Multi-omic microsampling captures health perturbations in a lifestyle context

Xiaotao Shen[1,2,+], Ryan Kellogg[1,2,+], Daniel J. Panyard[1,2,+], Nasim Bararpour[1,2,+], Kevin Erazo Castillo[1], Brittany Lee-McMullen[1], Alireza Delfarah[1], Jessalyn Ubellecker[1], Sara Ahadi[1], Yael Rosenberg-Hasson[3], Ariel Ganz[1], Kevin Contrepois[1], Basil Michael[1], Ian Simms[1], Chuchu Wang[4], Daniel Hornburg[1], and Michael P. Snyder[1,2,*]

[1] Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA.
[2] Stanford Center for Genomics and Personalized Medicine, Stanford, CA 94305, USA.
[3] Human Immune Monitoring Center, Microbiology and Immunology, Stanford University Medical Center, Stanford, CA 94305, USA.
[4] Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA.

[+] These authors contributed equally.
[*] Corresponding authors: Michael P. Snyder (mpsnyder@stanford.edu).

Abstract

The discovery and profiling of blood biomarkers in a clinical or research setting is challenging due to the cost and inconvenience of in-clinic venipuncture, geographical barriers, limited sampling frequency, and low depth of molecular measurements. In addition, current healthcare practices are reactive and based on limited physiological and clinical information, often collected months or years apart. Here, we present a multi-omic microsampling approach that enables the measurement of thousands of metabolite, lipid, cytokine, and protein molecules in frequently-collected 10 µL blood samples in conjunction with wearable sensors. Using this approach, we conducted two dense sampling case studies to (1) perform dynamic metabolic assessments of the response to a complex mixture of dietary interventions and discover highly individual-level inflammatory and metabolic responses, and (2) perform deep personal profiling to reveal large-scale dynamic molecular fluctuations as well as thousands of molecular associations including many that are associated with intraday variation in physiology (e.g heart rate), clinical markers (glucose, cortisol), and activity. The presented methodology achieves fully remote, scalable, high-temporal-resolution omics and sensor monitoring and has the potential for comprehensive biomarker discovery and dynamic health profiling.

Introduction

Multi-omic technologies enable the quantification of thousands of molecules and can provide new insights into the molecular landscape of health and disease[1,2]. Despite major advances in omics technologies, the upstream sample collection and processing still requires travel to a clinic, access to a phlebotomist, and physical and emotional discomfort. These current sample collection strategies do not meet the desired flexibility and noninvasiveness to conduct comprehensive longitudinal profiling independent of access to a clinic. Furthermore, the high sample volume needed (often 10-50 mL of venous blood) prohibits frequent collections, which precludes high-resolution analysis of dynamic metabolic and biological processes that occur on the scale of minutes or hours. Finally, the high sample collection and processing costs can be prohibitive for performing large studies in remote environments.

Previous studies have investigated dried blood spot (DBS) sampling[3–6] and volumetric absorptive microsampling (VAMS)[7,8,9] for metabolite and protein analyses[10]. In principle, DBS allows individuals to

collect a blood drop sample at home and return the sample by mail at room temperature. However, DBS sampling is often irreproducible since volumetric amounts can vary considerably, and, to date, the number of analytes analyzed from DBS has generally been modest[11].

To circumvent these challenges, we devised a streamlined multi-omics profiling system that uses fingerprick blood drop collection, minimizes pain, and enables sampling frequencies on the timescale of minutes without needing clinic access. Our method collects fixed 10 µL volumes and, following extraction, enables the simultaneous analysis of proteins, metabolites, lipids, and targeted cytokines/hormones from a single sample enabling broad analyte profiling. In two proof-of-principle studies, we first demonstrate the profiling of a dynamic response to ingestion of a mixed meal shake and discover high heterogeneity in individual metabolic and immune responses, and second, we perform high-resolution profiling of an individual for over one week enabling the identification and quantification of thousands of molecular changes and associations across omes at a personal level. Our platform is scalable, enabling high-frequency molecular profiling for broad utility in research and clinical studies.

Results

Overview of the multi-omic microsampling approach. The blood microsampling and multi-omics data acquisition workflow we devised are shown in **Figure 1a**. After testing numerous devices, we settled on collecting 10 µL blood microsamples using a Mitra device, a solid matrix that collects fixed blood volumes. We tested a wide variety of extraction conditions and further developed a method for efficiently extracting proteins, a broad range of lipids, and metabolites from a single microsample using biphasic extraction with Methyl tert-Butyl Ether (MTBE). This extraction procedure yields an organic phase containing hydrophobic metabolites and lipids, an aqueous phase containing hydrophilic metabolites, and methanol precipitated protein pellet processed for proteomics data acquisition. Using a separate microsample, we performed an aqueous extraction for performing multiplexed immunoassays on the Luminex platform (see **Methods**). All the omics data sets are then processed, annotated, and curated for detailed omics analysis.

To evaluate the microsampling method, we first examined the stability of proteins, metabolites, and lipids in microsamples under multiple conditions, including testing storage duration and temperature (**Figure 1b**, **Figure S1a**). We then compared microsampling to conventional intravenous sampling methods (**Figure 1b**). Finally, two pilot case studies were performed to demonstrate how microsampling can capture important health and biological perturbations in a lifestyle context (**Figure 1b**).

Protein, metabolite, and lipid stability in microsamples in multiple conditions. A study was first designed to evaluate the stability of proteins, metabolites, and lipids in blood microsamples (**Figure S1a**). In brief, two participants were recruited, and blood samples were collected using the 10-µL Mitra devices. A total of 36 microsamples were generated for each participant, with the microsamples stored in duplicate at 3 temperatures (4, 25, and 37°C) and for 5 durations at the given temperature (3, 6, 24, 72, and 120 hours) before being stored at -80°C until analysis. An additional set of samples was immediately stored at -80°C. Proteomics, metabolomics, and lipidomics data were acquired from the microsamples (see **Methods**). After quality control, imputation, and annotation of the data, there were 66 proteomics samples with 128 proteins, 71 metabolomics samples with 1,461 annotated features, and 72 lipidomics samples with 776 lipids (**Supplementary Data 1**). Each omics data set was assessed individually to examine analyte stability concerning storage duration, storage temperature, and the interaction of storage duration and temperature. The metrics of stability assessed were (1) the average coefficient of variation (CV) across both participants' samples (estimated using the formula for log-scale data[12]), (2) the presence of significant effects of storage

conditions on analyte level in a linear regression analysis (excluding the baseline samples that were not stored at any temperature), and (3) relative importance measures (partial R^2 and the Lindemann, Merenda and Gold measure, LMG1, see **Methods**).

For proteins, the median CV was 0.397 with a range of 0.149-1.728 (**Figure 1c**). The regression analysis yielded three proteins (2.3%) nominally associated with the duration of storage, eight proteins (6.3%) nominally associated with temperature, and six proteins (4.7%) nominally associated with the interaction effect (**Figure 1d**). In each case, the number of observed associations was near or less than that expected by chance at the alpha level of 0.05, indicating weak evidence for an effect of storage conditions on the protein levels. The overall model R^2 similarly showed that the models captured little (mean = 0.149) of the observed variance in protein levels. Of the storage condition covariates, the effects of storage duration, temperature, and interaction were all low (mean partial R^2 = 0.014, 0.022, 0.019, respectively; **Figure S1b**).

For the metabolite features, the median CV was 0.378 with a range of 0.054-54.328 (**Figure 1c**). Unlike the proteins, the metabolite features showed more significant effects of storage conditions, with all three conditions showing enrichment of significant associations above what would be expected by chance: 194 (13.3%) of the 1,461 features were associated with duration, 389 (26.6%) features were associated with temperature, and 193 (13.2%) features were associated with the interaction effect (**Figure 1d**). The overall model R^2 values captured more variation (mean = 0.327) than for the proteins. The mean partial R^2 values for duration, temperature, and the interaction effect showed a higher R^2 for temperature compared to the other terms (**Figure S1b**, mean partial R^2 = 0.030, 0.062, and 0.030, respectively).

For the lipids, the median CV was 0.335 with a range of 0.088-2.218 (**Figure 1c**). There were 150 lipids (19.3%) significantly associated with storage duration (p-value < 0.05), 513 lipids (66.1%) associated with storage temperature, and 173 lipids (22.3%) associated with the interaction term. The overall model R^2 values for lipids were the highest of the three classes (mean = 0.547). The mean partial R^2 values for duration, temperature, and the interaction effect were 0.043, 0.181, and 0.044, respectively, which similarly highlighted temperature as having a greater effect on the lipid levels as it did for the metabolites **(Figure S1b**).

In summary, there was little evidence for an effect of storage duration and temperature on proteins. There was an effect of all three storage effects on metabolites and lipids (especially temperature). Overall, the majority of all analytes were not significantly associated with storage duration, and the majority of proteins and metabolites were not associated with storage temperature.

Comparison between microsample and intravenous plasma sample. We next examined the similarity between the molecular profiles derived from microsamples of whole blood compared to venipuncture plasma. Blood samples were collected from 34 participants using both microsampling and conventional intravenous blood sampling approaches (**Figure S2a,** see **Methods**), and metabolomics and lipidomics data were acquired from each participant (**Supplementary Data 2**). The median intensity of every feature in the 34 participants was calculated separately in two datasets, microsampling and intravenous plasma collection samples, and compared via correlation graphs (**Figure 1e**). Interestingly, the results of the two collection types (microsampling and intravenous collection) were quite similar in that the Spearman correlations were 0.81 (p-value < 0.001) and 0.94 (p-value < 0.001) for 642 metabolites and 616 lipids, respectively. Metabolites and lipids that were not correlated well (Spearman correlation < 0.5) were enriched for amino acids and triglycerides (TAGs), respectively (**Figure S2b,c**). However, most classes of molecules were very similar between the microsampling and venous blood draw, including most of the amino acids,

carbohydrates, free fatty acids (FFAs), TAGs, diglycerides (DAGs), phosphatidylcholines (PCs), and other molecules.

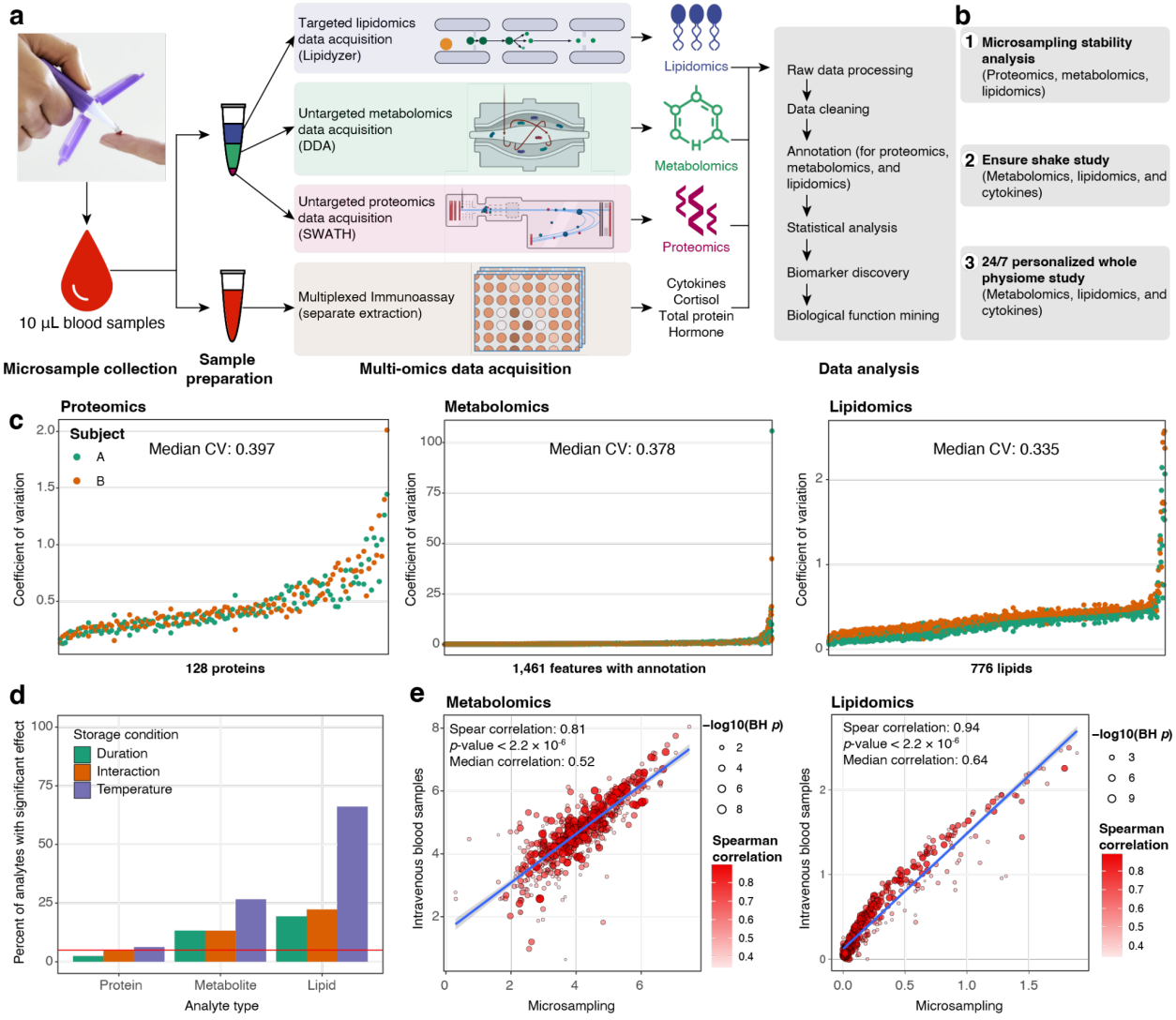

Figure 1. Overview of the microsampling multi-omics workflow and stability analysis. a, The samples were collected using microsampling devices, and then multi-omics data (proteomics, metabolomics, lipidomics, cytokine, *etc.*) were acquired. **b**, Outline of the primary microsampling analyses. **c**, The coefficient of variation distribution for proteins, metabolites, and lipids across all the samples in the stability analysis. **d**, The percentage of analytes is significantly affected by storage duration, temperature, and interactions. The red line shows the expected proportion of nominally significant results at the alpha level of 5% (p-value= 0.05). **e**, The Spearman correlations between microsamples and intravenous blood samples for metabolites and lipids, respectively. The icons used in this figure are from iconfont.cn.

Case studies. As a demonstration of the power of microsampling, we performed two case studies while participants were in their native environments. The first was to examine the effect of drinking a complex mixture on metabolic profile. The second was to perform very densely "24/7" profiling (98 microsamples) across periods of just greater than seven days.

Case Study #1: Metabolic phenotyping responses to Ensure shake consumption. Individuals can differ markedly in their metabolic response to food based on their epigenomic, microbiome, metabolomic, and other factors[13,14,15,16], yet the heterogeneity of this response is not well understood. Determining these differences at an individual level is important in order to optimize diet and lifestyle changes for overall health, weight reduction, and/or management of metabolic disease. Biomarkers are typically measured at a single time point because of the difficulty of collecting high-frequency blood samples using a conventional blood sampling approach, but the rapid and dynamic nature of metabolism in response to food intake requires higher resolution. To follow the diversity of metabolic responses to complex dietary mixtures, we measured the multi-omic responses to a defined mix of carbohydrates, lipids, proteins, and micronutrients. We analyzed metabolomics, lipidomics, cytokines, and hormones in 28 participants with diverse backgrounds (**Figure 2a, Figure S3a**) and developed six metabolic responses metrics: (1) carbohydrate, (2) lipid, (3) amino acid (protein), (4) insulin secretion, (5) free fatty acid (related to insulin sensitivity), and (6) immune (cytokines).

