Integrative Personal Omics Profiles during Periods of Weight Gain and Loss

Downloaded from: https://research.chalmers.se, 2024-08-16 10:23 UTC

Citation for the original published paper (version of record):
Piening, B., Zhou, W., Contrepois, K. et al (2018). Integrative Personal Omics Profiles during Periods of Weight Gain and Loss. Cell Systems, 6(2): 157-170.e8.
http://dx.doi.org/10.1016/j.cels.2017.12.013

N.B. When citing this work, cite the original published paper.

research.chalmers.se offers the possibility of retrieving research publications produced at Chalmers University of Technology.
It covers all kind of research output: articles, dissertations, conference papers, reports etc. since 2004.
research.chalmers.se is administrated and maintained by Chalmers Library

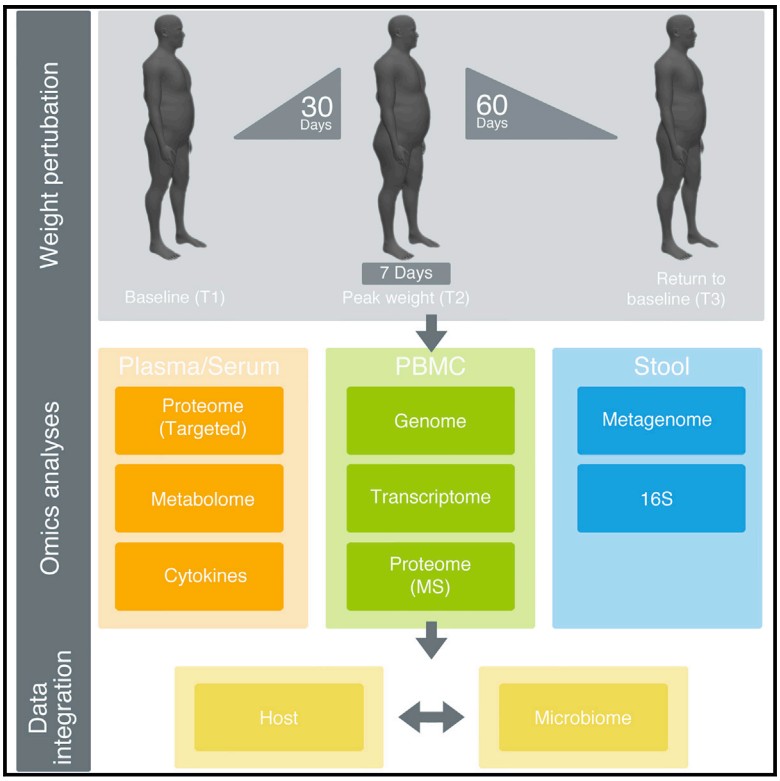
(article starts on next page)

Article

Cell Systems

Integrative Personal Omics Profiles during Periods of Weight Gain and Loss

Graphical Abstract



Highlights

- Analysis of insulin-sensitive and -resistant individuals across multiple omes
- We reveal molecular changes in the human body after modest short-term weight gain
- Reversal of short-term weight gain eliminates negative molecular alterations
- Open-access multi-omic resource for further bioinformatics developments

Authors

Brian D. Piening, Wenyu Zhou, Kévin Contrepois, ..., Tracey L. McLaughlin, George M. Weinstock, Michael P. Snyder

Correspondence

tmclaugh@stanford.edu (T.L.M.), george.weinstock@jax.org (G.M.W.), mpsnyder@stanford.edu (M.P.S.)

In Brief

Extensive multi-omic profiling of the blood and microbiomes of healthy and insulin-resistant humans as they gain and lose weight reveals insights into the systemic impacts of weight gain.

