and used in clinical trials for stratification.[4] One important reason could be that, unlike the stability observed in genetics (a certain single nucleotide polymorphism, etc.), metabolites and proteins as biomarkers tend to be dynamic. Although these fluctuations are considered to be a tremendous source of information in fields such as personalized medicine,[5,6] they simultaneously hamper translation to clinical practice.[7]

Variance components

Fluctuations in biomolecule measurements can be divided into three mutually exclusive, commonly exhaustive variance components. Intersubject variability is generated by structural (biological) differences *between* experimental units (here: individuals). This component is a key predictor in many investigations, for example, in clinical studies discriminating cases vs. controls, in clinical trials or in clinical diagnostics predicting (non-)responders to a particular treatment. Therefore, intersubject variability is a desirable form of fluctuation when the goal is to differentiate individuals. Intrasubject variability finds its origin *within* the experimental unit. Two sources of intrasubject variability can be identified: *time* and *noise*. If an individual is measured at separate occasions in time, the *time* variance component describes the structural (biological) fluctuation of biomolecule abundance. Such changes could, for instance, be caused by food intake or the circadian rhythm. Without accounting for these patterns, the clinical reliability of biomarkers used to discriminate *between* groups may be reduced substantially. Finally, the *noise* component of variation is determined by a set of remaining factors impacting reproducibility of biomolecule abundance from the same subject at the same time of day. These factors could be biological (e.g., in samples collected on different days) and/or technical (e.g., due to small uncontrolled changes of analytical settings). Even if variability over time is known or under control, clinical decision making becomes challenging whenever the *noise* component is relatively large with respect to that *between* subjects. Summarizing, the fluctuation of biomolecule abundance can be described by: *between*, *time*, and *noise* variance.

Related literature

Recently, attention for biomarker variability has increased.[8–11] Kim *et al.*[8] studied the proportions of variance components in 121 plasma metabolites and 294 urine metabolites measured several times over 3 days. The study involved 13 healthy volunteers and 13 with autosomal dominant polycystic kidney disease not taking medication (46% men, 58% Caucasian, age 33 ± 10.9, and body mass index (BMI) 24.3 ± 3.3) in a tightly controlled environment. Their main findings were that the *noise* source of variation is relatively large, most *time* variability occurs in the morning, and metabolites show more variation in urine than plasma on average. Maitre *et al.*[9] studied 44 urine metabolites in 20 Caucasian children, measured in the morning and at night during 6 days. As in ref. 8, these authors report that variance proportions range widely from one metabolite to another and that the *noise* tends to outweigh the *between* subject and *time* component. Floegel *et al.*[10] link the *between* variance proportion in plasma to biomolecule family (high for hexose, sphingolipids, amino acids, and glycerophospholipids; low for most acylcarnitines). Nicholson *et al.*[11] studied fasting plasma (198 metabolites) and urine (328 metabolites) of 154 twins, all postmenopausal healthy Caucasian women. The authors conclude 47% (urine) to 60% (plasma) of biological variation to be stable on average. Sampson *et al.*[12] measured 385 plasma metabolites in 60 Asian postmenopausal women. Their findings suggest technical reliability to be high, but again *within* subject fluctuation dominated that *between* subjects. To our knowledge, our paper is the first to simultaneously investigate different sources of fluctuation (*between*, *time*, and *noise*) of a wide array of biomolecules (metabolites and proteins) in different sample compartments (plasma and urine) by an integrative approach using *absolute* variance components.

Study design

This paper studies 10 healthy, Caucasian men (age 23.8 ± 3.1; BMI 23.3 ± 2.4) in a tightly controlled environment comparable to a phase I clinical trial design. The study was conducted by the Centre for Human Drug Research Leiden in conformity with the study plan approved by the Ethical Committee. EDTA plasma and urine were collected several times over the course of 12 hours. Volunteers entered the facility 1 day prior to sample collection, upon which temperature, food intake, and exercise were standardized. **Table 1** lists sample collection and food intake times used

Table 1 Study design: sample and food intake times

			09:00				12:00				16:30			19:00		
Panel	−2	−0.5	0	0.5	1.5	3	3.5	4	4.5	6	7.5	8	9.5	10	11	12
A		x		x	x	x			x	x			x		x	x
B	x		**Breakfast**				**Lunch**	x			**Snack**	x		**Dinner**		x
C		x		x	x	x				x			x			x

A, plasma metabolites; B, urine amines; C, plasma proteins.

to obtain our results. Three platforms were deployed to analyze the samples: 749 plasma metabolites were measured on a global metabolomics platform (A: Metabolon), 62 urine biogenic amines were measured using a targeted metabolomics platform of the Biomedical Metabolomics Facility Leiden (B: BMFL), and 1,263 proteins were measured in plasma using an aptamer-based proteomics platform (C: SomaLogic). Design and protocol details are included as Supplemental Information.

RESULTS

The Introduction has identified three important explanations of biomolecule variability: *between*, *time*, and *noise*. In this section, these components are analyzed and their relation to underlying biochemical processes and clinical practice is studied. One of the novelties of this paper is the use of a statistical method that allows preventing overestimation of *noise*. It does so by accounting for covariates age and BMI, nonequally spaced measurement times, and dynamic covariance matrix structures that capture correlations in samples from the same individual.