Thirty-two participants were mailed a kit containing microsampling Mitra devices, an Ensure shake, and instructions for microsampling sample collection. Each participant collected one microsample (defined as 0 min), consumed the Ensure shake, and collected additional blood microsamples at 30, 60, 120, and 240 minutes after consumption (**Figure 2a**). Participants returned their microsamples by overnight mail on the same day of microsample collection. The microsamples were used for multi-omics data acquisition, namely metabolomics, lipidomics, and cytokines/hormones. Four subjects without metabolomics data were removed from the dataset (see **Methods**, **Figure 2b**). After data cleaning, curation, and annotation, 768 analytes were detected from the microsamples, including 560 metabolites, 155 lipids, and 54 cytokines/hormones for each of the 28 participants at each of the five time points (a total of 140 data points) (**Figure 2b**) (**Supplementary Data 3**).

Clustering of altered molecules. We first determined if the microsampled multi-omics data reflected the consumption of the Ensure shake. For each time point post-consumption, the Wilcox rank test was utilized to define the significantly dysregulated molecules compared to time point 0 (baseline). Interestingly, the majority of significantly increased metabolites and lipids peaked at approximately 60 and 120 min, respectively, and then recovered to some extent in each case. Seven cytokines/hormones peaked at around 30 min and then recovered, indicative of a rapid immune reaction (**Figure 2c**). These results indicate that many molecules significantly responded to Ensure shake, and the response kinetics differed based on the classes of molecules. To quantify the molecules that shifted their levels upon Ensure shake consumption, an analysis of variance (ANOVA) test was utilized. The results show that the levels of 99 of 560 metabolites (17.7%, permutation test p-value < 0.001), 115 of 155 lipids (74.2%, permutation test p-value < 0.001), and 7 of 54 cytokines/hormones (13.0%, permutation test p-value < 0.001) significantly shifted following the Ensure shake consumption (**Supplementary Data 4,** see **Methods**). For the metabolites whose levels changed, the signals of analytes that differed from baseline were greater than those affected by storage duration. These results demonstrate that multi-omics analysis from microsamples can be used to measure the metabolic response to Ensure shake.

The molecules significantly affected by Ensure shake were then clustered using fuzzy c-means clustering to reveal and summarize the pattern of changes associated with consumption time (**Figure 2d,** see **Methods**). The shifted molecules were grouped into 3 major clusters across 5-time points (**Figure 2d**). Cluster 1 contained 39 metabolites, one lipid, and 4 cytokines that increased and then decreased with a peak at approximately 60 minutes following the Ensure shake consumption. Cluster 2 contained 19 metabolites

and 106 lipids that increased more gradually than cluster 1, peaking at approximately 60-120 minutes. Molecules in cluster 3 decreased after consuming the Ensure shake and then recovered, including 23 metabolites, 8 lipids, and 3 cytokines (**Figure 2d**). These results show that the molecules have very different response patterns following the Ensure shake consumption.

Altered metabolic pathway and physiological responses to Ensure shake. We next explored the pathways and physiological responses represented by the molecules in each cluster (**Figure 2d** and **Figure S3**). Cluster 1 was primarily comprised of metabolites (39 metabolites, one lipid, and 4 cytokines). Metabolic functional annotation defined several biological pathways such as aminoacyl-tRNA biosynthesis, phenylalanine, tyrosine and tryptophan biosynthesis, and phenylalanine metabolism pathways (**Figure S3c**). The two major chemical classes captured in cluster 1 were amino acids (*e.g.* alloisoleucine, alanine, isoleucine, methionine, norvaline, phenylalanine, tryptophan, and tyrosine) and carbohydrates (*e.g.* fructose, lactic acid, and pyruvic acid). Both compound classes likely come directly from the Ensure shake or are metabolized quickly and return to baseline levels by 240 min (**Figures 2e,f**). On the other hand, for cluster 3, acetylcarnitine was the main metabolite class (nine acetylcarnitine: decanoylcarnitine, decenoylcarnitine, dodecanoylcarnitine, tetradecadiencarnitine, L-octanoylcarnitine, decanoyl-L-carnitine, hexanoyl-L-carnitine, lauroyl-L-carnitine, and octanoylcarnitine) which dramatically decreased upon Ensure shake consumption and then recovered gradually by 240 min (**Figure 2g**). This is expected because acetylcarnitine is known to be broken down in the blood by plasma esterases to carnitine, and carnitine helps free fatty acids to be transported into the mitochondria for beta-oxidation and energy production and hence maintaining whole-body energy homeostasis[17]. Consistent with this interpretation, eight free fatty acids (FFAs) detected in cluster 3 (**Figure S3d**) decreased following the Ensure shake consumption. Notably, in cluster 2, we found 106 lipid species (**Figure 2d**), and most of them were triglycerides (102 TAGs with 48-52 carbons chains and 1-3 unsaturations, **Figure S3e**).

To better understand the molecules in the Ensure shake that might be directly detected in the participants' microsamples we also analyzed the composition of the Ensure shake using the same MS procedure. Nearly 50% of the compounds found in the Ensure shake can be detected in the blood, and most of the remainder were of low abundance (**Figure S3f**). Importantly, of 21 high-interest metabolites that changed in the blood (**Figure 2e,f,g**), 17 are present in the Ensure shake. This result demonstrates that the microsampling approach is able to detect the ingested molecular signatures from blood samples.

It is well known that both connecting-peptide (C-peptide) and insulin are co-secreted from the pancreas and correlate with increased carbohydrates[18,19,20]. As expected, the C-peptide and insulin were in the same cluster with carbohydrates (cluster 1, **Figure 2h**). Moreover, we found both gastric inhibitory polypeptide (GIP) and pancreatic polypeptide (PP) in the same cluster with insulin (cluster 1) (**Figure 2h, left panel)**. GIP is an inhibiting hormone of the secretin family of hormones[21], and its main role is stimulating insulin secretion[22]. Increased secretion of PP is reported to be associated with protein meal consumption, fasting, exercise, and acute hypoglycemia[23]. In cluster 3, we found leptin, interferon-gamma (IFNG), and interleukin 4 (IL4) decreased quickly following Ensure shake consumption (**Figure 2h, right panel**). The primary function of leptin is regulating adipose tissue mass through central hypothalamus-mediated effects on hunger[24]; its levels are expected to decrease after food consumption. IFNG and IL4 contribute to generating and regulating immune responses, including allergies and antibacterial responses. Interestingly, this suggests that the Ensure shake may have anti-inflammatory properties. In summary, these results demonstrate that the kinetics of the biochemical responses, including hormones, to complex mixture ingestion can be revealed using microsampling (**Supplementary Data 5**).

Metabolic phenotyping reveals unique individual responses. How people respond to different foods is an area of great interest. The Ensure shake is a simple yet complex mixture of many types of simple molecules that can be quickly absorbed by the small intestine. To examine how different people respond to different metabolites, we explored the diversity in the kinetics and magnitude of the molecular responses among the different participants. Analysis of the samples using a t-distributed stochastic neighbor embedding (tSNE) plot shows that the samples were clustered by the participant, indicating that each participant had a unique molecular profile and that the difference between participants was greater than that of the effect of the shake (**Figure S4a**). Nonetheless, a clear timewise separation of data points was observed (**Figure S4b**). Our study suggested that by 240 min, the metabolic levels tend to return closer to their baseline level (**Figure S4b**). We then used unsupervised consensus clustering to cluster participants into different groups. Our result suggested that there were two major groups based on the molecules altered in response to the shake consumption (**Figure S4c**, see **Methods**). In those two groups, we calculated the level of changes in metabolic features, comparing each time point to the baseline (time point 0) for each participant (see **Methods**). This result also suggested that participants of those two groups had different responses to the Ensure shake (**Figure S4d**): Group 2 responded more slowly than the participants in Group 1 (**Figure S4d**), indicating the kinetics of their responses were different. Interestingly, for the thirteen individuals with a measure of insulin resistance (steady state plasma glucose (SSPG); see **Methods**), although statistically insignificant, we noticed a trend for IR patients to be included in Group 1 over Group 2 (Wilcoxon test: p-value = 0.29, **Figure S4e**).

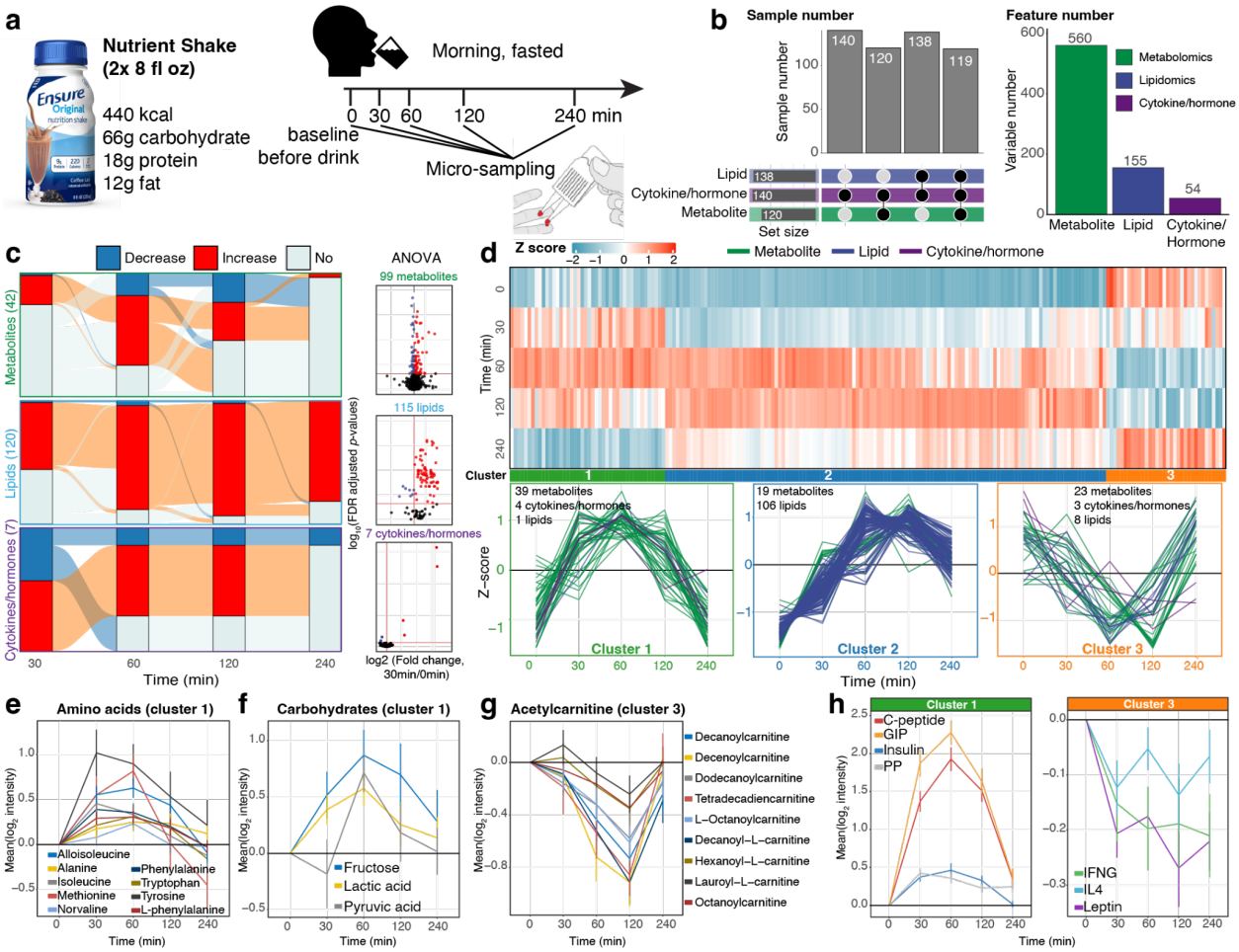

Figure 2. The overview of Ensure shake study and molecular response to ensure shake. a, The study design and overview of the Ensure shake study. **b**, The summary of multi-omics data from microsample. **c**, Responses of metabolites, lipids, and cytokines/hormones after the Ensure shake consumption. **d**, The clustering of dysregulated molecules following Ensure shake consumption. **e,** Amino acids responding to Ensure shake consumption. **f**, Three dysregulated carbohydrates responding to the Ensure shake consumption. **g**, Acylcarnitines responding to the Ensure shake consumption. **h**, Cytokine/hormone responding to Ensure shake consumption. The points are represented by mean ± SD. C-peptide: Connecting peptide; GIP: gastric inhibitory polypeptide; PP: pancreatic polypeptide; IFNG: interferon-gamma; IL4: interleukin 4. The icons used in this figure are from iconfont.cn.

Metabolic phenotyping based on the multi-omics response to the Ensure shake. Since individuals are known to vary in their response to different foods, and we found heterogeneity in response to the Ensure shake for each participant, we next examined the response of each class of molecules, carbohydrates, lipids, cytokines/hormones, and proteins to shake ingestion.

We derived a "metabolic score" for the degree of an individual's carbohydrate, lipid, FFA and protein response to the Ensure shake, along with insulin secretion and inflammatory response (cytokines) (see **Methods**). Briefly, for each molecule in each participant, after the Ensure shake consumption, the area under the curve (AUC) was used to represent its cumulative value (**Figure 3a**). The AUCs of molecules for each molecular class (lipids, carbohydrates, amino acids, inflammatory molecules) were then used to calculate the response score for each participant (**Figure 3b**). The final metabolic scores were normalized and ranged from 0 to 1, where 0 means the lowest relative metabolic level and 1 means the highest relative

metabolic level. One participant was recognized as an outlier subject and excluded during the score calculation (**Figure S5,** see **Methods**). For each participant, we observed a consistent distribution pattern of the molecular species within each metabolic score indicative of similar response patterns to Ensure shake consumption. However, those patterns differed greatly across subjects demonstrating inter-individual variability in the metabolism of nutrients (**Figure S6, 7, 8, 9, 10,** and **Figure S7a**).

The six metabolic scores were calculated for each participant (**Supplementary Data 6**). As expected, we found a negative correlation between FFA score and steady-state plasma glucose level (SSPG), a marker of insulin resistance[25] (**Figure S7b**). Previous studies have demonstrated that elevated plasma levels of FFA are associated with insulin resistance[26]. The participants were classified into five groups based on their metabolic scores using the hierarchical clustering method (**Figure 3c**). We found that individuals varied significantly in their response to the shake for each of the different areas; examples selected from each of the five groups are shown in **Figure 3d**. Within each group, we observed variations in the scores from the average score per metabolic class. For example, participant S30 in group 1 presented lower metabolic scores for fats and amino acids compared to the average level of the entire group. In comparison, S34 in group 5 showed higher scores for those classes (*i.e.* carbohydrates and amino acids) than the average scores. These differences may be due to a variety of underlying mechanisms, including levels of digestive enzymes, transporters, hormones (*e.g.* incretins), and/or intestinal microbes required to process particular molecules in the Ensure shake. Such underlying causes can be investigated in the future through additional analyses (such as metabolic flux analysis). Interestingly, S29 and S35 in groups 3 and 5, respectively, had higher scores in hormones and cytokines. The latter is particularly interesting as some individuals appear to have a strong inflammatory response, whereas others have a different response to appetite-suppressing hormones. Thus, the multi-omics data from microsamples reveal the enormous heterogeneity in the biochemical responses of each individual to a complex mixture. Such information can be defined using microsampling and is important for precision nutrition diets, including inflammatory responses to food.