Piening et al., 2018, Cell Systems 6, 157–170
February 28, 2018 © 2017 Elsevier Inc.
https://doi.org/10.1016/j.cels.2017.12.013

CellPress

Cell Systems
Article

Integrative Personal Omics Profiles during Periods of Weight Gain and Loss

Brian D. Piening,[1,16] Wenyu Zhou,[1,16] Kévin Contrepois,[1,16] Hannes Röst,[1,16] Gucci Jijuan Gu Urban,[1,10] Tejaswini Mishra,[1] Blake M. Hanson,[2] Eddy J. Bautista,[2,15] Shana Leopold,[2] Christine Y. Yeh,[1,4,5,11] Daniel Spakowicz,[2] Imon Banerjee,[12] Cynthia Chen,[12] Kimberly Kukurba,[1] Dalia Perelman,[3] Colleen Craig,[3] Elizabeth Colbert,[3] Denis Salins,[1] Shannon Rego,[1] Sunjae Lee,[7] Cheng Zhang,[7] Jessica Wheeler,[1] M. Reza Sailani,[1] Liang Liang,[1] Charles Abbott,[1] Mark Gerstein,[6,13,14] Adil Mardinoglu,[7,8] Ulf Smith,[9] Daniel L. Rubin,[12] Sharon Pitteri,[4,5] Erica Sodergren,[2] Tracey L. McLaughlin,[3,16,*] George M. Weinstock,[2,16,*] and Michael P. Snyder[1,16,17,*]

[1]Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA
[2]The Jackson Laboratory for Genomic Medicine, Farmington, CT 06032, USA
[3]Division of Endocrinology, Stanford University School of Medicine, Stanford, CA 94305, USA
[4]Department of Radiology, Stanford University School of Medicine, Stanford, CA 94305, USA
[5]Canary Center at Stanford, Stanford University School of Medicine, Stanford, CA 94305, USA
[6]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA
[7]Science for Life Laboratory, KTH - Royal Institute of Technology, Stockholm, Sweden
[8]Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden
[9]Department of Molecular and Clinical Medicine, University of Gothenburg, Gothenburg, Sweden
[10]Department of Immunology Genetics and Pathology, Uppsala University, Uppsala, Sweden
[11]Biomedical Informatics Program, Stanford University School of Medicine, Stanford, CA 94305, USA
[12]Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, CA 94305, USA
[13]Department of Computer Science, Yale University, New Haven, CT, USA
[14]Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT, USA
[15]Present address: Corporación de Investigación Agropecuaria Corpoica, Centro de Investigación Tibaitatá, Mosquera, Colombia
[16]These authors contributed equally
[17]Lead Contact
*Correspondence: tmclaugh@stanford.edu (T.L.M.), george.weinstock@jax.org (G.M.W.), mpsnyder@stanford.edu (M.P.S.)
https://doi.org/10.1016/j.cels.2017.12.013

SUMMARY

Advances in omics technologies now allow an unprecedented level of phenotyping for human diseases, including obesity, in which individual responses to excess weight are heterogeneous and unpredictable. To aid the development of better understanding of these phenotypes, we performed a controlled longitudinal weight perturbation study combining multiple omics strategies (genomics, transcriptomics, multiple proteomics assays, metabolomics, and microbiomics) during periods of weight gain and loss in humans. Results demonstrated that: (1) weight gain is associated with the activation of strong inflammatory and hypertrophic cardiomyopathy signatures in blood; (2) although weight loss reverses some changes, a number of signatures persist, indicative of long-term physiologic changes; (3) we observed omics signatures associated with insulin resistance that may serve as novel diagnostics; (4) specific biomolecules were highly individualized and stable in response to perturbations, potentially representing stable personalized markers. Most data are available open access and serve as a valuable resource for the community.

INTRODUCTION

With 34% of the US adult population classified as obese (Flegal et al., 2010) and 3 million obesity-attributable deaths worldwide (Finucane et al., 2011), it is imperative that we gain a better understanding of the factors contributing to obesity-associated morbidities, especially cardiovascular disease, cancer, and type 2 diabetes mellitus (T2DM). These diseases are mediated in part by insulin resistance (Kahn et al., 2006). While insulin resistance is overall positively correlated with measures of adiposity, there is great heterogeneity in metabolic phenotypes between individuals who are similarly obese; as such, our ability to predict these phenotypes is poor. One reason for this is that the onset of metabolic disease and related complications involves numerous pathways and complex interactions between metabolically active tissues such as fat, liver, and muscle (Kahn et al., 2006) (and most recently implicated, the microbiome; Janssen and Kersten, 2017). As such, a more comprehensive molecular profiling approach may offer novel insights into the diversity of systemic responses to the increase in body fat stores.