Quantifying variability

The statistical model estimates *between*, *time*, and *noise* variance components for each of the (749 + 62 + 1,263 = 2,074) biomolecules. Importantly, the relative concentration data have been normalized such that the *absolute* variance components may be compared across metabolites and proteins in plasma and urine.

Figure 1 illustrates *between* subject variability on the vertical axis against *noise* variability on the horizontal axis. Variability over *time* (σ_τ) is visualized proportional to the area of the circle, colored by platform.

Clinical translation

The overview of variance components in **Figure 1** can be used to identify biomolecules that have potential for clinical translation, especially if the effect size of a biomolecule as biomarker is known. Generally speaking, biomolecules with a low *noise* variability may have potential to detect changing conditions *within* a person. **Table 2** lists the "top" 10 (of 2,074) biomolecules with the lowest *noise* square root variance component (σ_ε). Note that a biomolecule with a low *noise* component can still have high *between* and/or *time* variability.

If in addition to the *noise*, also the variability *between* subjects is low, biomolecules may be able to differentiate conditions *between* persons as biomarkers. In **Figure 1**, biomolecules with a low sum of the *between* and *noise* square root variance component ($\sigma_b + \sigma_\varepsilon$) can be found in the bottom left corner. **Table 3** again lists their "top" 10. A complete list of results is available in **Table S1**.

To understand the practical implications of these variance components, **Figure 2** exhibits four biomolecule profiles as an example, including 3-carboxy-4-methyl-5-propyl-2-furanpropanoate (CMPF; from **Table 2**), cathepsin D (from **Table 3**), lactate (a

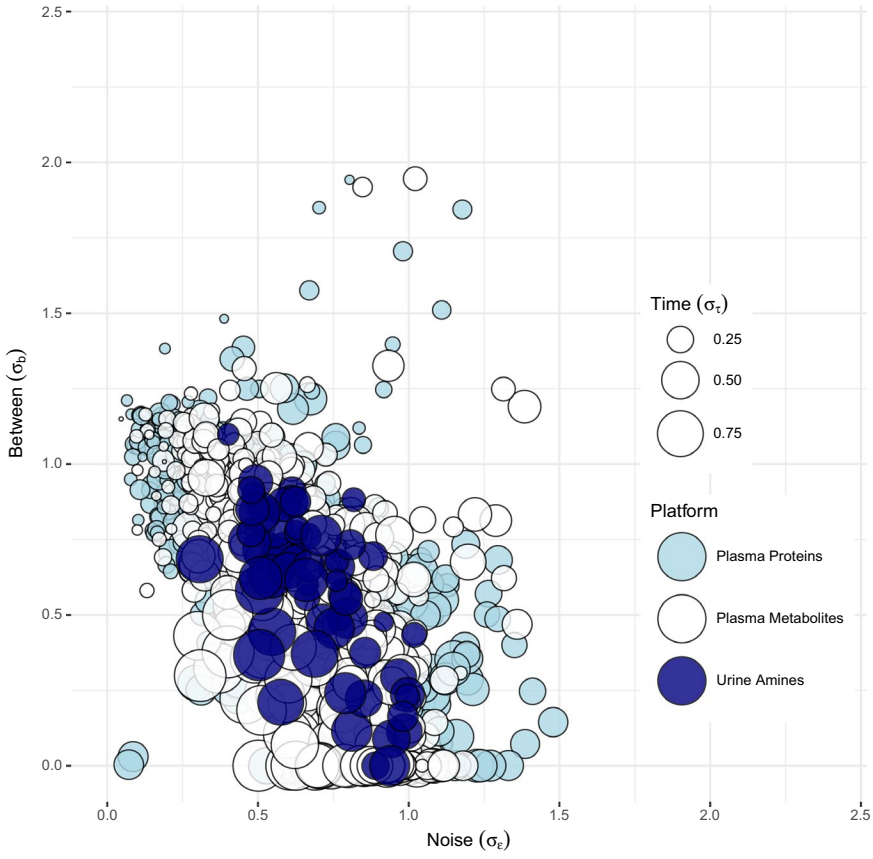


Figure 1 Variance components of each biomolecule on the square root scale (i.e., standard deviations). The area of each circle is proportional to its standard deviation over time (σ_τ), and colors indicate the analytical platform.

Table 2 The top 10 biomolecules with lowest σ_ε

Biomolecule	Panel	Inter σ_b	Intra σ_τ	Intra σ_ε
3-Carboxy-4-methyl-5-propyl-2-furanpropanoate	A	1.15	0.02	0.05
Platelet-derived growth factor C	C	1.21	0.05	0.07
Cathepsin D	C	0.00	0.31	0.07
Mitogen-activated protein kinase 11	C	0.95	0.06	0.08
PILR alpha-associated neural protein	C	1.16	0.05	0.08
Anterior gradient protein 2 homolog	C	1.02	0.10	0.08
CD27 antigen	C	0.03	0.32	0.09
Delta-like protein 4	C	1.07	0.12	0.09
Artemin	C	1.06	0.12	0.10
Androsterone sulfate	A	1.09	0.07	0.10

All parameters are defined in detail in the Methods section.
A, plasma metabolites; B, urine amines; C, plasma proteins; σ_b represents the square root variance component *between* subjects, σ_τ that over *time*, and σ_ε the *noise*.