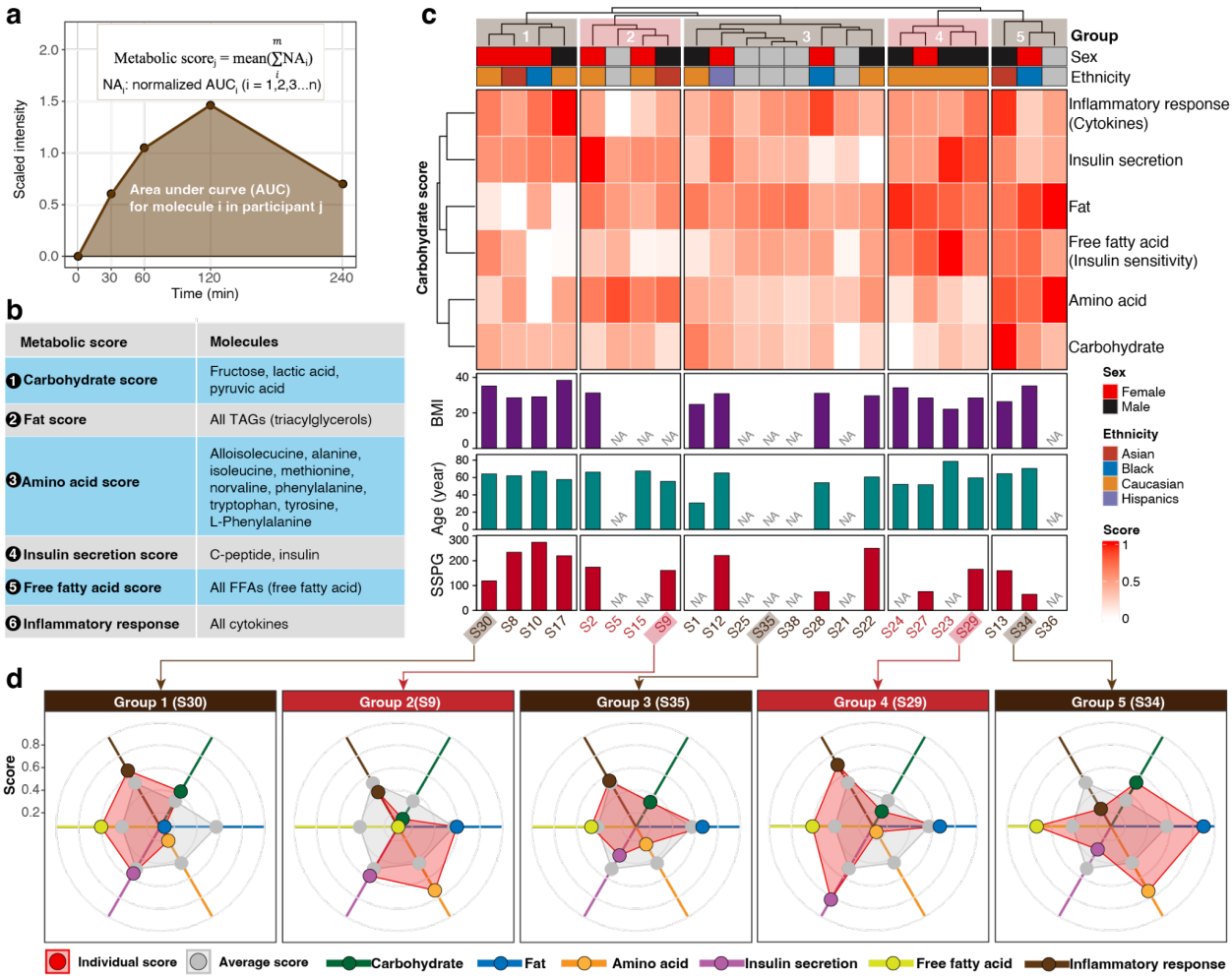

Figure 3. Metabolic phenotyping based on the multi-omics response to the Ensure shake. a, The visualization of the AUC metric for each analyte used in a metabolic score. **b**, The analytes used in calculating each of the six metabolic scores are shown. **c,** Participants were grouped into 5 groups based on 6 metabolic scores. **d,** five participant examples for each group.

Case Study #2: 24/7 personalized whole physiome profiling by wearable and multi-omics data. Longitudinal measurements of internal multi-omics profiles from blood and digital data using wearable devices are expected to provide a more comprehensive understanding of human health and disease. Several studies have demonstrated that longitudinal individualized molecular profiles, clinical tests, and digital data can monitor health and enable early disease detection at an individual level [27,28,29]. However, limitations still exist. For example, most studies use low-frequency/high volume blood sampling (*e.g.* weekly or monthly), which does not enable the detection of circadian metabolic changes. Consequently, integrating those data with wearable data is biased by missing information at the biochemical levels, thus hindering the study of causal physiological relationships. Knowing those limitations and pitfalls, we performed a detailed personalized profiling study to explore if our microsampling method can capture relationships between multi-omic molecular measurements and wearable data at the individual level. Higher frequency data collection would enable monitoring of health status at high resolution in real-time, uncover relationships between molecules with each other and physiological and lifestyle activities, and decipher causal associations between them at the personal level.

As a proof of principle we performed a study called the 24/7 study, during which a single participant collected blood microsamples usually every one to two hours during waking hours for over seven days, with some samplings as short as 30 minutes apart (**Figure 4a** and **Figure S8a**). The collection of 98 samples over seven days is one of the most densely sampled studies ever performed. This dense sampling enabled the collection of high resolution omics data that could be compared and integrated with physiological data to reveal circadian patterns (**Figure S8b**). The participant performed food logging and wore two wearable devices: (1) a smartwatch that recorded heart rate (HR) and step count, and (2) a continuous glucose monitor (CGM)[30] (**Figure 4a**, **Supplementary Data 7**). The microsamples were collected while the participant performed their native activities and immediately stored samples on portable dry ice after collection.

A total of 98 microsamples were collected and used to perform in-depth multi-omics profiling, including untargeted proteomics, untargeted metabolomics, targeted lipidomics, and targeted cytokine, hormone, total protein, and cortisol assays (**Figure 4b, upper panel)**. After data acquisition and annotation, we detected a total of 2,213 analytes that included 1,051 metabolites, 811 lipids, 291 proteins, 45 cytokines, 13 metabolic panels (cytokines/hormones), 1 total protein, and 1 cortisol measurement (**Figure 4b, lower panel)** resulting in a total of 214,661 biochemical measurements in addition to wearable physiological data (**Supplementary Data 8**). Overall, the prospective collection of internal molecular and wearable data resulted in comprehensive, high-frequency, and abundant longitudinal data on the human whole physiome and lifestyle (**Figure 4b** and **c, Figure S8b, c**), allowing us to explore how the internal molecules and physiology change on an hourly-scale and their relationships at a personal level.

To explore if multi-omic microsampling captures real biological signatures (such as food intake), we selected two days that the participants ate high carbohydrate food (131.8 g) and low carbohydrate food (31.9 g), respectively (**Figure S9a**). Then the two carbohydrate metabolites (fructose, and pyruvic acid) in microsamples were extracted and analyzed, as shown in the boxplot in **Figure S9b**. The median values of Carbohydrate metabolites are 7.8 and 4.7, respectively. This result demonstrates that the omics data from microsamples roughly reflect the concentration of the food the participants consumed.

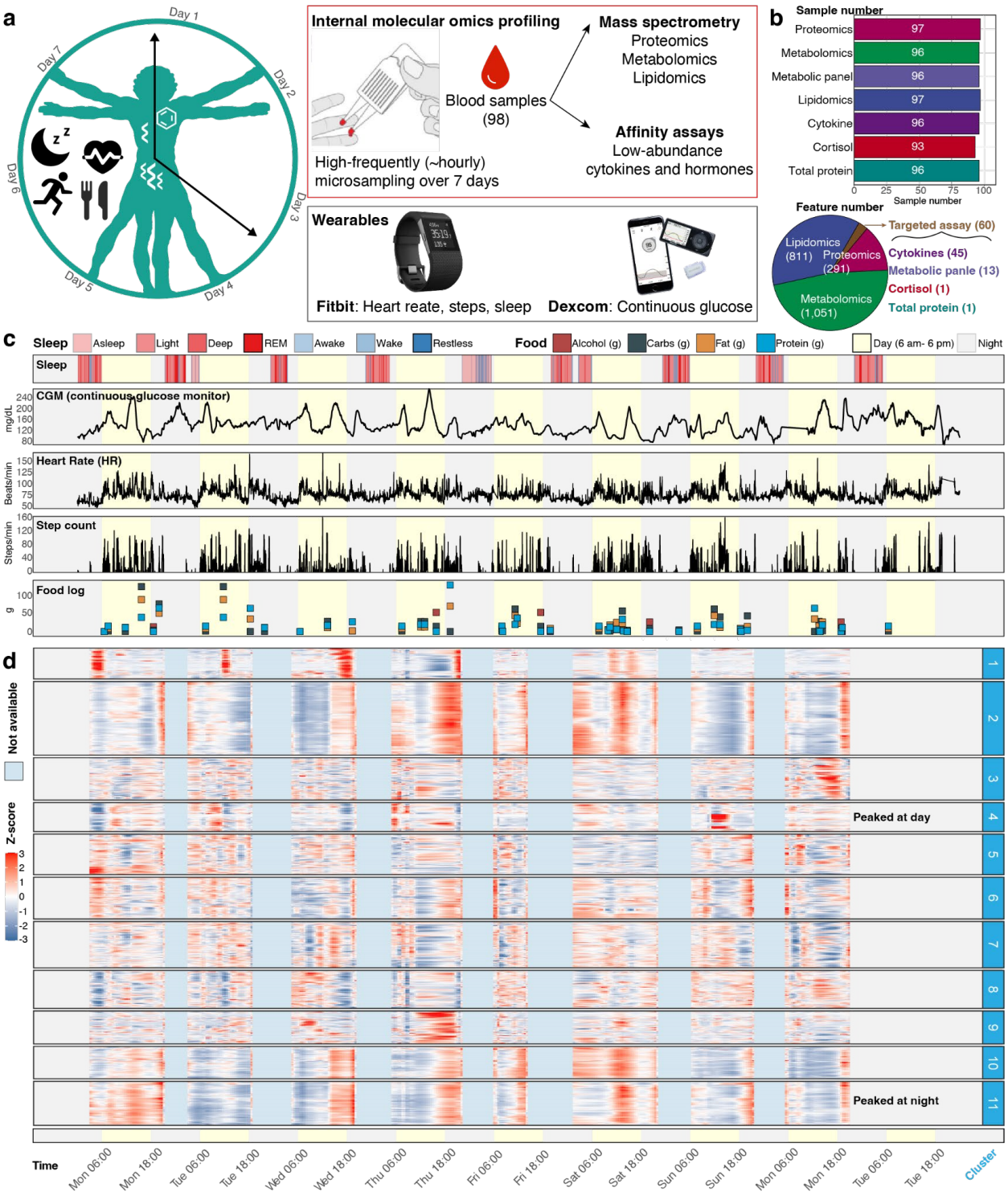

Figure 4. Overview of the study design, sample collection, and data acquisition for the 24/7 project. **a**, One participant was closely monitored using wearable devices and high-frequency microsampling (~ hourly) across 7 days. Microsamples were then analyzed for internal multi-omics data measurements. **b**, Molecular information was detected from the high-frequency microsamples. **c**, Wearable data from the smartwatch (sleep and step count) and Dexcom (CGM glucose). Legend defines the status of sleep, the category of consumed foods, and the day/night period at every record. The yellow background represents

the day (6:00 AM - 6:00 PM). **d**, The internal molecules were grouped into 11 clusters using fuzzy c-means clustering. The icons used in this figure are from iconfont.cn.

Wearable and internal multi-omics data reflect the individual physiological status. We first explored whether wearable and high-frequency internal multi-omics data can monitor and reflect the participant's health status. The 2,213 internal molecular profiles were smoothed (see **Methods, Figure S9c,d**) and then grouped into 11 clusters using fuzzy c-means clustering analysis. We found that two clusters followed circadian patterns. For example, cluster 4, which is enriched by a high number of metabolites (**Figure S10a),** generally peaked at day time, while cluster 11, which mostly includes lipids (**Figure S10a)** peaked primarily at night (**Figure 4d**). Other clusters were not necessarily tied to circadian patterns and thus may reflect other events. The components of the different clusters were unique, indicating that the molecules have different temporal patterns (**Figure S10b**). To get more tight and distinct molecular modules, the community analysis method[31] was utilized to extract modules from each cluster (see **Methods, Figure S10c, d, e**). Interestingly, obvious peaks were found in some modules (**Figure S10e,** see **Methods**), indicating that the molecules in modules may be triggered by specific events (**Figure 4d** and **Figure 5a**).

As we have the detailed food (nutrition) and exercise logs, we next analyzed if and how molecular fluctuations relate to daily nutrition intake[32,33] (see **Methods**). Briefly, nutrients in the food log were classified into several major classes based on their content level: amino acids, vitamins, fat, electrolytes, calories, carbs, and fiber. Next, we calculated the association between those classes with internal molecules presented by the Jaccard index depicted in the heatmap (**Figure S10g**). Interestingly, we captured a high association between classes of amino acids and fat with several modules highly enriched in amino acids, free fatty acids, and lipids (**Figure S10g**). This result is consistent with a previous study[34]. Our data also suggested the potential usage of internal molecular data from microsamples reflective of daily nutrition intake. For example, the participant consumed the same meal shake every morning during the study, and we captured a clear link between daily shake consumption and temporal increase of several compounds such as 1,2,3-benzenetriol sulfate and hydroxyphenyllactic acid, which are listed as the shake's ingredients (**Figure 5a, upper panel**).

Cortisol is believed to follow a circadian pattern, with levels higher in the morning that decrease towards the evening[35]. However, events during the day related to stress, activity, and diet can impact cortisol levels[36]. Although peak levels of cortisol were evident on three days, we observed large day-to-day variations in cortisol patterns, demonstrating that within-day cortisol levels may not represent accurate inter-day cortisol patterns for this individual (**Figure 5a**). This result suggests the importance of high-frequency sampling for monitoring health status.

Importantly, this study demonstrates the potential usage of microsamples to measure the pharmacokinetics of a drug at an individual level. Our participant took a low dose of aspirin in the morning for four days. Microsampling accurately captured the pharmacokinetics of salicylic acid (hydrolyzed product of aspirin) and revealed a clearance period of about 24 hours in this person, which is similar to previous studies[37] (**Figure 5a, lower panel**). In addition, we found a negative correlation between caffeine and sleep quality (**Figure S11a,b**). This might be expected and has been reported in other studies[38,39]; however, the participant always consumed coffee before noon, indicating its long-lasting effect.

Interestingly, our detailed monitoring also revealed an unidentified inflammatory event in the middle of the week, spanning 3 days, with a subset of inflammatory cytokines with apparent rising (TNF α, CD40L, *etc.*) and decreasing (Eotaxin, *etc.*), respectively (**Figure S11c,d**). This event was subclinical, as no symptoms were reported, but it may represent an asymptomatic infection or other stress event.