While most human studies have simply examined plasma markers that occur in association with increasing adiposity, another way to approach this problem is to compare equally obese individuals who are either insulin resistant (IR) or insulin sensitive (IS) to characterize differences between these metabolically disparate groups. Many metabolic risk markers differ markedly as a function of IR/IS independent of adiposity

(McLaughlin et al., 2002, 2007, 2006). Induction of obesity in mice via high-fat diets has revealed important insights into the biological links between weight gain and metabolic disease. While inflammation and oxidative stress are clear players in mice, there are very few studies on determinants of insulin resistance/metabolic disease in response to weight gain and loss in humans, and causality is harder to determine given the relative difficulty in manipulating genes or proteins in human subjects.

The integration of multiple large-scale omics profiling technologies across biological fluids and tissues has recently been demonstrated to offer novel insights into disease development and progression, including T2DM (Chen et al., 2012; Hood et al., 2015; Lee et al., 2016; Price et al., 2017; Williams et al., 2016). In this context, we performed a controlled short-term weight gain and weight loss intervention in humans. Overweight to moderately obese (BMI 25–35 kg/m^2) healthy, nondiabetic participants were selected to span a wide range of insulin resistance. This group was targeted since they are most at risk for T2DM, and, by subjecting individuals of different insulin resistance profiles to identical weight perturbations, we sought to identify pathways involved in obesity-mediated insulin resistance. Furthermore, subjects already overweight require little additional weight gain to trigger stress responses in adipose tissue (McLaughlin et al., 2014), which contributes to systemic responses mediating disease states such as type 2 diabetes and cardiovascular disease (McLaughlin et al., 2016). Unlike many prior studies, the current human study entailed a perturbation in body weight such that changes with gain and loss could be evaluated and, in particular, with insulin resistance, which was quantified by a gold-standard physiologic measure (modified insulin suppression test; Greenfield et al., 1981; Pei et al., 1994; Shen et al., 1970). Multi-omics profiling including genomics; transcriptomics; proteomics; metabolomics from blood peripheral blood mononuclear cells (PBMCs), plasma, and serum; and microbiomics from stool was performed over the course of the study to generate a wealth of personal longitudinal data, thereby enabling a detailed map of the individual molecular changes that occur in response to weight gain and weight loss in the context of insulin sensitivity and resistance. Our study revealed a number of important findings, including (1) modest weight gain in overweight humans was associated with the activation of inflammatory signatures in the blood, induction of markers and pathways for cardiovascular disease, and significant changes to the microbiome; (2) many of these changes affected IR participants differently from metabolically healthy controls (e.g., microbiome and metabolome differences); (3) the longitudinal nature of the designed perturbation results in statistical power increases that are equivalent to orders-of-magnitude larger cross-sectional studies, an important consideration for the design of future large-scale (and costly) multi-omics studies.

RESULTS

Overview of the Multi-omics Study

The overall goal of the study was 2-fold: (1) assemble a comprehensive map of the molecular changes in humans (in circulating blood as well as the microbiome) that occur over the course of a carefully controlled weight gain and their reversibility with weight loss; and (2) determine whether IS and IR individuals who are matched for degree of obesity demonstrate unique biomolecular signatures and/or pathway activation during similar weight gain. Identifying specific molecules and/or pathways that characterize IR versus IS individuals may reveal the fundamental mechanisms by which obesity potentiates insulin resistance and associated diseases such as type 2 diabetes and cardiovascular disease. Participants were recruited as part of the current iteration of the Human Microbiome Project (iHMP) (Integrative HMP (iHMP) Research Network Consortium, 2014), and omic data are open access (http://hmp2.org/).