Table 3 The top 10 biomolecules with lowest ($\sigma_b + \sigma_\varepsilon$)

Biomolecule	Panel	Inter σ_b	Intra σ_τ	Intra σ_ε
Cathepsin D	C	0.00	0.31	0.07
CD27 antigen	C	0.03	0.32	0.09
Prolylglycine	A	0.00	0.93	0.50
Neuronal cell adhesion molecule	C	0.00	0.47	0.53
Retinol-binding protein 4	C	0.25	0.33	0.31
Cystatin-SA	C	0.27	0.45	0.29
Ubiquitin	C	0.00	0.43	0.59
Fructose	A	0.00	0.85	0.61
4-Acetylphenyl sulfate	A	0.30	0.95	0.31
Fatty acid-binding protein, liver	C	0.00	0.52	0.61

All parameters are defined in detail in the Methods section.
A, plasma metabolites; B, urine amines; C, plasma proteins; σ_b represents the square root variance component *between* subjects, σ_τ that over *time*, and σ_ε the *noise*.

known biomarker in plasma[13], treated in the Discussion), and L-valine (a known biomarker in urine[14]). In each plot, the intuition being that ($\sigma_b + \sigma_\varepsilon$) is roughly proportional to the length of the error bars, whereas σ_τ is linked to the fluctuation in mean profile. One may observe the remarkable difference of CMPF (large σ_b, very low σ_τ, and very low σ_ε) vs. cathepsin D (very low σ_b, moderate σ_τ, and very low σ_ε) due to *between* subject variability, and that of CMPF vs. lactate (low σ_b, high σ_τ, and moderate σ_ε) due to variability over *time*. The profile of L-valine (moderate σ_b, very low σ_τ, and moderate σ_ε) behaves roughly like that of CMPF, except it has been collected in urine less frequently. These profiles can guide the translation of biomarkers into clinical trial use and clinical decision support, and several strategies are provided in the Discussion section. For instance, a large variability over *time* does not necessarily dismiss a biomolecule from clinical translation because its impact can be mitigated by the designation of appropriate sampling times.

Time effects including food intake

A large *time* variance component is caused by fluctuation in the mean profile of a biomolecule, whereas a σ_τ close to zero corresponds to a nearly constant normalized relative concentration. A biomolecule can show interesting (nonconstant) behavior around times surrounding food intake or due to the circadian rhythm, for example. This can be clearly seen in a heatmap of the top 10 biomolecules with highest σ_τ across all platforms, as depicted in **Figure S1**. Here, several mean profiles seem to fluctuate around the morning and/or evening. More formally, **Figure S2** pinpoints statistically significant changes per data type (plasma metabolites, urine amines, and plasma proteins) based on comparisons of all pairwise differences over time along the estimated mean profile. By comparing all pairwise differences, including those between nonadjacent times, the analysis allows for heterogeneous dynamics. For instance, the effect of food intake may be noticeable instantly in metabolites related to energy, but could take several hours in lipids. The results suggest the plasma metabolites have relatively many pairwise differences, which is consistent with evidence from **Figure 1** and **Figure S1**. In particular, it seems, on average, that the profiles are different in the morning compared with both the afternoon as well as the evening. Moreover, for the plasma proteins, a clear distinction between the −0.5-hour and 1.5-hour samples appears. Again, the −0.5-hour vs. 12-hour difference stands out as well.

Variance components by biochemical class

To assess whether differences in underlying biochemical processes have a general impact on variance components, such as observed for the time effects, **Figure 3** illustrates variance components in plasma metabolites segmented by their biochemical class. Note that in each of the three dimensions, the segmentation is informative and separates most classes by median variability, except for lipids and amino acids, whose median *between*, *time*, and *noise* variance components are close. For example, cofactors and vitamins vary more *between* subjects than carbohydrates do, which may be explained by the fact that carbohydrate levels (aided by quick absorption and dissimilation) are more tightly controlled than endogenous cofactor and vitamin levels.

Variance proportions

Converting *absolute* variance components to variance proportions allows for a comparison with earlier findings reported by other studies in **Table 4**. Here, π_b, π_τ, and π_ε denote the average of the *between*, *time*, and *noise* variance proportions for each metabolomics/proteomics dataset. For comparability purposes, $\pi_w = \pi_\tau + \pi_\varepsilon$ is included, the *within* (or *intra*) subject variance proportion. The intraclass correlation coefficient (ICC), defined as $\sigma_b^2/(\sigma_b^2+\sigma_\tau^2+\sigma_\varepsilon^2)$, is summarized by its median across all