Together, these results show the power of high-frequency monitoring to record daily measures and events not evident to the patient.

Circadian rhythms of internal molecules in human blood. Circadian rhythms are endogenous oscillators in physiological and behavioral processes over a 24-hour cycle, and they play a critical role in human health and diseases[40]. The circadian molecules in the body participate in physiological phenomena such as cell division, energy metabolism, and blood pressure[36,41]. The circadian analysis of molecules requires high-frequency (~hourly) biological time-series data. However, it is difficult to collect high-frequency time-series data over multiple days in a real-life setting because of the limitations of the traditional blood sampling approach. Using the high-frequency data collected from the microsampling method, we were able to explore and evaluate molecules associated with circadian rhythms in the human body[42].

Each molecule was first searched for those that exhibited a consistent pattern across all 7 days and removed those that lacked a consistent daily pattern (see **Methods**, **Figure S12a**). The circadian rhythms analysis (JTK_CYCLE algorithm[43]) was then utilized for quantitative analysis of all the molecules (see **Methods**). We identified 332 circadian molecules (Benjamini–Hochberg adjusted p-values < 0.05) that show clear circadian patterns (**Figure S12b**, **Supplementary Data 9**). The circadian molecules were grouped into five major clusters using fuzzy c-means clustering (**Figure S12c**) to identify the major patterns. Interestingly, all clusters, except cluster 4 (enriched by protein), were dominated by lipids (**Figure S12d**). We focused on the molecules that exhibited a complete one-day cycle (those in clusters 1, 2, and 3; **Figure 5b,c**) and removed clusters 4 and 5, whose molecules had different levels at the beginning and end of the day (**Figure S12e,f**). Cluster 1 was dominated by Phosphatidylcholine (PC, 32.56%) and Lysophosphatidylcholine (LPC, 25.58%), cluster 2 was dominated by Triacylglycerols (TAG, 93.65%), and cluster 3 was dominated by both TAGs (49.15%) and Phosphatidylethanolamine (PE, 22.03%) (**Figure 5d**). Examples for each cluster are shown in **Figure 5e**.

To explore the in-depth functions of the rhythmic molecules in each cluster, we performed lipid enrichment analysis using Lipid Mini-on[44]. LPC, PC, Sterol, and Cholesterol ester (CE) were significantly enriched in cluster 1. Previous work has shown that LPC and PC have circadian rhythms with peak concentrations in the evening, consistent with our result[45]. For cluster 2, TG and Glycerolipid were significantly enriched, and for cluster 3, PE was significantly enriched (**Figure 5f**). Thus, the different classes of lipids exhibit distinct circadian patterns. To explore if the circadian lipids were affected by the food intake, we then examined the food logging data. We found that the fat nutrition intakes differed across eight days, meaning that the circadian lipids are not driven by the food intake. It is plausible that circadian lipids were driven by the individual rhymic kinetics or gut microbes, which can be investigated using additional experiments. In summary, multi-omics analyses from the high-frequency microsamples revealed rhythmic molecules and demonstrated that lipids related to energy metabolism have distinct circadian patterns.

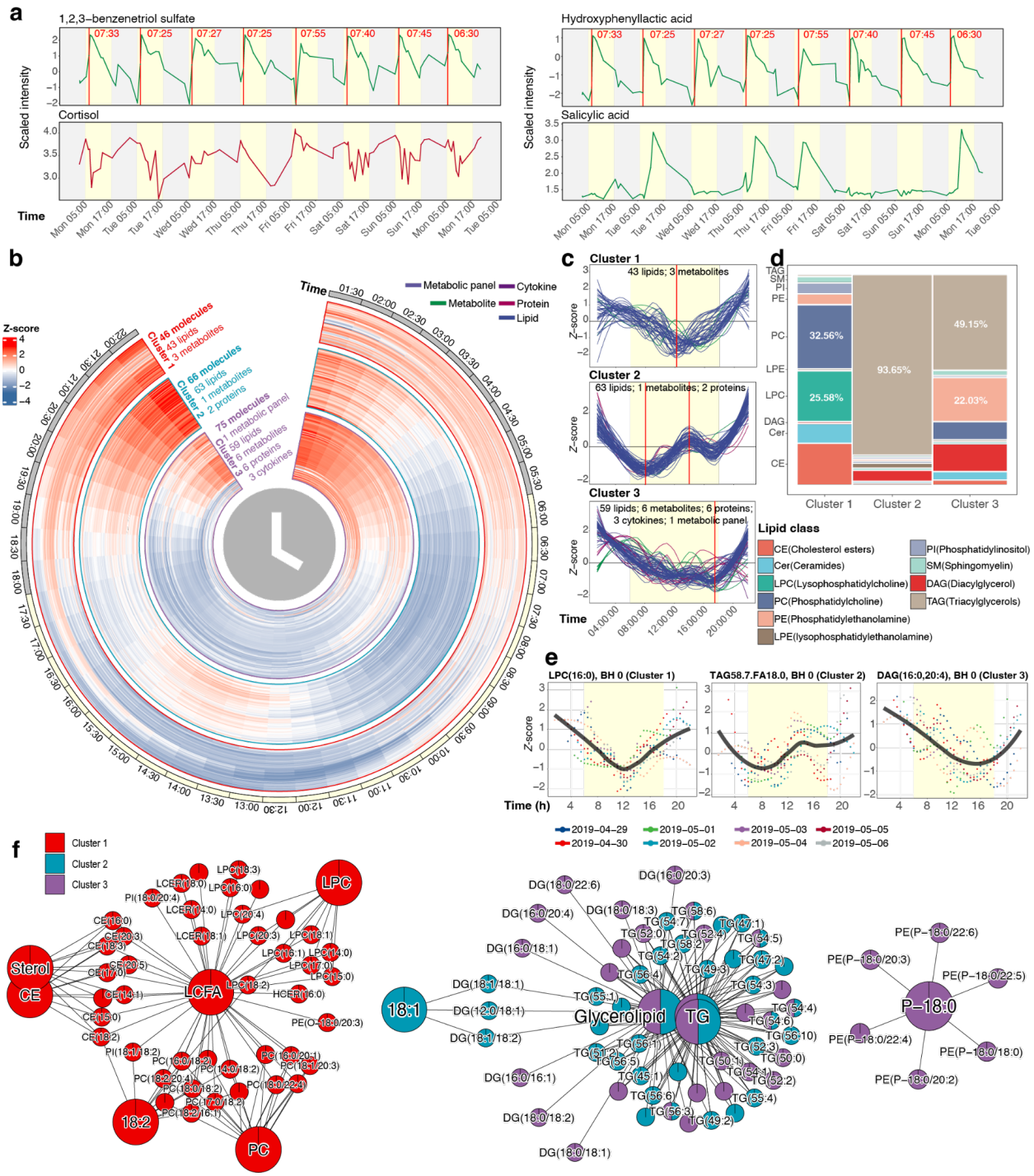

Figure 5. Wearable and internal multi-omics data reflect the individual physiological status and circadian rhythm analysis of multi-omics data. a, Four molecules reflect the participant's lifestyle. **b, c,** Three clusters that have strong rhythmic patterns. **d**, Lipid class distributions of lipids in three clusters. **e**, Examples for each cluster. **f,** Lipid enrichment results for lipids in clusters 1-3. The yellow background represents the day (6:00 AM - 6:00 PM).

The wearable data reflect internal molecular changes. Over the past several years, longitudinal monitoring of physiological data has garnered considerable interest[30,46,47,48,49]. However, the ability of

wearable data to predict clinical labs has been limited[30]. Several studies have demonstrated that wearable data can reflect and predict the internal molecules (multi-omics data), including laboratory clinic tests and metabolites on a weekly or monthly scale[29,30]. However, due to the low-frequency sampling of multi-omics data, the circadian patterns and causal relationships between digital and internal molecular data cannot be discerned[49]. We explored the relationship between wearable data and internal molecular changes on an hourly scale at an individual level, including building predictive models.

Because of the different sampling frequencies of wearable and internal multi-omics data, we first attempted to match the wearable and internal multi-omics data using different window sizes (**Figure S8**). The matching windows were set as 5, 10, 20, 30, 40, 50, 60, 90, and 120 minutes. For each wearable data type (Heart rate, step count, and CGM) in the matched windows, a feature engineering pipeline[30] was utilized to convert different data types into eight features (for example, the standard deviation of heart standard, maximum heart rate; see **Methods**) resulting in a total of 24 wearable features. The 24 wearable features were utilized to predict each analyte using the Random Forest model. Of the 2,223 molecules, we found 447 molecules with at least one $R^2 > 0.3$ (**Figure S13a**). Interestingly, we also found that most molecules have higher prediction accuracy with the larger matching window, consistent with a previous study[30]. Most of the 447 molecules were lipids, and enrichment analysis showed that TGs were the most predictable by wearable data (**Figure S13b**). Heart rate-related features (*e.g.* HR range, HR maximum, and HR SD) contributed the most to the predictive models (**Figure S13c**). Using the Random Forest model, we then found that several cytokines (C-peptide, GIP, insulin, and PP) could also be predicted by wearable data. The most contributed wearable features were CGM and heart rate-related features (**Figure S13d**). All those results demonstrate that wearable data could predict our high-frequency multi-omics data from microsamples.

Biochemical changes in the body can occur on the order of minutes and hours[50], and thus low-frequency multi-omics data (weekly or monthly) can find the associations with physiological measurements but not causal relationships[29,30]. Using the high-frequency microsampling approach, we next explored if we could deduce the potential causal relationships between wearable data and internal molecules through temporal relationships; causal events are expected to precede downstream effects[51]. For each wearable data and molecule, we first matched them with different lagged times. Then, the Spearman correlation and *p*-value were calculated for each matching time-series data. Only the correlations with similar shapes and lagged time were scored as significant lagged correlations (see **Methods**). To enable this analysis, the *laggedCor* (lagged correlation) algorithm was developed as an R package (https://jaspershen.github.io/laggedcor/). We then utilized it to demonstrate that this algorithm could capture \the known lagged correlation and causal relationships between step count and heart rate. Interestingly, we found a lagged correlation of 0.6 (BH adjusted p-value < 0.0001) with a shift time of -1 min (step count - heart rate, **Figure S14a**), which means that one minute after the step count increases, the heart rate begins to increase. This expected result demonstrates that our lagged correlation algorithm can capture and quantify potential causal relationships. Next, a lagged correlation network between wearable and internal molecular data was generated (**Figure S14b, Figure 6a**), including 1,217 nodes (3 wearable data and 1,214 molecules) and 1,895 edges (**Figure 6b, Supplementary Data 10**), demonstrating that the high degree of association between wearable and multi-omics data. An example with the top 100 edges for each pair of wearable and omics data is provided in **Figure 6a**. Step count and heart rate have most of the edges (57.3% and 42.6%, respectively) in the lagged correlation network (**Figure 6b**). We also found that CGM correlates more with cytokines than heart rate and step count (**Figure S14c**), indicating that glucose levels strongly correlate with immune responses. This result has been demonstrated by other studies[52]. In

addition, we also observed that step count and heart rate have many (669) overlapping correlations (**Figure S14d**), as expected since they have a significant positive correlation (**Figure S14a**).

Interestingly, the immunity-related pathways contained some proteins that negatively correlated with CGM, which was not expected (**Figure S14e**). This demonstrates the importance of following these responses at the individual level. As expected, we also found that glucagon signaling, oxidative phosphorylation pathways, and free fatty acids positively correlate with CGM (**Figure S14f,g**). Glucagon breakdown can raise the concentration of glucose and fatty acids in the bloodstream, and oxidative phosphorylation can oxidize nutrients to release chemical energy. We found that the caffeine metabolism pathway positively correlates with heart rate (**Figure S14h**), consistent with previous studies[53]. We also found that the blood coagulation pathway positively correlates with heart rate (**Figure S14i**), and the neutrophil degranulation pathway negatively correlates with heart rate (**Figure S14j**). To the best of our knowledge, these associations provide new biological insights that should be validated in future studies. Overview, these results demonstrate that the wearable data can reflect the physiological status of the participant and reveal useful insights at a personal level.

We extracted the CGM glucose subnetwork from the entire lagged correlation network to further explore how glucose associates with internal molecules (**Figure 6c**). We observed that CGM glucose has a significant lagged correlation with α-Synuclein (lagged correlation: 0.36, BH-adjusted p-value < 0.05) (**Figure 6d**), and the shift time is -55 min (α-Synuclein - CGM), indicating that α-Synuclein may directly or indirectly upregulate glucose levels in the blood. This result has been demonstrated by previous studies[54,55]. Previous studies have shown that higher C-peptide levels correlate with increased CGM glucose values[56]. Our data found that CGM glucose significantly lagged correlations with C-peptide (**Figure 6d** and **Figure S15a**) and insulin (**Figure S15b**). The shift time between CGM and C-peptide in this individual is 10 minutes (lagged correlation 0.36, BH-adjusted p-values < 0.05), which means that CGM glucose precedes the concentration of C-peptide in blood by 10'. We also observed that CGM significantly correlates with several cytokines, including TNF-β (**Figure 6d**), FLT3L, IL15, IP10, and TNF-α (**Figure S15c;** time shift 0 to 15 min), and four of them are proinflammatory cytokines. These results indicate that glucose can cause a rapid proinflammatory response. In summary, our results show that based on the high-frequency multi-omics data from microsamples, we find potential causal associations between wearable and multi-omics data. The potential causal relationships we found using the *laggedcor* algorithm can be validated by future experiments.