We sampled 23 carefully selected healthy participants with BMI 25–35 kg/m^2 at baseline before perturbation and during periods of short-term weight gain followed by weight loss, with samples drawn and metabolic measurements made at baseline (T1), peak weight (T2), and post weight loss (T3) (Figure 1), and after 3 months of weight stability following return to baseline weight (T4). All subjects provided written informed consent and all evaluations and blood/stool samples were obtained in the Clinical and Translational Research Unit (CTRU) after an overnight fast. Metabolic phenotyping included quantification of insulin-mediated glucose uptake using the modified insulin suppression test (Pei et al., 1994; Shen et al., 1970), which replaces endogenous insulin secretion with a controlled intravenous infusion of insulin and glucose such that, at steady state, all individuals will have the same insulin concentration but different plasma glucose concentrations, which reflect the relative ability of insulin to dispose of a glucose load. In this test, which has been validated against the euglycemic clamp method (Greenfield et al., 1981), a high steady-state plasma glucose (SSPG) level indicates relative resistance to insulin action. Thirteen fasted overweight IR participants (defined by SSPG $\geq$ 150 mg/dL; Yeni-Komshian et al., 2000, Table S1, and STAR Methods) and ten BMI-matched healthy overweight IS participants completed the study. Three others dropped out after baseline tests and were not included in the analyses. Comprehensive anthropometric, clinical, and plasma measures of general health (waist circumference, blood pressure, hematocrit, cholesterol, triglycerides, liver function tests, creatinine, etc.) were also performed (Table S1).

Blood and stool were sampled from fasted subjects at baseline (T1, Figure 1A); participants then underwent a controlled hypercaloric diet for a period of 30 days, with each participant supplementing their normal diets with high-caloric foods provided by the Stanford CTRU Research Kitchen (see STAR Methods) to achieve an average excess of 880 kcal per day as determined for each individual by resting metabolic caloric requirement and an activity factor (STAR Methods). At the end of the 30-day weight gain period, participants maintained a eucaloric diet for 7 days, at which point a second fasted sample of blood and stool was collected (T2, Figure 1A). Participants gained an average of 2.8 kg over the course of the perturbation (Figure S1). Each participant then underwent a caloric-restricted diet under nutritionist supervision (see STAR Methods) for a subsequent 60-day period designed to return each participant back to his/her initial baseline weight, at which point a third set of fasted samples of blood and stool were collected (T3, Figure 1A). A subset of participants returned for a follow-up sampling approximately 3 months after the end of the perturbation (T4). A total of 90 time points were sampled. The large-scale multi-omics