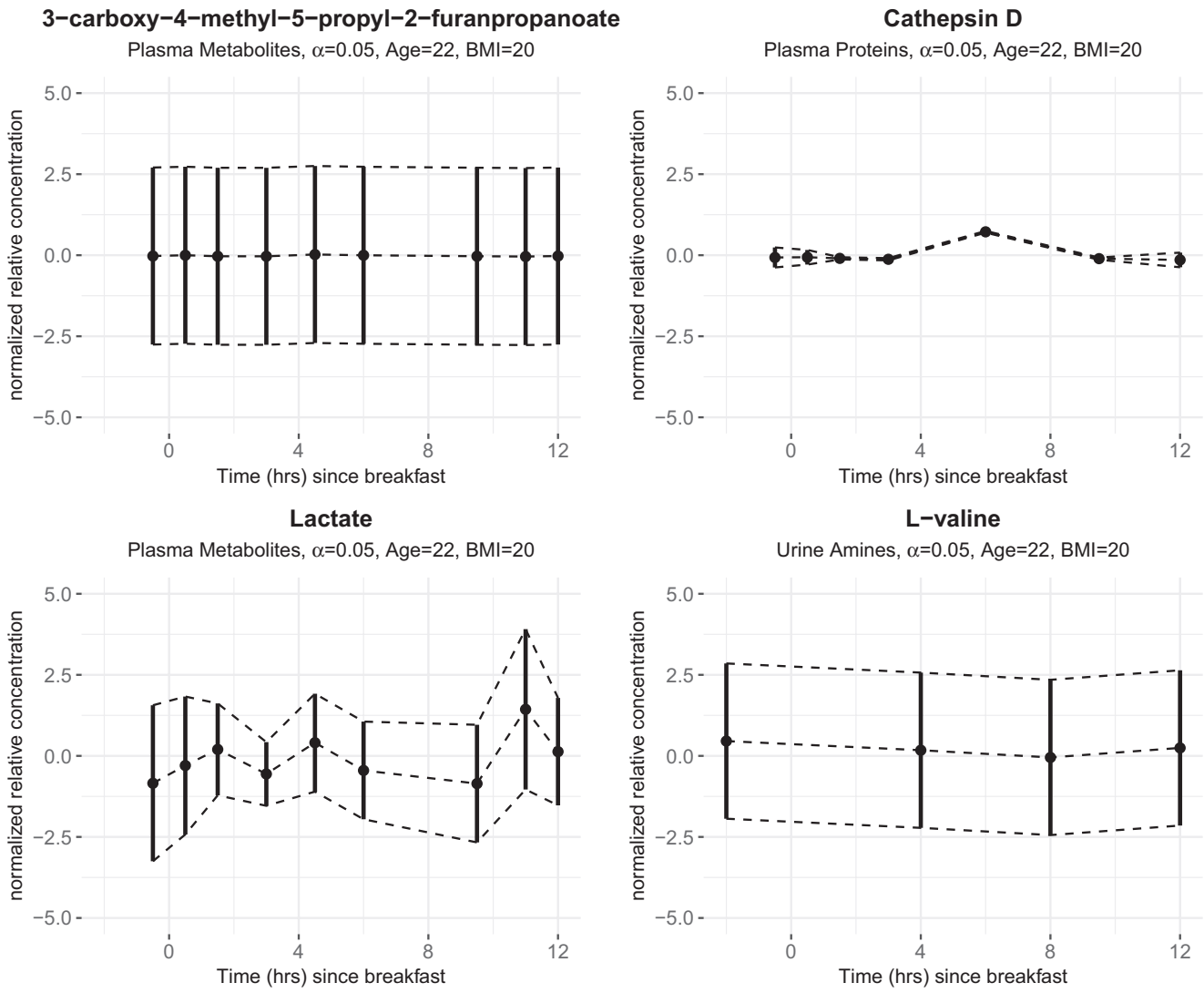


Figure 2 Mean profiles (dots) of normalized relative concentrations for a healthy man (age 22, body mass index (BMI) 20) including 95% pointwise prediction intervals for observed measurement occasions. Dashed lines only serve as illustration in between these sampling times. The following biomolecules (σ_b, σ_τ, and σ_ε) are depicted: 3-carboxy-4-methyl-5-propyl-2-furanpropanoate (1.15, 0.02, and 0.05) in plasma, cathepsin D (0.00, 0.31, and 0.07) in plasma, lactate (0.38, 0.73, and 0.67) in plasma, and L-valine (0.79, 0.21, and 0.63) in urine.

biomolecules in the respective dataset. A high π_b or ICC shows that the *between* variance component is relatively important in explaining total variability of a biomolecule as opposed to the *time* and *noise* components.

Table 4 shows that our variance proportions are comparable with those reported in the literature. This comparison of variance component proportions is merely indicative given differences in targeted biomolecules, populations under study, and deployed analytical methods. Furthermore, different preprocessing and normalization strategies were used. Nevertheless, the median ICC found by refs. 9,12 is comparable with the results established here; the median ICC of ref. 10 is somewhat higher. Compared against ref. 8, which is closest in study design, on average, our results seem to have captured more variance *between* subjects and *noise* has reduced relatively. A possible explanation could be that our statistical approach based on linear mixed models (see Methods) describes several aspects in detail that would otherwise have been classified as *noise* by analysis of variance (ANOVA)-style methods. Taking effects due to covariates of age and BMI, correlations between observations from the same individual, and mathematical flexibility for nonequally spaced sampling times (**Table 1**) into account, allows to filter more signal (*between*) from *noise* (*within*). **Figure S3** illustrates the distribution of variance proportions for the plasma metabolites, urine amines, and plasma proteins. However, although presenting a relevant summary, this paper intends to emphasize the added value contained in studying *absolute* variance components (e.g., **Figure 1**) in addition to these proportions. For instance, there is no direct relation between a biomolecule proportion or ICC and its position on the *absolute* (σ_ε, σ_b)-plane, which in many practical settings is of great importance as the Discussion will argue.