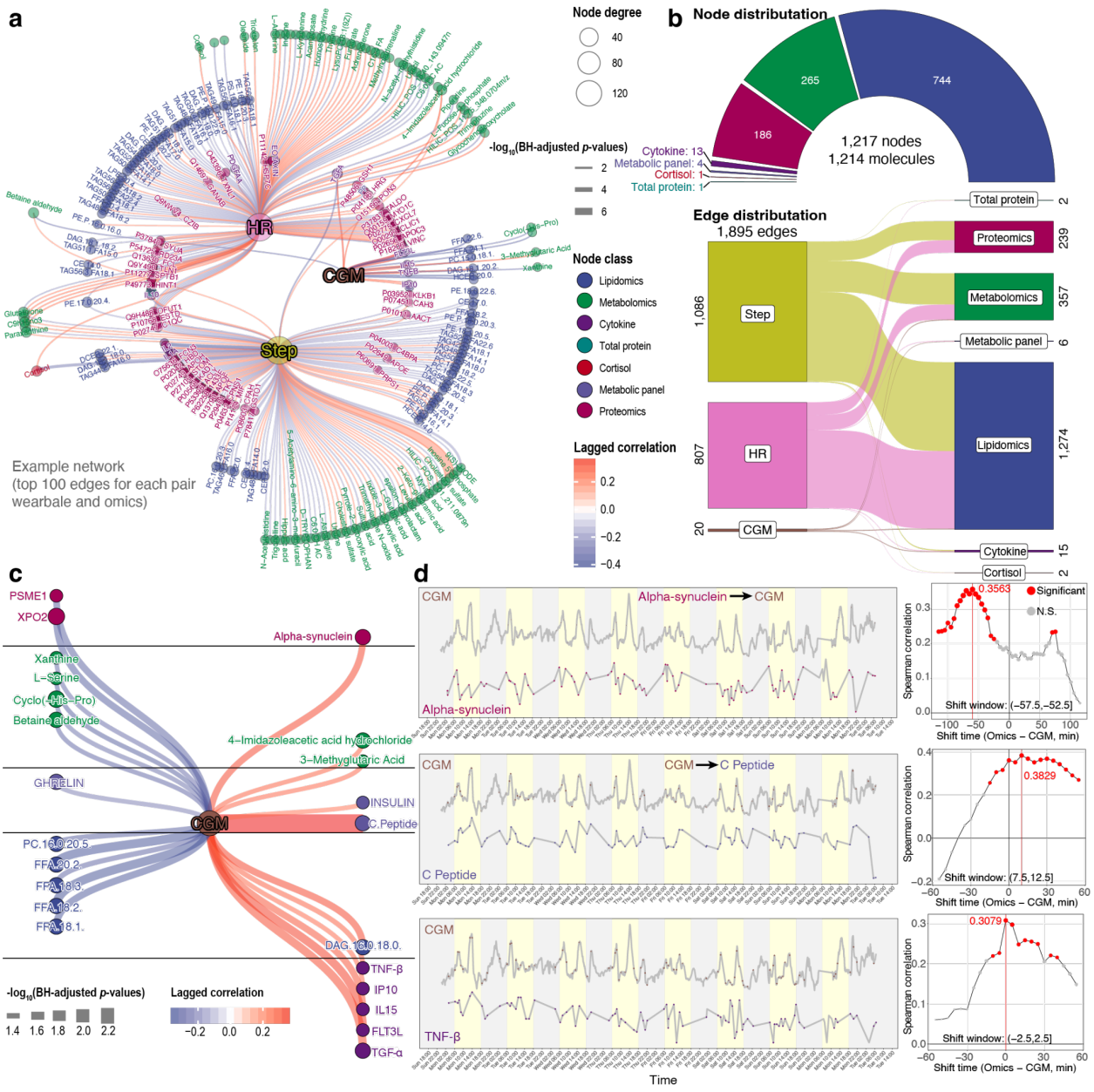

Figure 6. Wearable data and internal molecule causal association network. a, Example association network between wearable data and internal molecules. **b**, Node and edge distribution of association network. **c**, The CGM glucose subnetwork from the whole network. **d**, Three examples are shown to represent the causal relationships between CGM glucose and internal molecules.

Discussion

Current healthcare practices are reactive and based on limited physiological and/or clinical information, often collected months or years apart. Here, we built a multi-omic microsampling approach that enables the measurement of thousands of metabolite, lipid, cytokine, and protein molecules in frequently-collected 10 µL blood samples. We demonstrated that many of the molecules from the microsamples (VAMS) are stable and reliable. In addition, most of the molecules from the microsampling are consistent with the classic blood sampling approach (Spearman correlation: 0.8-0.9). Compared to DBS,

VAMS can achieve good analytical performance for targeted compound and protein analysis[57,58]. However, for the DBS, the hematocrit effect affects the resulting spot size, which can introduce variation in analysis. In addition, DBS requires a drying period before the sample can be sealed and shipped, which may also introduce variations in the analysis compared to VAMS[59]. Since the microsampling approach is less invasive and can be used remotely and without specific training, it enables high-frequency blood sample collection (~ hourly) in a native setting, which is difficult to perform using the classic blood sampling approach. Based on the multi-omic microsampling workflow, we carried out two case studies to demonstrate the dense in situ samplings, and analytic capabilities to (1) perform dynamic and individualized metabolic assessments after response to dietary (Ensure shake) intervention and (2) reveal large-scale intraday molecular fluctuations as well as thousands of molecular associations including those associated with intraday variation in heart rate, glucose levels and activity.

For the first case study, multi-omics profiling after Ensure shake consumption revealed that 46.6% of detected molecules were significantly altered after consumption, demonstrating extensive biochemical changes. The altered molecules are grouped into 3 clusters with distinct patterns across time points, demonstrating that the molecules have different patterns and kinetics of the biochemical responses to complex mixture ingestion. In addition, utilizing the multi-omics data from microsamples, six metabolic scores, namely carbohydrate score, fat score, amino acid (protein) score, insulin secretion score, free fatty acid score (insulin sensitivity), and immune response score (cytokines), were calculated to represent people's individual quantitative response to the complex mixture (**Figure 3c**). Interestingly, some individuals exhibited a rapid inflammation response, whereas others exhibited a fast insulin response. Correlating these individual responses with medical phenotypes (*e.g.* LDL levels, HbA1c levels) will be important for personalized nutrition management nutrition in the future.

The combination of molecular phenotypes and quantitative wearable lifestyle measurements to monitor human health in real time has increased[49]. In the second 24/7 case study, a total of 98 microsamples over seven days were collected from a single participant and used to perform in-depth multi-omics personal profiling. A total of 2,213 analytes (metabolites, lipids, proteins, cytokines, hormones, total protein, and cortisol) were measured, resulting in a total of 214,661 biochemical data points in addition to wearable physiological data (from a smartwatch, CGM, and food logging; **Figure 4**). Overall, the prospective collection of internal molecular and wearable data resulted in comprehensive, high-frequency, and abundant longitudinal data on the human whole physiome and lifestyle, which is a precious resource for personal profiling and putative utility in predicting responses and diseases. We found that the multi-omics data from microsamples could accurately reflect nutrition intake on an hourly scale. In particular, two specific metabolites (1,2,3-benzenetriol sulfate, and Hydroxyphenyllactic acid) have a high temporal association with daily shake consumption. We can also measure the external drug kinetics (Aspirin) at an individual level (**Figure 5a**). Finally, we detected a non-symptomatic 3-day inflammation event. Together, these results show the power of high-frequency monitoring to record daily measures as well as subclinical events. The latter is particularly important for the early detection of disease[60].

Circadian rhythms are endogenous oscillators in physiological and behavioral and play a critical role in human health and diseases. The circadian analysis of molecules requires high-frequency biological time-series data. Here, utilizing the high-frequency multi-omics data collected from the microsampling method, we explored and evaluated molecules associated with circadian rhythms in the human body and found that lipids related to energy metabolism have distinct circadian patterns. We also demonstrated that wearable data could predict our high-frequency multi-omics data from microsamples on an hourly scale. In addition, taking advantage of the high-frequency samples from the microsampling approach, we developed

a novel algorithm, *laggedCor* (lagged correlation), which enables the ascertainment of the potential causal relationships between wearable and internal multi-omics data. We found α-Synuclein levels precede the glucose levels in the blood, which was previously demonstrated[54,55]. We also found that CGM glucose can increase the concentration of C-peptide in blood, as expected, but we captured the exact timing at a personal level for each of these associations. Finally, microsampling analysis reveals actionable information. For example, the negative correlation of caffeine levels with sleep, even though consumption was halted before noon, suggests further restricting that consumption to earlier in the day might be valuable to improve sleep for this individual. Together, based on the high-frequency multi-omics data from microsamples, we could explore the potential causal associations revealed using wearable data and multi-omics data at an individual level.

Finally, it is worth noting that many analytes we measure, particularly proteins, have some degree of stability with regard to time, temperature, and the combination of both. We also note too that since we tested for significant effects of storage conditions with a relatively low sample size, we do not rule out additional effects that may not have been observed here due to power challenges, which was evident from a sensitivity regression analysis that analyzed only one storage condition at a time (storage duration and storage temperature) and additional effects for storage duration were identified when the baseline samples were added to the analysis. For those molecules that are not stable, they can either be discarded from the analyses or quantification can be ascertained from unique degradation products. Alternatively, sample collection procedures could include rapid and cold shipping to minimize potential issues with less stable molecules. Indeed, we expect most samples can be collected and stored within 24 hours, thus minimizing degradation. Larger stability studies, especially in larger and more diverse populations, will help identify other potential issues. Regardless, reliable measurements can be made for thousands of molecules, including those present at very low abundance (*e.g.* cytokines).

In summary, the presented methodology achieves fully remote, scalable, high temporal resolution omics and sensor monitoring. It has the potential for large-scale comprehensive, dynamic molecular and digital biomarker discovery and monitoring as well as health profiling. Here, we used two case studies to show the potential of multi-omic microsampling in precision medicine. Many other applications can be envisioned. Examples include (1) Longitudinal biomarker discovery. The multi-omics microsampling is simple and unpainful compared to the traditional blood collection method and thus enables anyone to self-collect high-frequent and high-quality blood microsamples anywhere for longitudinal biomarker discovery. (2) Personalized health monitoring. The people can collect blood samples at home without any help and then send the samples to the laboratory for data acquisition and analysis. If a significant abnormality is detected, the result is sent immediately to a physician. The physician would then be able to validate the results and respond quickly with an intervention. (3) Therapeutic drug monitoring. The patients could collect microsamples frequently and remotely to monitor the drug-related compounds or biomarkers in the blood at a known time, to guide dosage, and result in optimized therapy. In our study, all the microsamples were prepared and run together as a batch to avoid batch effects. In the future, the microsamples collected in one day could be prepared and run in one day after sample collection. The users can receive their results within two days after sending their samples to the laboratory for analysis. Additionally, developing a clinical diagnostic based on microsampling requires additional validation steps for accuracy, precision, matrix effects, *etc,* and the use of standards such as isotopically labeled reference molecules. In addition, presently, only proteins, metabolites, lipids, and cytokines were measured using our microsampling approach, but other types of molecules can be measured, including, such as DNA, epigenomes, and RNA. For the 24/7 study, as a pilot study, only one participant was recruited to demonstrate the power of following

personalized responses. Enlargement of the sample size will enable the measurement of more generalized patterns but will also reveal new challenges in the processing and analysis of large numbers of samples. Indeed, our simple studies generated 98 data points in a single individual.

Here, the two pilot case studies (group study and individual study) were utilized to demonstrate the power and application of the approach. The molecular signatures found in our study provide vast testable hypotheses that should be validated using analytical and experimental approaches. We note that group analysis is usually performed to find the overall trend. However, it can be potentially used to identify individual outliers who may have underlying conditions[1]. When an individual profile differs greatly from the average, one needs to first check for sample mixups, systematic variation and batch effects. Then on normalized, data outlier detection can be further performed. Individuals who fall outside the overall pattern can be investigated for underlying causes for their molecular shift (medical conditions, medications, or lifestyle abnormalities). In addition, the confounders (e.g. sex, age, BMI, etc.) must be controlled and adjusted to find the real and expected biological variation. Similarly, we note that when an individual profile differs greatly from the average, overview conclusions from the whole cohort may not extend to individuals[33,61]. For the personalized analysis, the conclusion from the individual may not be extended to specific individuals[61], which can be revealed using our approach. Overall, we believe the multi-omic microsampling approach offers a promising opportunity to integrate with wearable data to improve precision healthcare.

Methods

Sample collection

Microsampling blood samples. The Mitra device (Neoteryx, CA, USA) is used to collect the microsampling blood samples. The blood microsampling method and multi-omics data acquisition workflow were established first (**Figure 1a**). We developed a method for extracting proteins, lipids, and metabolites from single microsamples, using biphasic extraction using methyl tert-butyl ether (MTBE). This extraction yields an organic phase processed for lipids, an aqueous phase processed for metabolites, and a protein pellet processed for proteomics. Using a separate microsample, we performed an aqueous extraction for performing multiplexed immunoassays on the Luminex platform. (**Figure 1a**).

Intravenous blood samples. Intravenous blood from the upper forearm was drawn from overnight-fasted participants. Specimens were immediately placed on ice after collection to avoid sample deterioration. Blood was collected in a purple top tube vacutainer (BD, NJ, UCA), layered onto Ficoll media (Thermo Fisher Scientific, USA), and spun at 2,000 r.p.m. for 25 min at 24 ℃. The top layer EDTA-plasma was pipetted off, aliquoted, and immediately frozen at 80 ℃. The peripheral blood mononuclear cells (PBMC) layer was collected and counted via the cell counter, and aliquots of PBMCs were further pelleted and flash-frozen.

Sample preparation

Microsampling blood samples. Mitra tip samples were thawed on ice, prepared, and analyzed randomly. Briefly, 300 µL of methanol spiked in with internal standards (provided with the Lipidyzer platform) was added to a Mitra tip and vortexed for 20 s. Lipids were solubilized by adding 1,000 µL of MTBE and incubated under agitation for 30 min at 4 °C. Phase separation was induced by the addition of 250 µL of ice-cold water. Samples were vortexed for 1 min and centrifuged at 14,000 g for 5 min at 20 °C. The upper phase containing the lipids was then collected, dried down under nitrogen, reconstituted with 200 µL of

methanol, and stored at -20 °C. After biphasic extraction, the Mitra tips were resuspended in 0.1 M Tris pH 8.6 buffer, along with 10% N-octyl-glucoside and 50 mM tris(2-carboxyethyl)phosphine (TCEP), followed by shaking at 60 ℃ for 1 hour (denaturation, solubilization, and reduction). The protein mixture was subsequently alkylated with 200 mM Indole-3-acetic acid (IAA) and incubated at room temperature (24 ℃) in the dark for 30 minutes. Proteins were digested with trypsin overnight at 37 ℃ and quenched the following day with 10% (v/v) formic acid the following day. 300 µL of metabolite layer was transferred and then added 1200 uL ice-cold MeOH: Acetone: ACN (1:1:1), then vortexed for 10 s. And the sample was incubated overnight at -20 ℃. The samples were vortexed for 10 s, and followed a centrifuge of 20,000 g for 10 min at 4°C. Then the sample was transferred to a new 2.0 mL tube and dried down. Finally, the samples were stored at -20 °C until data acquisition.

Intravenous blood samples. The sample preparation of venous blood samples for omics data acquisition is documented by Schüssler-Fiorenza Rose *et al.*[1], Zhou *et al.*[2], and Gao *et al.*[29].

Untargeted proteomics

Data acquisition. Approximately 8 µg of tryptic digest were separated on a NanoLC 425 System (Sciex, Redwood City, CA, USA). 5 µL/min flow was used with trap-elute setting using a ChromXP C18 trap column 0.5 × 10 mm, 5 µm, 120 Å (cat# 5028898, Sciex, Redwood City, CA, USA). Tryptic peptides were eluted from a ChromXP C18 column 0.3 × 150 mm, 3 µm, 120 Å (cat# 5022436, Sciex, Redwood City, CA, USA) using a 43-minute gradient from 4%–32% B with 1-hour total run. Mobile phase solvents consisted of 92.9% water, 2% acetonitrile, 5% dimethyl sulfoxide, 0.1% formic acid (A) and 92.9% acetonitrile, 2% water, 5% dimethyl sulfoxide, 0.1% formic acid (B). MS analysis was performed using Sequential Window Acquisition of all Theoretical (SWATH) acquisitions on a TripleTOF 6600 System equipped with a DuoSpray Source and 25 mm I.D. electrode (Sciex, Redwood City, CA, USA). Variable Q1 window SWATH Acquisition methods (100 windows) were built-in high sensitivity MS/MS mode with Analyst TF Software (v1.7).