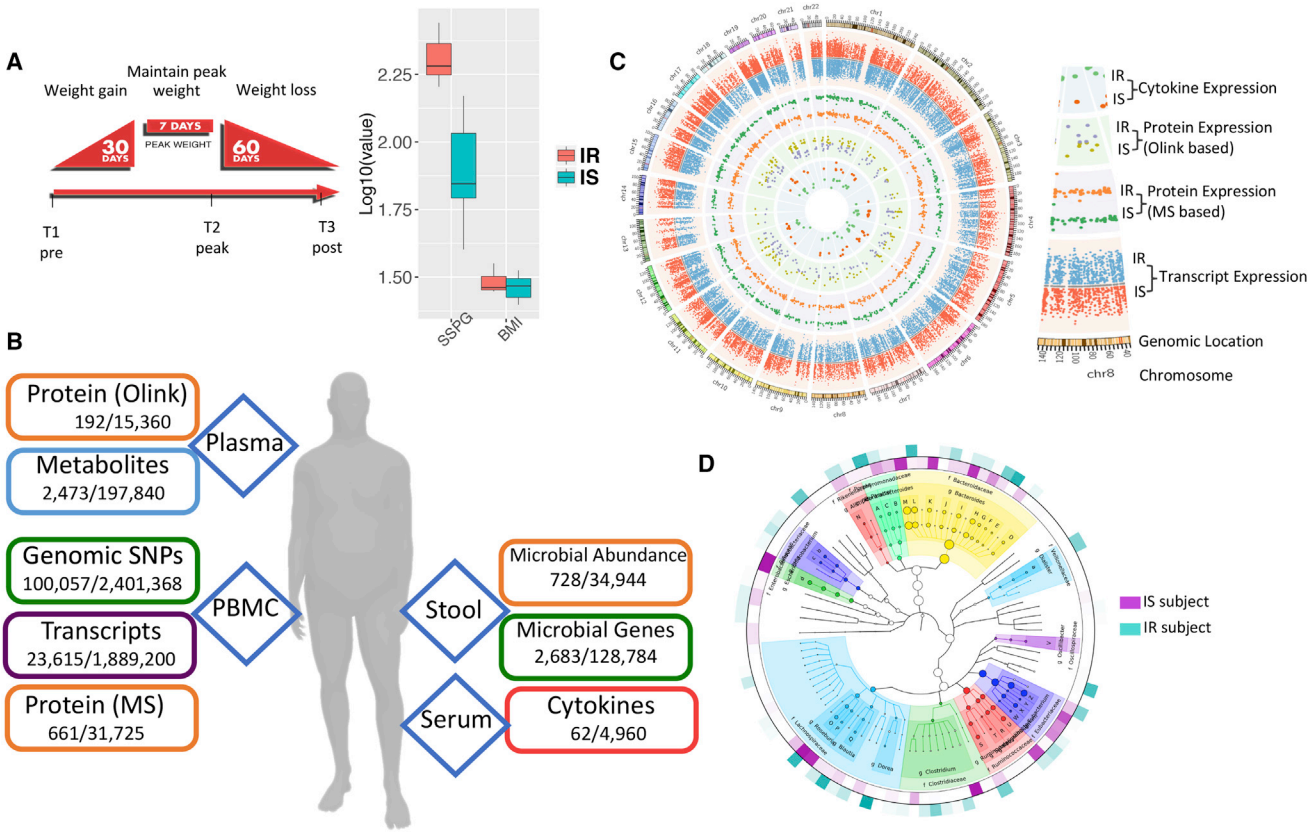


Figure 1. Overview of the Multi-omic Weight Perturbation Experiment

(A) Schematic of the weight gain and loss perturbation. The sampling time points (T1–T3) are indicated at the specific time in the perturbation when they occur. Inset: SSPG and BMI measurements for IS and IR subjects.

(B) The types of omics analyses performed are indicated along with the types of biological materials they are performed on. Listed below each data type are the number of analytes measured per time point for each individual, as well as the total number of analytes measured across the study.

(C) Circos plot of multi-omic data points from selected assays. The transcriptome, both targeted and untargeted proteome, and serum cytokine levels are plotted according to their genomic location as well as the average expression in IR and IS participants (see inset labels).