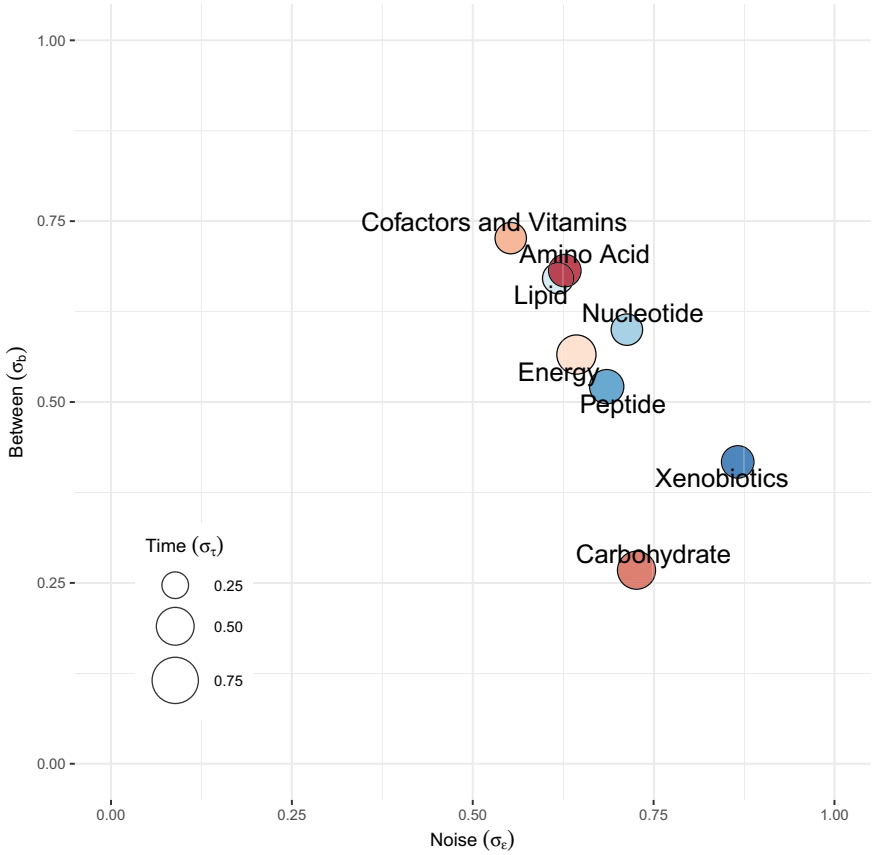


Figure 3 Median variance components for the plasma metabolites per biochemical class on the square root scale (i.e., standard deviations). The area of the circles is proportional to the class median standard deviation over time (σ_τ). Note the points for lipids (0.67, 0.62, and 0.34) and amino acids (0.68, 0.63, and 0.37) visually coincide.

Table 4 Variance proportions: overview of related results

			Mean				Median
Study	**Medium**	**Method**	π_b	π_τ	π_ε	π_w	**ICC**
Kim *et al.*[8]	Plasma	LC/MS	32	3	65	68	–
Kim *et al.*[8]	Urine	MS/MS	25	8	67	75	–
Maitre *et al.*[9]	Urine	^{1}H-NMR	–	–	–	–	40
Floegel *et al.*[10]	Plasma	MS/MS	51	–	–	49	57
Nicholson *et al.*[11]	Urine	^{1}H-NMR	43	7	51	58	–
Sampson *et al.*[12]	Plasma	(LC,GC)/MS	–	–	–	–	43
A: Metabolites	Plasma	MS	41	17	42	59	37
B: Amines	Urine	MS/MS	36	19	45	64	38
C: Proteins	Plasma	Aptamer	46	12	42	54	44

π_b, π_τ, and π_ε denote the average of the *between*, *time*, and *noise* variance proportions for each metabolomics/proteomics dataset. $\pi_w = \pi_\tau + \pi_\varepsilon$ represents the *within* (or *intra*) subject variance proportion. ICC is the intraclass correlation coefficient $(\sigma_b^2/(\sigma_b^2+\sigma_\tau^2+\sigma_\varepsilon^2))$, here, the median across all biomolecules in the respective dataset. Replication details for related literature are given as follows: in ref. 8 Table 2 (time of day), Patient = between; Meal = time; Residuals = noise; in ref. 10 Tables 2–5 were used to imply variance components; in ref. 11 plasma (all peaks) and urine (all peaks): $\pi_b = (\sigma_d^2+\sigma_m^2+\sigma_e^2)/\sigma^2$, $\pi_\tau = \sigma_w^2/\sigma^2$, $\pi_b = (\sigma_\varepsilon^2+\sigma_v^2)/\sigma^2$; with $\sigma^2 = \sigma_d^2+\sigma_m^2+\sigma_e^2+\sigma_w^2+\sigma_v^2+\sigma_\varepsilon^2$.
GC, gas chromatography; ICC, intraclass correlation coefficient; LC/MS, liquid chromatography/mass spectrometry; MS/MS, tandem mass spectrometry; NMR, nuclear magnetic resonance.

DISCUSSION

In this study, a widely used phase I clinical trial design was followed, with typical inclusion criteria, environmental control, and sample size. Variability components *between*, *time* (including food intake), and *noise* were quantified for a diverse set of metabolites in urine and plasma, and proteins in plasma. Knowledge of the *absolute* variance components per metabolite or protein is crucial in the process of biomarker development and translation as this

variation can jeopardize the use of a biomolecule as a biomarker. The size of the three variance components can predict whether a biomarker with a known or expected effect size will be able to differentiate *between* two human conditions, and whether it is necessary to control the variation due to *time*, or not. To illustrate the potential of such information in clinical practice, some use-cases and limitations are addressed.