Data processing. The spectra were analyzed with OpenSWATH using an in-house spectral library made from plasma and peripheral blood mononuclear cell (PBMC) samples. Peak groups were then statistically scored with the PyProphet tool (v2.0.1), and all runs were aligned using the TRIC strategy. A final data matrix was produced with 1% FDR at the peptide level and 5% FDR at the protein level. Several quality control steps were then applied to the output from SWATH2STATS. The correlation of peptide intensities between samples was calculated, and 2 samples with a mean sample correlation less than 2 SDs from the mean sample correlation were removed. An additional sample with a peptide count less than 3 SDs below the mean was removed. Poorly identified proteins and peptides were removed according to their m-scores using a target FDR of 0.05 (m-score threshold = 8.91×10^{-12}). Peptides matched to an unknown protein, non-proteotypic peptides, and peptides beyond the 10 most intense peptides for a given protein were all removed. Protein intensities were then calculated by first summing the intensities of all transitions mapped to each peptide and then all peptides mapped to each protein. Proteins that were missing for > 50% of samples were removed, as were proteins whose CV among a separate set of 3 QC samples was greater than 50%. Each missing protein value was imputed using k-nearest neighbors (k = 10; using only non-imputed data; R package VIM, version 6.1.0). Protein values were then $\log_2$-transformed.

Untargeted metabolomics

Data acquisition. Prepared samples were analyzed four times using HILIC and RPLC separation in both positive and negative ionization modes, respectively. Data were acquired on a Q Exactive Plus mass spectrometer for HILIC and a Q Exactive mass spectrometer for RPLC (Thermo Scientific, San Jose, CA, USA). Both instruments were equipped with a HESI-II probe and operated in full MS scan mode. MS/MS data were acquired on quality control samples (QC) consisting of an equimolar mixture of all samples in the study. HILIC experiments were performed using a ZIC-HILIC column 2.1 × 100 mm, 3.5 µm, 200 Å (cat# 1504470001, Millipore, Burlington, MA, USA) and mobile phase solvents consisting of 10 mM ammonium acetate in 50/50 acetonitrile/water (A) and 10 mM ammonium acetate in 95/5 acetonitrile/water (B). RPLC experiments were performed using a Zorbax SBaq column 2.1 × 50 mm, 1.7 µm, 100 Å (cat# 827700-914, Agilent Technologies, Santa Clara, CA, USA) and mobile phase solvents consisting of 0.06% acetic acid in water (A) and 0.06% acetic acid in methanol (B).

Data processing. Data from each mode were independently analyzed using Progenesis QI software (v2.3) (Nonlinear Dynamics, Durham, NC). Metabolic features from blanks that didn't show sufficient linearity upon dilution in QC samples ($r < 0.6$) were discarded. To reduce metabolic features of the metabolome profile, only metabolic features present in > 2/3 of the samples were kept for further analysis. Next, in the study samples, metabolic features present in > 50% of those samples were kept for further analysis. Missing values were imputed using k-nearest neighbors (KNN) with $k = 10$. Data was then $\log_2$-transformed. The batch effect was evaluated using the dbnorm package[62]. Applying several batch removal algorithms, the *ComBat* model[63], giving the best performance, was considered for correcting systematic variation associated with the batch. Data from each mode were independently analyzed using Progenesis QI software (v2.3, Nonlinear Dynamics). ComBat was used to do data normalization[62], and KNN was used for missing value imputation. Data from each mode were merged, and metabolites were formally identified by matching fragmentation spectra and retention time to analytical-grade standards when possible or by matching experimental MS/MS to fragmentation spectra in publicly available databases using metID[64]. We used the Metabolomics Standards Initiative (MSI)[65] level of confidence to grade metabolite annotation confidence (level 1 - level 2).

Semi-targeted lipidomics

Data acquisition. Prepared samples were analyzed using the Lipidyzer platform that comprises a 5500 QTRAP System equipped with a SelexION differential mobility spectrometry (DMS) interface (Sciex, Redwood City, CA, USA) and a high flow LC-30AD solvent delivery unit (Shimadzu, Kyoto, Japan). The detailed method can be found in our previous study[66]. In brief, lipid molecular species were identified and quantified using multiple reaction monitoring (MRM) and positive/negative ionization switching. Two acquisition methods were employed, covering 10 lipid classes; method 1 had SelexION voltages turned on, while method 2 had SelexION voltages turned off. Lipidyzer data were reported by the Lipidomics Workflow Manager (LWM) software which calculates concentrations for each detected lipid as the average intensity of the analyte MRM/average intensity of the most structurally similar internal standard (IS) MRM multiplied by its concentration.

Data processing. The final datasets were generated from the Lipidyzer platform, and the lipid abundances were reported as concentrations in nmol/g. Lipids detected in less than 2/3 of the samples were discarded, and missing values were imputed based on a lipid class-wise KNN-TN imputation method[67].

Cytokines and metabolic panel. Cytokines were analyzed using the HCYTMAG-60K-PX41 kit or the HSTCMAG28SPMX13 kit. For Metabolic hormone assays, the catalog number was HMHEMAG-34K. These assays were performed by the Human Immune Monitoring Center at Stanford University. All kits were purchased from EMD Millipore Corporation (Burlington, MA, USA) and used according to the manufacturer's instructions with the following modifications. Briefly, samples were mixed with antibody-linked magnetic beads on a 96-well plate and incubated overnight at 4 °C with shaking. Cold (4 °C) and room temperature incubation steps were performed on an orbital shaker at 500-600 r.p.m. Plates were washed twice with wash buffer in a Biotek ELx405 washer. Following one hour of incubation at room temperature with a biotinylated detection antibody, streptavidin-PE was added for 30 minutes with shaking. Plates were washed as described, and Phosphate buffered saline (PBS) was added to wells for reading in the Luminex FlexMap3D Instrument (Thermo Fisher Scientific, USA) with a lower bound of 50 beads per sample per cytokine. Each sample was measured in a singlet. Custom Assay Chex control beads were purchased from Radix BioSolutions, Georgetown, Texas, and added to all wells.

Cortisol. This assay was performed by the Human Immune Monitoring Center at Stanford University using the ProcartaPlex™ Simplex Kit (Catalog number: EPX010-12190-901, ThermoFisher, Santa Clara, California, USA) and used according to the manufacturer's instructions with modifications as described. Briefly: Beads were added to a 96-well plate and washed in a BioTek ELx405 washer. Samples were added to the plate containing the mixed antibody-linked beads, and 20 µL of the competitive conjugate was added and incubated overnight at 4 °C with shaking. Cold (4 °C) and Room temperature incubation steps were performed on an orbital shaker at 500-600 r.p.m. Following the overnight plate was washed as described, and PE was added for 30 minutes at room temperature. The plate was washed as above, and a reading buffer was added to the wells. Each sample was measured in a single well. Plates were read using a Luminex FM3D FlexMap instrument with a lower bound of 50 beads per sample per cytokine. Custom Assay Chex control beads (Radix BioSolutions, Georgetown, Texas) were added to all wells.

Total protein. The total protein was determined by BCA assay according to kit instructions (ThermoFisher, USA).

Wearable data. The smartwatch (Fitbit) was utilized to collect the sleep, heart rate (HR), and step count data. The Fitbit Intraday API through the My Personal Health Dashboard app[68] was used to retrieve sleep, HR, and step count data for the experiment period. The Dexcom G5 device was utilized to collect the continuous glucose monitoring (CGM) data. CGM data were transferred directly from the G5 device[50]. Dietary intake was logged manually using a notebook to track approximate meal timing and composition.

Stability analysis

Study design. All the microsamples were stored at -80 °C before they were prepared and analyzed. So the stability analysis was designed to explore if the molecules from the microsamples are stable in different storage conditions (temperature and duration time) before they are stored at -80 °C. Two individuals were enrolled under the institutional review board (IRB)-approved protocol (IRB-23602 at Stanford University) with written consent. By venipuncture, two individuals were asked to provide 10 mL of whole blood (in an EDTA purple top tube). The whole blood of each participant was poured into separate plastic reservoirs. Then 10 µL Mitra devices were touched to the surface of the blood to fill the microsample sponge. Thirty-six microsamples were generated for each participant, and microsamples were stored in duplicate at 3

temperatures (4, 25, and 37 °C) for 6 durations at the given temperature (3, 6, 24, 72, 120, and 0 hours [i.e., put into cold storage immediately]) before being stored at -80 °C until analysis. Then all the microsamples were prepared and used to acquire proteomics, metabolomics, and lipidomics data using the protocol described above. All the omics data were provided as **Supplementary Data 1**.

The first metric of stability. After the data generation, annotation, cleaning, imputation, and transformations, each of the omic data sets (proteins, metabolite features, and lipids) were assessed for analyte stability in storage. A total of 128 proteins (n = 66 samples), 1,461 metabolites (no redundant metabolite removal, n = 71 samples), and 776 lipids (n = 72 samples) were available for the stability analysis. The first metric assessed was the coefficient of variation (CV; estimated using the formula for log-transformed data[12]), which was calculated separately across all of the samples for each of the two participants from whom samples were taken. The mean of the two CVs (one from each participant's samples) was used as the CV for that analyte. The distribution of CVs was plotted.

The second metric of stability. The second stability metric was used to identify storage conditions' significant effects on the analyte level. Linear regression was performed for each analyte where the analyte level was regressed on storage duration, temperature, the duration×temperature interaction effect, and an indicator for one of the two participants (to remove the effect of the actual difference in analyte level between the participants). Since the samples that had 0 storage duration were never stored at any temperature, those samples were excluded from the analysis so that the effect of storage temperature could still be estimated, leaving 54, 59, and 60 samples for the protein, metabolite, and lipid analyses. The "lm" function in R was used, and since the objective of the study was to identify analytes that were stable under storage, a simple significance threshold of p-values = 0.05 was used to be more conservative since smaller p-value thresholds would exclude subtler potential effects of storage. The total model R^2 and the partial R^2 for each regression term were calculated using the "rsq" and "rsq.partial" functions of the "rsq" package (version 2.2). The Lindemann, Merenda and Gold (LMG) measure of variable importance[1] was also calculated using the "calc.relimp" function of the "relaimpo" package (version 2.2-6). The proportion of statistically significant effects of storage conditions on analyte level was evaluated against the expected number of significant results at the alpha level of 0.05 to gauge the extent of signal for significant storage effects on the analytes. For each omic data set and storage condition term, the top most associated analytes (according to p-value) were plotted over time and colored by storage temperature to visually examine the identified effects. Since a lack of power might have prevented the identification of some storage effects, each regression analysis was repeated but using two separate models, one testing only storage duration and one testing only storage temperature. The benefit of this change was that the baseline samples could be included in the models testing the effect of storage duration.

Comparison between microsamples and intravenous plasma. To compare the microsampling and conventional intravenous plasma collection approaches, 34 participants were enrolled under the institutional review board (IRB)-approved protocol (IRB-55689 at Stanford University) with written consent. Then one microsampling blood sample and one intravenous plasma sample were collected for each participant. All the samples were immediately saved at the -80 °C for subsequent sample preparation. Then all the samples were prepared and used to acquire untargeted metabolomics and lipidomics data according to the above protocols. For the metabolomics data, after data processing and data curation, 22,858 metabolic features were detected (RPLC positive mode: 7,487 features, RPLC negative mode: 4,662 features, HILIC

positive mode: 6,362 features, HILIC negative mode: 4,374 features). And only 642 features with annotations (MSI levels 1 & 2) remained for subsequent analysis. For the lipidomics data, 616 lipids were detected. All the omics data were provided as **Supplementary Data 2**.

Ensure shake study cohort. Twenty-eight participants were enrolled in the Ensure shake study under the institutional review board (IRB)-approved protocol (IRB-47966 at Stanford University) with written consent. 21 out of 28 participants have completed demographic data (**Figure S3**). The median steady-state plasma glucose (SSPG) is 166, the median age is 64.2 years, and the median body mass index (BMI) is 29.7. Among all the participants, 38% are male, 14.3% are Asian, 14.3% are Black, 66.7% are Caucasian, and 4.8% are Hispanic. All 28 participants were mailed a kit containing microsampling devices (Mitra device), Ensure shake (contains 440 kcal, 66 g carbohydrate, 18 g protein, and 12 g fat), and instructions for the microsampling sample collection. Each participant was instructed to consume the Ensure shake and then collected microsampling blood samples immediately before consuming Ensure shake (baseline, time point 0), and at 30, 60, 120, and 240 min following Ensure shake consumption (**Figure S3b**). Finally, we collected 5 time-point microsamples for each participant (**Figure S3b**). Participants were asked to return their microsamples by overnight mail the same day after blood sample collection. Then all the microsamples were used for multi-omics data acquisition, namely untargeted metabolomics, targeted lipidomics, and cytokine/hormone. Four participants (S6, S26, S31, and S37) without metabolomics data were removed from the final dataset (**Figure S3b**). After data cleaning, curation, and annotation, 768 analytes were detected from the microsamples, containing 560 metabolites, 155 lipids, and 54 cytokines/hormones. All the omics data were provided as **Supplementary Data 3**.

24/7 study cohort. Only one participant (Male, 64 years old) was enrolled in the 24/7 study under IRB-approved protocol (IRB-23602 at Stanford University) with written consent. The microsampling method enables frequent sampling on the order of minutes or hours. However, to make it acceptable and executable, the participant was instructed to perform self-collected finger prick microsamples approximately every hour during waking and every two hours overnight periods sporadically for 7 days (**Figure 4a** and **Figure S8a**). In addition, the participant was also instructed to leverage several wearable devices (Fitbit smartwatch, Dexcom) to acquire comprehensive digital data (wearable data), including the heart rate (HR), step count, continuous glucose monitoring (CGM), and food logging. The microsamples were immediately saved on dry ice upon collection by the participant and then shipped to the laboratory daily. Finally, 97 microsamples in total were collected. They were used to perform in-depth multi-omics data acquisition, including (1) untargeted proteomics, (2) untargeted metabolomics, (3) semi-targeted lipidomics, and (4) targeted assay (cytokine, hormones, total protein, and cortisol). After data processing, curation, and annotation, from the microsamples, we finally detected a total of 2,213 analytes that included 1,051 metabolites, 811 lipids, 291 proteins, 45 cytokines, 13 metabolic panels (cytokines/hormones), 1 total protein, and 1 cortisol. All the data was provided as a resource in **Supplementary Data 7 and 8**.

General statistical, bioinformatics analysis, and data visualization. Most statistical analysis and data visualization were performed using RStudio and R language (version 4.1.2). Most of the R packages and their dependencies used in this study are maintained in CRAN (https://cran.r-project.org/) or Bioconductor (https://bioconductor.org/). The detailed version of all the packages can be found in the **Supplementary Note**. The main script for analysis and data visualization was provided on GitHub (https://github.com/jaspershen/microsampling_multiomics).

In general, before all the statistical analysis, the data are log_2-transformed and then auto-scaled. All the multiple comparisons were adjusted by the Benjamini & Hochberg method (BH) using the "p.adjust" function in R. The R functions "cor" and "cor.test" were used to calculate the Spearman correlation coefficients. The R package "ggplot2" was used to perform most of the data visualization in this study. The R package "Rtsne" was used for the tSNE analysis in the Ensure shake study.

Differentially expressed molecules after consuming Ensure shake. In the Ensure shake study, the time point 0 (before consuming Ensure shake) was set as the baseline, and all the other 4 time points were compared to the baseline to get the differentially expressed molecules (metabolites, lipids, and cytokines/hormones). The paired Wilcoxon rank-sum test ("wilcox.test" function of R) was utilized to get the *p*-values. The multiple comparisons were adjusted using the Benjamini & Hochberg (BH) method ("p.adjust" function of R). And the adjusted *p*-values less than 0.05 were considered as significantly differentially expressed molecules. Then the number of significant molecules whose level had changed at different time points was visualized using a Sankey plot ("ggalluvial" package of R). Next, after consuming Ensure shake across all the time points, we identified the entire set of molecules whose levels changed. The ANOVA test ("anova_test" function from the "rstatix" package in R) was utilized to calculate the *p*-values and then adjusted using the BH method. To evaluate if the significantly expressed molecules we found were random or not, a permutation test was performed. In brief, the sample labels of omics data were randomly shifted to get the random datasets. Then the same method (ANOVA test) was utilized to find the altered molecules for the random dataset. This step was repeated 100 times to get a null distribution of differential molecules. Then the permutation *p*-value was calculated to evaluate if the expressed molecules were random.