Clinical trial design

Suppose that a clinical trial uses a stratification biomarker to include patients into the study and to predict the desired outcome of the intervention. In a translational setting, an effect size for humans is extrapolated from *in vitro*/*in vivo* studies and computational modeling. Combined with assumed (equal) case/control variances, this effect size is then used to estimate suitability of such a biomarker and/or power calculations of necessary sample sizes (e.g., refs. 11,12).

Instead, healthy baseline profiles (**Figure 2**) can be used to derive optimal sampling times in a clinical trial (i.e., to choose sampling times where the lowest sample size is needed or the statistical power is maximized). If one knows that the concentrations of a biomarker in question fluctuate heavily in the afternoon but do not around breakfast, it is reasonable to designate the sampling time accordingly to avoid variability of the concentration due to the chosen experimental design of the clinical study. Moreover, these concerns could be a reason for focusing attention on specific biomarkers (e.g., those with small variability associated with *time* would be preferable in pharmacokinetic/pharmacodynamic settings). Finally, biomarkers with small *between* subject variability are of greater concern in parallel designs, where each cohort is treated differently than in crossover clinical trials.

Clinical decision support and health monitoring

The second use-case is about disease diagnosis and/or prediction whether an intervention to treat, prevent, or retard a disease will result in the desired outcome. Again, knowing the *absolute* variation over *time* reveals whether choice of sampling time is important, and what the effect size has to be in relation to the known variation (*between* subjects and *noise*) in order for a biomarker to diagnose or predict properly at the individual level. One example is lactate in plasma, illustrated in **Figure 2**, which has been reported as indicative of acute illness.[13] As with clinical trial design, fluctuations of the mean profile and associated prediction interval lengths over time have a direct impact on whether an observed concentration should be classified as abnormal. Suppose a clinician diagnoses a Caucasian man of age 22 and BMI 20. On the normalized relative scale of **Figure 2**, a lactate level of 3.0 at 20:00 (10.5 hours since breakfast) seems fairly normal for a population of healthy, Caucasian, male volunteers of age 22 and BMI 20, whereas such a level around breakfast would fall well outside the estimated 95% prediction interval at that time, which could be a reason for concern. Given the low sample size, this example should be considered conceptual. Note that in practice it is more realistic to consider decision rules based on *absolute* concentrations, which is a straightforward next step after exploratory analysis using the methods presented in this paper.

Limitations

Although the study design resembles a typical phase I clinical trial, the small sample size ($N = 10$) could be a concern when generalizing findings. Although it is safe to assume trends in variance components observed here are a reasonable reflection of typical phase I clinical trials, one cannot claim to have estimated biomolecule profiles (e.g., **Figure 2**) over time including the influence of covariates of age and BMI precisely. A considerably larger sample size is required to describe an average biomolecule profile exactly than to claim a biomolecule changes (not) much *between* subjects, over *time*, or through *noise*. Furthermore, the study contained only healthy Caucasian men, which is possibly not generalizable to a broader population of (diseased) individuals.

CONCLUSION

Despite the aforementioned limitations, the statistical methodology and resulting biomolecule profiles contain very relevant information on potential fluctuations for drug and clinical researchers. Profiles indicating structural fluctuations over *time* on top of *between* subject variability and *noise* (**Figure 2**) are key in selection of sampling time(s) and conduct of pharmacokinetic/pharmacodynamic analysis. Therefore, we have developed an interactive, web-based tool to make these plots available for all biomolecules investigated here (http://hbt.analyticalbiosciences.nl). Looking ahead, it is worthwhile to investigate biological drivers of variability around food intake further. In addition to biomolecular class characteristics, pairwise differences observed in the data might result from diverging turnover rates between metabolites and proteins in the body.[15,16] Furthermore, our results (e.g., as visualized in our web-based tool) suggest that after meals, intermediates of glycolysis, taurocholate cycle, and branched chain amino acid metabolism increased, whereas acylcarnitines, glycerols, and free fatty acids decreased. This could indicate that the body depends more on glucose and amino-acid metabolism for energy production than beta-oxidation after meals. Summarizing, the obtained insights into variability, and the conceptual framework of how to study these fluctuations, can contribute to implement metabolite and protein profiling for clinical decision making. With knowledge of biomolecule variation in "healthy" individuals over *time*, the prediction of possible prevention interventions or early diagnosis of disease onset becomes feasible. Biomolecules such as CMPF, platelet-derived growth factor C, and cathepsin D (all in plasma) with low *noise* potentially detect changing conditions *within* a person. If the *between* component is also low, biomolecules may differentiate conditions *between* persons, for example, cathepsin D, CD27 antigen, and prolylglycine (all in plasma). Interestingly, some of these biomolecules have substantial variability over *time* potentially caused by food intake or circadian rhythm; this variation can be controlled in clinical practice when designating sampling times, as suggested by our method. It should be mentioned that this concept not only applies to single biomolecule biomarkers but also to biomarker "profiles" (i.e., multivariate combinations of biomolecules) for the diagnosis of diseases and patient stratification.

METHODS

This section contains the details underlying the Results and Discussion.

Summary of analytical techniques

A detailed description of all analytical techniques is included as Supplemental Information. In summary, the following procedures were followed:

- *Plasma metabolite analysis at Metabolon*. The metabolomics platform consisted of four technologies: (i) reverse phase optimized for more hydrophilic compounds; (ii) more hydrophobic compounds; (iii) measured in positive ion mode, reverse phase in negative ion mode; (iv) and hydrophilic interaction liquid chromatography separation analyzed with negative ionization. Sample preparation was conducted using methanol extraction to remove proteins. Metabolites were quantified using area under the curve (AUC) on the mass spectrometry chromatogram. Each metabolite was normalized relative to that in pooled plasma prepared from samples in all subjects.
- *Urine amine analysis at Biomedical Metabolomics Facility Leiden*. The amine platform was measured by the Biomedical Metabolomics Facility Leiden and covers amino acids and biogenic amines employing an Accq-tag derivatization strategy adapted from the protocol supplied by Waters (Etten-Leur, The Netherlands). Acquired data were evaluated using MultiQuant Software for Quantitative Analysis (version 3.0.2; AB Sciex, Framingham, MA), by integration of assigned multiple reaction monitoring peaks and normalization using proper internal standards. For analysis of amino acids their "13C15N"-labeled analogues were used. For other amines, the closest eluting internal standard was used. All urine measurements were normalized by their respective sample median to account for dilution.
- *Plasma protein analysis at SomaLogic*. Proteins covering a diverse set of major gene families, including receptors, kinases, growth factors, and hormones, and secreted, intracellular, and extracellular proteins or domains were measured by the SOMAscan proteomic assay (SomaLogic, Boulder, CO[17]). The target protein levels were quantified in relative fluorescent units. The assay was conducted by SomaLogic under their quality system, which comprises facility control and strict standard operating procedures.

Modeling strategy and data analysis

In addition to routine platform preprocessing and normalization, all data must be brought to the same scale for the statistical results to be interpretable in *absolute* terms. To this end, data from each biomolecule are normalized to have mean zero and variance one. No logarithmic transformation was applied.

Let $Y_{it}^{(j)}$ denote the normalized relative concentration of biomolecule j in (1,..., 2,074) from subject i in (1,..., $n = 10$) at occasion t. A linear mixed model[18] has been used to estimate effects of time, corrected for age and BMI, as well as variance components for each biomolecule j.

$$Y_{it}^{(j)} = b_i^{(j)} + \gamma_1^{(j)}\text{age} + \gamma_2^{(j)}\text{BMI} + \beta_t^{(j)} + \varepsilon_{it}^{(j)} \quad (1)$$

Here, $b_i^{(j)} \sim N(0, \boldsymbol{\sigma}_{\boldsymbol{b}}^{\mathbf{2}(j)})$ is a random effect accounting for dependence between measurements of the same individual. Fixed effects $\gamma_1^{(j)}$ and $\gamma_2^{(j)}$ control the relationship for differences in age and BMI. Furthermore, each measurement occasion t is associated with its own fixed effect $\beta_t^{(j)}$, leaving the metabolite or protein profile completely flexible. Depending on the platform, a total T of 9, 4, or 7 time-effects is included (**Table 1**). Finally, the length-T residual vector $\boldsymbol{\varepsilon}_i^{(j)} \sim N(\mathbf{0}, \boldsymbol{\Sigma}^{(j)})$ is assumed to be multivariate normally distributed with $T \times T$ covariance matrix $\boldsymbol{\Sigma}^{(j)}$. This model allows for correlation between errors *within* the same individual. Such flexibility could be important because measurement times are nonequally spaced (**Table 1**). The Supplemental Information elaborates on the statistical methods used for parameter estimation (Eq. 1) and significance testing (**Figure S2**). All statistical analyses were conducted in R version 3.4.3[19] and are available through an open Github repository (https://github.com/KevinDuisters/HumanBaseline). The linear mixed models, including covariance structures, were estimated using the nlme package version 3.1-137.[20]

Definitions

The following parameter estimates based on the statistical model (Eq. 1) underpin the presented results for each biomolecule j.

- *Between* (inter): $\boldsymbol{\sigma}_{\boldsymbol{b}}^{\mathbf{2}(j)}$
- *Time* (intra): $\boldsymbol{\sigma}_{\boldsymbol{\tau}}^{\mathbf{2}(j)} = \frac{1}{T-1}\sum_{t=1}^{T}\left(\beta_t^{(j)} - \frac{1}{T}\sum_{t=1}^{T}\beta_t^{(j)}\right)^2$
- *Noise* (intra): $\boldsymbol{\sigma}_{\boldsymbol{\varepsilon}}^{\mathbf{2}(j)} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\sigma}_{\boldsymbol{t}}^{\mathbf{2}(j)}$

Here, $\boldsymbol{\sigma}_{\boldsymbol{t}}^{\mathbf{2}(j)}$ is the $[t,t]$ diagonal element of the matrix $\boldsymbol{\Sigma}^{(j)}$, and $\beta_t^{(j)}$ are the estimated fixed effects for time. **Figure 2** uses the $\Sigma^{(j)}$ covariance matrix in detail instead of averaging over the diagonal as $\boldsymbol{\sigma}_{\boldsymbol{\varepsilon}}^{\mathbf{2}(j)}$ does. Next to depicting a (new) subject's expected mean profile, a $(1 - \alpha)$ prediction interval $\left[m_t - \frac{n+1}{n}c_\alpha s_t, m_t + \frac{n+1}{n}c_\alpha s_t\right]$ is included. Here, c_α denotes the $(1-\alpha/2)$ quantile in Student's t distribution with $n - 1 = 9$ degrees of freedom. Dropping the (j) superscript for convenience, $m_t = \beta_t + \gamma_1\text{age} + \gamma_2$ BMI is the expected normalized relative biomolecule concentration at occasion t and $s_t = \sqrt{\sigma_b^2 + \sigma_\varepsilon^2}$ its estimated SE. Whenever a heterogeneous Σ structure is selected, σ_t^2 differs over time t and, hence, the pointwise prediction intervals (i.e., the error bars) in **Figure 2** vary in width.