Consensus clustering. In the Ensure shake study, the unsupervised K-means consensus-clustering of all samples was performed with the R packages "CancerSubtypes" and "ConsensusClusterPlus" using the significantly shifted molecules that were discovered after consuming the Ensure shake[69]. The data was log_2-transformed first and then auto-scaled. Samples clusters were detected based on K-means clustering, Euclidean distance, and 1,000 resampling repetitions in the "ExecuteCC" function in the range of 2 to 6 clusters. The generated empirical cumulative distribution function (CDF) plot initially showed the optional separation of 2 clusters for all samples. To further decide how many groups (k) should be generated, the silhouette information from clustering was extracted using the "silhouette_SimilarityMatrix function". We compared k = 2, 3, 4, and 5 and found that when k = 2, we got high stability for clustering (**Figure S4c**). And from the consensus matrix heatmaps, 2 groups seem to have the best good clustering (**Figure S4d**). So finally, all the samples were assigned to 2 groups.

Fuzzy c-means clustering. The R package "Mfuzz" was used for fuzzy c-means clustering[70]. In brief, the omics data was first log_2-transformed and auto-scaled, and then the minimum centroid distances were calculated for cluster numbers from 2 to 22 by step 1. The minimum centroid distance is used as the cluster validity index. Then the optimal cluster number was selected according to rule[71]. To get a more accurate cluster number, the clusters whose center expression data correlations are more than 0.8 were merged as one cluster. Then the optimal cluster number was used to do the fuzzy c-means clustering. For each cluster, only the molecules with memberships of more than 0.5 were retained for subsequent analysis.

Metabolic scores. Participant S18 was considered as an outlier in the baseline and removed from the dataset for subsequent analysis (**Figure S5**). Then 5 metabolic scores were calculated: (1) Three carbohydrates (fructose, lactic acid, and pyruvic acid) were detected and used to calculate the carbohydrate score, which represents the human's ability to metabolize carbohydrates (**Figure S6**). (2) Nine amino acids (alloisoleucine, Alanine, isoleucine, methionine, norvaline, phenylalanine, tryptophan, tyrosine, and L-phenylalanine) were detected and used to calculate the amino acid score (protein), which represents the human's ability to metabolize proteins (**Figure S6**). (3) 103 TAGs were detected and used to calculate the fat score, representing the human's ability to metabolize the fat (**Figure S6**). (4) The C-peptide and insulin were detected and used to calculate the insulin secretion score, representing the human's ability to secrete insulin (**Figure S6**). (5) The 8 free fatty acids (FFA 16:0, FFA 16:1, FFA 18:1, FFA 18:2, FFA 18:3, FFA 22:2, FFA 22:5, and FFA 22:6) were detected and used to calculate free fatty acid (insulin sensitivity) score, which represents the human's ability to respond to insulin sensitivity (**Figure S6**). (6) All the cytokines were used to calculate the immune response score representing the human's immune response (**Figure S7a**).

For each metabolic score *MS*, the molecules M_i (i = 1, 2, 3 … m) in this group were first defined and selected (**Figure 3b**), and then the dataset was $\log_2$-transformed and auto-scaled. For each participant and molecule, the intensity values across all the time points were subtracted by the baseline value, so the baseline value was 0. Then the area under the curve (AUC) $A_{i,j}$ was calculated for molecule M_i (i = 1, 2, 3 … m) and participant P_j (j = 1, 2, 3 … n). To normalize the $A_{i,j}$, the $A_{i,j}$ were subtracted by the minimum min($A_{i,j}$) and divided by the range of all the AUCs (max($A_{i,j}$) - min($A_{i,j}$)). The normalized $A_{i,j}$ is labeled as $NA_{i,j}$ and is from 0 to 1. Then for each metabolic score MS_j in each participant j, it is calculated as below:

$$\text{Metabolic score}_j = \text{mean}(\sum_i^m \quad NA_i)$$

Where MS_j is the metabolic score for participant j, and NA_i is the normalized AUCs of molecule i (i = 1, 2, 3 … m). For the carbohydrate score, amino acid (protein) score, fat score, and free fatty acid score (insulin sensitivity), the high AUCs of molecules mean that the person's ability to metabolize the molecules is low, so the final metabolic scores were calculated as 1 - MS_j. For the insulin secretion score and immune response score, the final score is the same as the MS_j.

Enrichment analysis

Metabolomics pathway enrichment. To do the metabolomics pathway enrichment, the human KEGG pathway database was downloaded from KEGG using the R package massDatabase[72]. The original KEGG database has 275 metabolic pathways. Then we separated them into metabolic pathways or disease pathways based on the "Class" information for each pathway. The pathways with the "Human Disease" class were assigned to the disease pathway database, which contains 74 pathways, and remained 201 pathways were assigned to the metabolic pathway database. The pathway enrichment analysis is used in the Hypergeometric distribution test from the tidyMass project[73]. The BH method was used to adjust *p*-values, and the cutoff was set as 0.05 (BH-adjusted *p*-values < 0.05).

Lipidomics data enrichment analysis. The Lipid Mini-on software was utilized to do the lipid enrichment analysis[44]. In brief, the lipids' names were first modified to meet the requirement of the tool. The dysregulated lipids were uploaded as query files, and all the detected lipids were uploaded as universe files. The default Fisher's exact test was used as the enrichment test method. The category, main class, subclass, individual chains, individual chain length, and a number of double bonds were selected for general

parameters to test. Finally, the enrichment result containing detailed tables and networks was downloaded for subsequent analysis.

Proteomics pathway enrichment. The R package "clusterProfiler" was utilized for proteomics pathway enrichment. We first converted the gene ID of proteins to ENTREZID ID, and then the GO database was utilized for GO term enrichment analysis. The *p*-values were adjusted using the BH method, and the cutoff was set as 0.05. Only the enriched GO terms with at least mapped 5 proteins remained to ensure that the enriched GO terms have enough genes. To reduce the redundancy of enriched GO terms, the similarity between GO terms was calculated using the "Wang" algorithm from the R package "simplifyEnrichment"[74]. And only the connections with similarities > 0.3 remained to construct the GO term similarity network. Then the community analysis (R package "igraph") was utilized to divide this network into different modules. The GO term with the smaller enrichment adjusted *p*-values was selected for each module as the representative.

LOESS smoothing data. In the 24/7 study, the time points of microsamples for each day differ. However, the circadian analysis requires enough time points for each day. So we leveraged the locally estimated scatterplot smoothing (LOESS) method to smooth and predict the multi-omics data in the specific time points (every half hour) described in another publication[75]. In brief, for each molecule, we fitted it with the LOESS regression method for each day ("loess" function in R). During the fitting, LOESS's argument "span" was optimized by cross-validation. As the gap between two days is always more than 4 hours, we didn't fit the time between two days for an accurate and robust fitting and prediction. After getting the LOESS prediction model, we predicted each molecule's intensity every half hour during the days (**Figure S8**).

Correlation network and community analysis. In the 24/7 study, we constructed a correlation network for each cluster that we got using the fuzzy c-means clustering. In brief, the Spearman correlation was calculated for every two molecules. Only the correlations with coefficient > 0.7 and BH adjusted *p*-values < 0.05 remained for subsequent analysis. All the remained correlations were used to construct the correlation network. To get more accurate and distinct modules, we use the community analysis to extract subnetworks (modules) from the correlation network[31]. Here, we used the fast greedy modularity optimization algorithm ("cluster_fast_greedy" function from the R package "igraph"). Finally, 11 clusters and 83 modules were detected. The R packages "igraph" and "ggraph" were used to visualize the network.

Associations between molecular modules and nutrition intake. In the 24/7 study, to evaluate the associations between molecular modules and nutrition intake, peak detection (Gaussian distribution fitting) was first utilized to find the "peaks" in each module (**Figure S10f**). If there is a peak, then it is marked as "1" at this time. If not, it is marked as "0". For food, if the participant consumes this food at this time point, then this time point will be marked as "1" for this food. And then, for each food and module, the Jaccard index was calculated, and only the pairs with a Jaccard index > 0.3 were retained for subsequent analysis (**Figure S10g**).

Consistency score for molecules. In the 24/7 study, the consistency score was designed and calculated for each molecule to assess if one molecule is consistent daily. LOESS smoothed data was used for consistency score calculation. For each molecule, the Spearman correlations between two days were calculated, and the

median correlation value was calculated and considered as the consistency score for this molecule. Only the molecules with consistency scores > 0.6 were retained for the next circadian analysis.

Circadian rhythm analysis. In the 24/7 study, the R package "MetaCycle" is used to do the circadian rhythm analysis[42]. The LOESS smoothed omics data were $\log_2$-transformed and auto-scaled. And then, the times for samples were set as the time points in the "meta2d" function. The Lomb-Scargle was selected for circadian rhythm analysis[76]. The *p*-values were adjusted using the BH method. Only the molecules with BH-adjusted *p*-values < 0.05 were considered statistically significant circadian molecules and retained for subsequent analysis.

Wearable data predicts internal molecules. In the 24/7 study, to evaluate if the wearable data could be utilized to predict internal molecules, the method from a previous publishment[30] was used. Because the frequency of wearable data and internal molecules are different, we need to match the internal molecule and wearable data first. The matching windows were set as 5, 10, 20, 30, 40, 50, 60, 90, and 120 mins, respectively. And for the wearable data points that matched with internal molecules, a feature engineering pipeline[30] was utilized to convert the wearable data into eight features: mean value, median value, standard, maximum, minimum, skewness, kurtosis, and range. So for each wearable data, it was converted into 8 features. The wearable data (heart rate, step count, and CGM) were converted to 24 features in total and were used as independent variables to predict each internal molecule. The random forest model (R package "caret" and "RandomForest"), which has been proven to have the best prediction accuracy, was utilized[30]. The 24 wearable features were combined for each internal molecule to construct the prediction model. The 7-fold cross-validation method was used during the prediction model construction. The importance of each wearable feature was saved for subsequent analysis.

Lagged correlation. In the 24/7 study, to calculate the lagged correlation between wearable data and internal molecules, we have developed the *laggedCor* algorithm (lagged correlation) and an R package named "laggedcor" (https://jaspershen.github.io/laggedcor/). *The laggedCor* algorithm can be used to extract potential causal relationships. Let's assume that X is wearable data and Y is internal omics data. In a real biological system, if X and Y have a causal relationship (X causes Y), Y often responds to X after a certain lapse of time. Such a lapse of time is called a lag time. So it means that X and Y change asynchronously. To explore if X and Y have a potential causal relationship, we just shift the lag time between X and Y for matching and then calculate the correlation between them. Suppose the X and Y have a potential causal relationship and the lag time is T, then we can get the highest lagged correlation between X and Y at the lag time T.

Briefly, two time-series data are used as the inputs for *laggedcor*. The lower frequency time-series data (in the 24/7 study is the omics data) is labeled as X_t ($t \in Ti$), and the higher frequency time-series data (in the 24/7 study is the wearable data) is labeled as Y_t ($t \in Tj$). To make sure that there are overlaps between X_{ti} and Y_{tj}, they should meet the below equation:

$$Ti \cap Tj \neq \emptyset$$

then the two series data, X_t and Y_t are used to calculate the lagged correlation as described below steps.

Step 1: Matching between X_t and Y_t. Every sample point Y_{tj} in Y is used to match the sample points in X_t. The shift time is labeled as Ts (Ts is set based on the frequency of X_t and Y_t), and the matching time window

is labeled as Tw. So the sample points X_{ti} in X_t that meet the below equitation are labeled as matched sample points for Y_{tj} in Y.

$$tj + Ts - \frac{Tw}{2} \leq ti < tj + Ts + \frac{Tw}{2}\ ;\ i \in (1, 2, 3 \dots m)$$

The the matched sample points X_{ti} are averaged as X_{tj} that matched with Y_{tj} in Y.

$$X_{tj} = \sum_{ti}^{tm} X_{ti}$$

Then we get the new time-series data X_t ($t \in Tj$).

Step 2: Correlation calculation. Then the Spearman correlation between X_t and Y_t ($t \in Tj$) is calculated with the shift time Ts. And the correlation rho and *p*-value are recorded as Cor_{ts} and p_{ts}.

Step 3: Repeat step 1 and step 2 with different shift time. Then, step 1 and step 2 are repeated for a series shift times Ts_i, i = 1, 2, 3 … n; $Ts_1 < 0$ and $abs(Ts_1) = abs(Ts_n)$. Then we can get a series Cor_{ts} and a series p_{ts}, $ts \in Ts_i$.

Step 4: Evaluation of the significance of lagged correlation. The max correlation of Cor_{ts} and related *p*-value are extracted as the lagged correlation for time-series data X_t and Y_t. To evaluate if the lagged correlation is significant, the Gaussian distribution is used to fit the Cor_{ts}, and the correlations in all the shift times are calculated using the fitted Gaussian distribution and labeled as $PCor_{ts}$. The quality score was then calculated as the absolute Spearman correlation score between $PCor_{ts}$ and Cor_{ts}. Only the lagged correlation with a quality score was considered a real lagged correlation and used for subsequent analysis.

Data availability

All the data used in this study are provided as **Supplementary Data**.

Code availability

R version 4.1.2 was used with the base packages and other packages, and detailed information has been provided in the **Supplementary Note**. All the custom scripts for data analysis and data visualization were provided and open source via https://github.com/jaspershen/microsampling_multiomics. The *laggedCor* algorithm and package were developed for lagged correlation calculation and are open-source via https://jaspershen.github.io/laggedcor/.

Acknowledgments

This work was supported by grants from the National Institute of Health (NIH) 5RM1HG00773508 and 5R01AT01023204.

Author contributions

R.K., D.H., X.S., and M.P.S. conceived and designed the study; D.H., B.L., K.E., R.K. prepared samples and acquired lipidomics, metabolomics, and proteomics data; Y.R-H. prepared samples and generated Luminex data. D.P., N.B., and X.S. performed the stability analysis. X.S. and R.K. analyzed the data of the Ensure shake study; X.S., R.K., and C.W. analyzed the data of the 24/7 study. X.S. and C.W. developed the *laggedCor* algorithm and built the R package. X.S., C.W., and D.P. prepared all the figures. X.S., R.K., N.B., D.H., D.P., C.W., and M.P.S. wrote the manuscript. All the authors contributed to the final version of the manuscript.

Competing interests

M.P.S. is a cofounder and scientific advisor of Personalis, SensOmics, Qbio, January AI, Fodsel, Filtricine, Protos, RTHM, Iollo, Marble Therapeutics, and Mirvie. He is a scientific advisor of Genapsys, Jupiter, Neuvivo, Swaza, and Mitrix. D.H. has a financial interest in Seer Inc. and Prognomiq Inc. All other authors have no competing interests.

Reference

1. Schüssler-Fiorenza Rose, S. M. *et al.* A longitudinal big data approach for precision health. *Nat. Med.* **25**, 792–804 (2019).
2. Zhou, W. *et al.* Longitudinal multi-omics of host-microbe dynamics in prediabetes. *Nature* **569**, 663–671 (2019).
3. Guthrie, R. & Susi, A. A SIMPLE PHENYLALANINE METHOD FOR DETECTING PHENYLKETONURIA IN LARGE POPULATIONS OF NEWBORN INFANTS. *Pediatrics* **32**, 338–343 (1963).
4. Antunes, M. V., Charão, M. F. & Linden, R. Dried blood spots analysis with mass spectrometry: Potentials and pitfalls in therapeutic drug monitoring. *Clin. Biochem.* **49**, 1035–1046 (2016).
5. Corso, G., D'Apolito, O., Gelzo, M., Paglia, G. & Dello Russo, A. A powerful couple in the future of clinical biochemistry: in situ analysis of dried blood spots by ambient mass spectrometry. *Bioanalysis* **2**, 1883–1891 (2010).
6. Bennett, M. J. *et al.* Newborn screening for metabolic disorders: how are we doing, and where are we going? *Clin. Chem.* **58**, 324–331 (2012).
7. Volani, C. *et al.* Pre-analytic evaluation of volumetric absorptive microsampling and integration in a mass spectrometry-based metabolomics workflow. *Anal. Bioanal. Chem.* **409**, 6263–6276 (2017).
8. van den Broek, I. *et al.* Application of volumetric absorptive microsampling for robust, high-throughput mass spectrometric quantification of circulating protein biomarkers. *Clinical Mass Spectrometry* vols 4-5 25–33 Preprint at https://doi.org/10.1016/j.clinms.2017.08.004 (2017).
9. Molloy, M. P. *et al.* Proteomic Analysis of Whole Blood Using Volumetric Absorptive Microsampling for Precision Medicine Biomarker Studies. *J. Proteome Res.* **21**, 1196–1203 (2022).
10. Lei, B. U. W. & Prow, T. W. A review of microsampling techniques and their social impact. *Biomed. Microdevices* **21**, 81 (2019).
11. Zhuang, Y.-J., Mangwiro, Y., Wake, M., Saffery, R. & Greaves, R. F. Multi-omics analysis from archival neonatal dried blood spots: limitations and opportunities. *Clin. Chem. Lab. Med.* (2022) doi:10.1515/cclm-2022-0311.
12. Canchola, J. A. Correct Use of Percent Coefficient of Variation (%CV) Formula for Log-Transformed Data. *MOJ Proteomics & Bioinformatics* vol. 6 Preprint at https://doi.org/10.15406/mojpb.2017.06.00200 (2017).
13. Hannon, B. A. *et al.* Single Nucleotide Polymorphisms Related to Lipoprotein Metabolism Are Associated with Blood Lipid Changes following Regular Avocado Intake in a Randomized Control Trial among Adults with Overweight and Obesity. *The Journal of Nutrition* vol. 150 1379–1387 Preprint at https://doi.org/10.1093/jn/nxaa054 (2020).
14. Berry, S. E. *et al.* Human postprandial responses to food and potential for precision nutrition. *Nat. Med.* **26**, 964–973 (2020).
15. Wang, D. D. *et al.* The gut microbiome modulates the protective association between a Mediterranean

diet and cardiometabolic disease risk. *Nat. Med.* **27**, 333–343 (2021).
16. Turnbaugh, P. J. *et al.* The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci. Transl. Med.* **1**, 6ra14 (2009).
17. Spiekerkoetter, U. Mitochondrial fatty acid oxidation disorders: clinical presentation of long-chain fatty acid oxidation defects before and after newborn screening. *Journal of Inherited Metabolic Disease* vol. 33 527–532 Preprint at https://doi.org/10.1007/s10545-010-9090-x (2010).
18. Gannon, M. C., Nuttall, F. Q., Westphal, S. A. & Seaquist, E. R. The effect of fat and carbohydrate on plasma glucose, insulin, C-peptide, and triglycerides in normal male subjects. *J. Am. Coll. Nutr.* **12**, 36–41 (1993).
19. Ludwig, D. S. *et al.* The carbohydrate-insulin model: a physiological perspective on the obesity pandemic. *The American Journal of Clinical Nutrition* vol. 114 1873–1885 Preprint at https://doi.org/10.1093/ajcn/nqab270 (2021).
20. Hall, K. D. A review of the carbohydrate–insulin model of obesity. *European Journal of Clinical Nutrition* vol. 71 323–326 Preprint at https://doi.org/10.1038/ejcn.2016.260 (2017).
21. Meier, J. J. & Nauck, M. A. Glucagon-like peptide 1(GLP-1) in biology and pathology. *Diabetes/Metabolism Research and Reviews* vol. 21 91–117 Preprint at https://doi.org/10.1002/dmrr.538 (2005).
22. Pederson, R. A. & McIntosh, C. H. S. Discovery of gastric inhibitory polypeptide and its subsequent fate: Personal reflections. *Journal of Diabetes Investigation* vol. 7 4–7 Preprint at https://doi.org/10.1111/jdi.12480 (2016).
23. Katsuura, G., Asakawa, A. & Inui, A. Roles of pancreatic polypeptide in regulation of food intake. *Peptides* vol. 23 323–329 Preprint at https://doi.org/10.1016/s0196-9781(01)00604-0 (2002).
24. Kelesidis, T., Kelesidis, I., Chou, S. & Mantzoros, C. S. Narrative review: the role of leptin in human physiology: emerging clinical applications. *Ann. Intern. Med.* **152**, 93–100 (2010).
25. Kim, S. H. & Reaven, G. M. Insulin Resistance and Hyperinsulinemia. *Diabetes Care* vol. 31 1433–1438 Preprint at https://doi.org/10.2337/dc08-0045 (2008).
26. Xin, Y. *et al.* Elevated free fatty acid level is associated with insulin-resistant state in nondiabetic Chinese people. *Diabetes Metab. Syndr. Obes.* **12**, 139–147 (2019).
27. Chen, R. *et al.* Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell* **148**, 1293–1307 (2012).
28. Piening, B. D. *et al.* Integrative Personal Omics Profiles during Periods of Weight Gain and Loss. *Cell Systems* vol. 6 157–170.e8 Preprint at https://doi.org/10.1016/j.cels.2017.12.013 (2018).
29. Gao, P. *et al.* Precision environmental health monitoring by longitudinal exposome and multi-omics profiling. *Genome Res.* **32**, 1199–1214 (2022).
30. Dunn, J. *et al.* Wearable sensors enable personalized predictions of clinical laboratory measurements. *Nat. Med.* **27**, 1105–1112 (2021).
31. Price, N. D. *et al.* A wellness study of 108 individuals using personal, dense, dynamic data clouds. *Nat. Biotechnol.* **35**, 747–756 (2017).
32. Burton-Pimentel, K. J. *et al.* Discriminating Dietary Responses by Combining Transcriptomics and Metabolomics Data in Nutrition Intervention Studies. *Mol. Nutr. Food Res.* **65**, e2000647 (2021).
33. Contrepois, K. *et al.* Molecular Choreography of Acute Exercise. *Cell* **181**, 1112–1130.e16 (2020).
34. DiNicolantonio, J. J. & O'Keefe, J. H. Effects of dietary fats on blood lipids: a review of direct comparison trials. *Open Heart* **5**, e000871 (2018).
35. Mohd Azmi, N. A. S. *et al.* Cortisol on Circadian Rhythm and Its Effect on Cardiovascular System.

Int. J. Environ. Res. Public Health **18**, (2021).
36. Stachowicz, M. & Lebiedzińska, A. The effect of diet components on the level of cortisol. *European Food Research and Technology* vol. 242 2001–2009 Preprint at https://doi.org/10.1007/s00217-016-2772-3 (2016).
37. Cao, W. *et al.* Drug-drug interactions between salvianolate injection and aspirin based on their metabolic enzymes. *Biomed. Pharmacother.* **135**, 111203 (2021).
38. Watson, E. J., Coates, A. M., Kohler, M. & Banks, S. Caffeine Consumption and Sleep Quality in Australian Adults. *Nutrients* **8**, (2016).
39. Clark, I. & Landolt, H. P. Coffee, caffeine, and sleep: A systematic review of epidemiological studies and randomized controlled trials. *Sleep Med. Rev.* **31**, 70–78 (2017).
40. Fishbein, A. B., Knutson, K. L. & Zee, P. C. Circadian disruption and human health. *Journal of Clinical Investigation* vol. 131 Preprint at https://doi.org/10.1172/jci148286 (2021).
41. Patke, A., Young, M. W. & Axelrod, S. Molecular mechanisms and physiological importance of circadian rhythms. *Nat. Rev. Mol. Cell Biol.* **21**, 67–84 (2020).
42. Wu, G., Anafi, R. C., Hughes, M. E., Kornacker, K. & Hogenesch, J. B. MetaCycle: an integrated R package to evaluate periodicity in large scale data. *Bioinformatics* **32**, 3351–3353 (2016).
43. Hughes, M. E., Hogenesch, J. B. & Kornacker, K. JTK_CYCLE: An Efficient Nonparametric Algorithm for Detecting Rhythmic Components in Genome-Scale Data Sets. *Journal of Biological Rhythms* vol. 25 372–380 Preprint at https://doi.org/10.1177/0748730410379711 (2010).
44. Clair, G. *et al.* Lipid Mini-On: mining and ontology tool for enrichment analysis of lipidomic data. *Bioinformatics* **35**, 4507–4508 (2019).
45. Chua, E. C.-P. *et al.* Extensive diversity in circadian regulation of plasma lipids and evidence for different circadian metabolic phenotypes in humans. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 14468–14473 (2013).
46. Alavi, A. *et al.* Real-time alerting system for COVID-19 and other stress events using wearable data. *Nat. Med.* **28**, 175–184 (2022).
47. Mishra, T. *et al.* Pre-symptomatic detection of COVID-19 from smartwatch data. *Nature Biomedical Engineering* **4**, 1208–1220 (2020).
48. Marabita, F. *et al.* Multiomics and digital monitoring during lifestyle changes reveal independent dimensions of human biology and health. *Cell Syst* **13**, 241–255.e7 (2022).
49. Miller, I. J. *et al.* Real-time health monitoring through urine metabolomics. *NPJ Digit Med* **2**, 109 (2019).
50. Hall, H. *et al.* Glucotypes reveal new patterns of glucose dysregulation. *PLoS Biol.* **16**, e2005143 (2018).
51. Bizzarri, M. *et al.* A call for a better understanding of causation in cell biology. *Nat. Rev. Mol. Cell Biol.* **20**, 261–262 (2019).
52. Shomali, N. *et al.* Harmful effects of high amounts of glucose on the immune system: An updated review. *Biotechnol. Appl. Biochem.* **68**, 404–410 (2021).
53. Bitar, A., Mastouri, R. & Kreutz, R. P. Caffeine Consumption and Heart Rate and Blood Pressure Response to Regadenoson. *PLoS One* **10**, e0130487 (2015).
54. Rodriguez-Araujo, G. *et al.* Alpha-synuclein elicits glucose uptake and utilization in adipocytes through the Gab1/PI3K/Akt transduction pathway. *Cell. Mol. Life Sci.* **70**, 1123–1133 (2013).
55. Wijesekara, N. *et al.* α-Synuclein Regulates Peripheral Insulin Secretion and Glucose Transport. *Front. Aging Neurosci.* **13**, 665348 (2021).

56. Buckingham, B. *et al.* CGM-measured glucose values have a strong correlation with C-peptide, HbA1c and IDAAC, but do poorly in predicting C-peptide levels in the two years following onset of diabetes. *Diabetologia* **58**, 1167–1174 (2015).
57. Paniagua-González, L. *et al.* Comparison of conventional dried blood spots and volumetric absorptive microsampling for tacrolimus and mycophenolic acid determination. *J. Pharm. Biomed. Anal.* **208**, 114443 (2022).
58. Andersen, I. K. L., Rosting, C., Gjelstad, A. & Halvorsen, T. G. Volumetric absorptive MicroSampling vs. other blood sampling materials in LC–MS-based protein analysis – preliminary investigations. *Journal of Pharmaceutical and Biomedical Analysis* vol. 156 239–246 Preprint at https://doi.org/10.1016/j.jpba.2018.04.036 (2018).
59. Spooner, N. *et al.* A device for dried blood microsampling in quantitative bioanalysis: overcoming the issues associated blood hematocrit. *Bioanalysis* **7**, 653–659 (2015).
60. Li, X. *et al.* Digital Health: Tracking Physiomes and Activity Using Wearable Biosensors Reveals Useful Health-Related Information. *PLoS Biol.* **15**, e2001402 (2017).
61. Lancaster, S. M. *et al.* Global, distinctive, and personal changes in molecular and microbial profiles by specific fibers in humans. *Cell Host Microbe* **30**, 848–862.e7 (2022).
62. Bararpour, N. *et al.* DBnorm as an R package for the comparison and selection of appropriate statistical methods for batch effect correction in metabolomic studies. *Scientific Reports* vol. 11 Preprint at https://doi.org/10.1038/s41598-021-84824-3 (2021).
63. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
64. Shen, X. *et al.* metID: an R package for automatable compound annotation for LC−MS-based data. *Bioinformatics* vol. 38 568–569 Preprint at https://doi.org/10.1093/bioinformatics/btab583 (2022).
65. Sumner, L. W. *et al.* Proposed minimum reporting standards for chemical analysis. *Metabolomics* vol. 3 211–221 Preprint at https://doi.org/10.1007/s11306-007-0082-2 (2007).
66. Contrepois, K. *et al.* Cross-Platform Comparison of Untargeted and Targeted Lipidomics Approaches on Aging Mouse Plasma. *Sci. Rep.* **8**, 17747 (2018).
67. Shah, J. S. *et al.* Distribution based nearest neighbor imputation for truncated high dimensional data with applications to pre-clinical and clinical metabolomics studies. *BMC Bioinformatics* **18**, 114 (2017).
68. Bahmani, A. *et al.* A scalable, secure, and interoperable platform for deep data-driven health management. *Nat. Commun.* **12**, 5757 (2021).
69. Xu, T. *et al.* CancerSubtypes: an R/Bioconductor package for molecular cancer subtype identification, validation and visualization. *Bioinformatics* vol. 33 3131–3133 Preprint at https://doi.org/10.1093/bioinformatics/btx378 (2017).
70. Kumar, L. & E Futschik, M. Mfuzz: a software package for soft clustering of microarray data. *Bioinformation* **2**, 5–7 (2007).
71. Schwämmle, V. & Jensen, O. N. A simple and fast method to determine the parameters for fuzzy c-means cluster analysis. *Bioinformatics* **26**, 2841–2848 (2010).
72. Shen, X., Wang, C. & Snyder, M. P. massDatabase: utilities for the operation of the public compound and pathway database. *Bioinformatics* **38**, 4650–4651 (2022).
73. Shen, X. *et al.* TidyMass an object-oriented reproducible analysis framework for LC-MS data. *Nat. Commun.* **13**, 4365 (2022).
74. Gu, Z. & Hübschmann, D. *simplifyEnrichment*: an R/Bioconductor package for Clustering and

Visualizing Functional Enrichment Results. Preprint at https://doi.org/10.1101/2020.10.27.312116.
75. Lehallier, B. *et al.* Undulating changes in human plasma proteome profiles across the lifespan. *Nat. Med.* **25**, 1843–1850 (2019).
76. Glynn, E. F., Chen, J. & Mushegian, A. R. Detecting periodic patterns in unevenly spaced gene expression time series using Lomb-Scargle periodograms. *Bioinformatics* **22**, 310–316 (2006).