Data access

The authors have developed a Human Baseline Tool, in which profile plots (e.g., **Figure 2**) for each of the 2,074 normalized biomolecules can be visualized interactively. The Human Baseline Tool is illustrated in **Figure S4** and publicly accessible via http://hbt.analyticalbiosciences.nl, including an option to download all data used to generate the results.

SUPPORTING INFORMATION

Supplementary information accompanies this paper on the *Clinical Pharmacology & Therapeutics* website (www.cpt-journal.com).

Figure S1.
Figure S2.
Figure S3.
Figure S4.
Table S1.
Supplemental Information. Study design.
Supplemental Information. Analytical methods.
Supplemental Information. Statistical methods.

ACKNOWLEDGMENT

The authors would like to thank Willem Birkhoff (CHDR, Leiden) for his contributions to the clinical study.

FUNDING

Sample collection and measurement were funded by Astellas Pharma, Takeda Pharmaceutical Company, and Daiichi Sankyo Company.

CONFLICT OF INTEREST

During this project, S.O., Ta.A., Y.H., Y.O., and K.S. were employed by Astellas; To.A., A.A., and H.K. by Takeda; and K.I., N.W., H.Y., and N.K.

by Daiichi Sankyo. All other authors declared no competing interests for this work.

AUTHOR CONTRIBUTIONS

K.D. and T.H. wrote the manuscript. S.O., To.A., K.I., Ta.A., A.H., M.M., Y.H., A.A., Y.O., N.W., H.Y., J.M., H.K., N.K., K.S., and T.H. designed the research. K.D. analyzed the data.

1. Roberts, L.D., Koulman, A. & Griffin, J.L. Towards metabolic biomarkers of insulin resistance and type 2 diabetes: progress from the metabolome. *Lancet Diabetes Endocrinol.* **2**, 65–75 (2014).
2. Ngo, D. *et al*. Aptamer-based proteomic profiling reveals novel candidate biomarkers and pathways in cardiovascular disease. *Circulation* **134**, 270–285 (2016).
3. Geyer, P.E., Kulak, N.A., Pichler, G., Holdt, L.M., Teupser, D. & Mann, M. Plasma proteome profiling to assess human health and disease. *Cell Syst.* **2**, 185–195 (2016).
4. Poste, G. Bring on the biomarkers. *Nature* **469**, 156–157 (2011).
5. Beger, R.D. *et al*. Metabolomics enables precision medicine: "A White Paper Community Perspective." *Metabolomics* **12**, 149 (2016).
6. Van der Greef, J., Hankemeier, T. & McBurney, R.N. Metabolomics-based systems biology and personalized medicine: moving towards n = 1 clinical trials? *Pharmacogenomics* **7**, 1087–1094 (2006).
7. Senn, S. Statistical pitfalls of personalized medicine. *Nature* **563**, 619–621 (2018).
8. Kim, K. *et al*. Mealtime, temporal, and daily variability of the human urinary and plasma metabolomes in a tightly controlled environment. *PLoS One* **9**, e86223 (2014).
9. Maitre, L. *et al*. Assessment of metabolic phenotypic variability in children's urine using 1H NMR spectroscopy. *Sci. Rep.* **7**, 46082 (2017).
10. Floegel, A. *et al*. Reliability of serum metabolite concentrations over a 4-month period using a targeted metabolomic approach. *PLoS One* **6**, e21103 (2011).
11. Nicholson, G. *et al*. Human metabolic profiles are stably controlled by genetic and environmental variation. *Mol. Syst. Biol.* **7**, 525 (2011).
12. Sampson, J. *et al*. Metabolomics in epidemiology: sources of variability in metabolite measurements and implications. *Cancer Epidemiol. Biomarkers Prev.* **22**, 631–640 (2013).
13. Kjelland, C. & Djogovic, D. The role of serum lactate in the acute care setting. *J. Intensive Care Med.* **25**, 286–300 (2010).
14. Manoli, I. & Venditti, C. Disorders of branched chain amino acid metabolism. *Trans. Sci. Rare Dis.* **1**, 91–110 (2016).
15. Niklas, J., Sandig, V. & Heinzle, E. Metabolite channeling and compartmentation in the human cell line AGE1.HN determined by 13C labeling experiments and 13C metabolic flux analysis. *J. Biosci. Bioeng.* **112**, 616–623 (2011).
16. Boisvert, F.M. *et al*. A quantitative spatial proteomics analysis of proteome turnover in human cells. *Mol. Cell Proteomics* **11**, M111.011429 (2012).
17. Gold, L. *et al*. Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS One* **5**, e15004 (2010).
18. McCullogh, C.E., Searle, S.R. & Neuhaus, J.M. *Generalized, Linear, and Mixed Models*, *2nd edn*. (John Wiley & Sons, Hoboken, NJ, 2008).
19. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2017). <https://www.R-project.org>.
20. Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. & R Core Team. nlme: linear and nonlinear mixed effects models <https://CRAN.R-project.org/package=nlme> (2017).