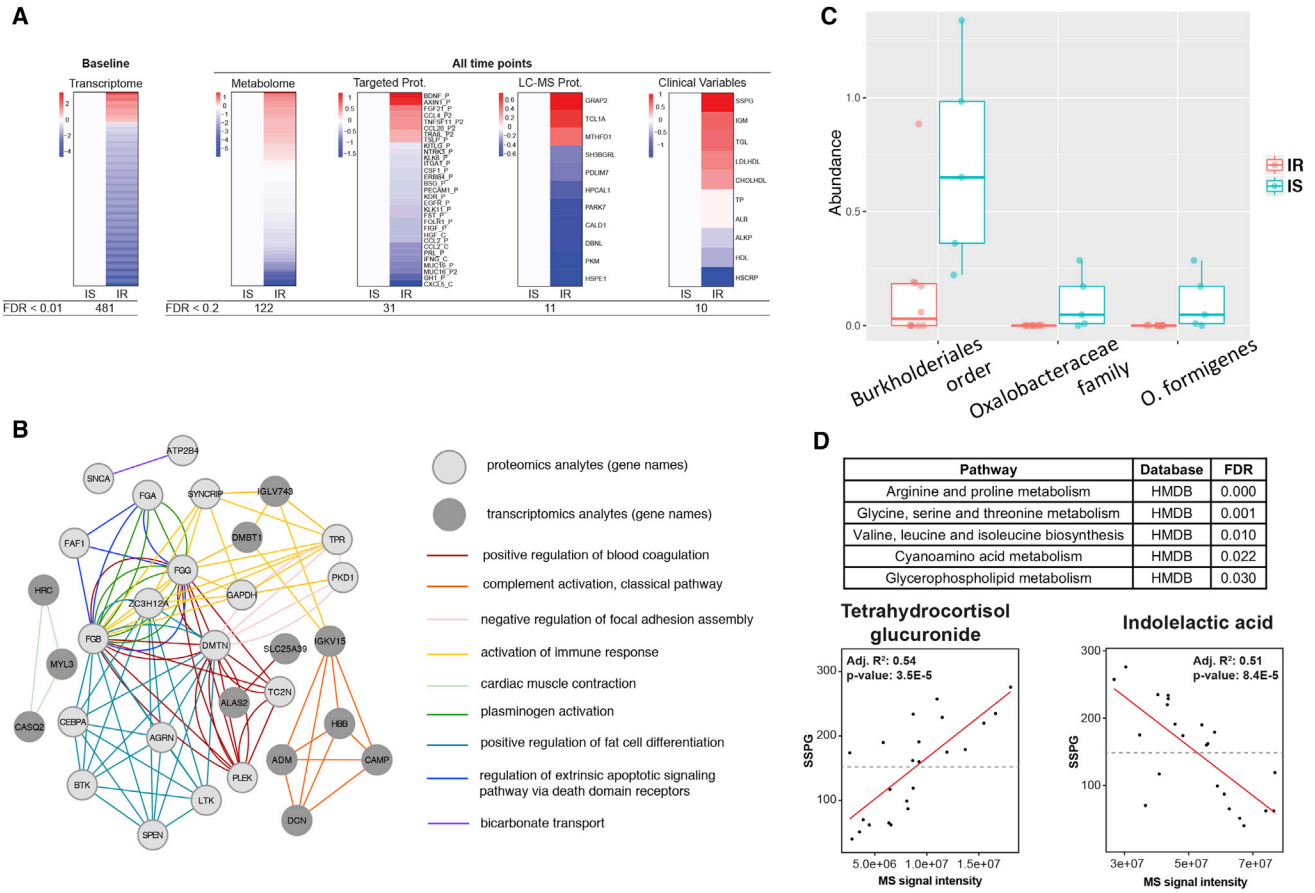
(D) The stool microbiome phylogenetic tree is visualized by GraPhlAn for taxonomies present across all participants along with the respective relative abundance in IR and IS (outer layers).

assays performed at all time points were: genomics (germline exome sequencing from whole blood, performed once for each individual), transcriptomics (RNA sequencing [RNA-seq] from blood PBMCs performed over the time course), proteomics performed three different orthogonal ways (untargeted liquid chromatography [LC] mass spectrometry [MS] from blood PBMCs, 276 targeted protein assays from plasma [Proseek multiplex], and a 63-protein cytokine/chemokine/adipokine panel from serum), metabolomics (untargeted LC-MS from plasma using a pipeline we recently developed; Contrepois et al., 2015) and microbiomics done using two different methods (16S and whole-metagenome sequencing from stool) (Figures 1B and 1C). In total, across all time points, over 2 million unique data points were measured.

Differences in Omics Profiles and Biochemical Pathways between IR and IS Participants

We first examined whether baseline differences were detectable in omics profiles between insulin-sensitive and insulin-resistant individuals; such differences may influence how individuals respond to the dietary perturbation. In order to identify any individuals that may harbor germline mutations in known metabolic disease susceptibility genes as well as enable accurate mapping of the different omics data, exome sequencing was performed on each subject followed by reconstruction of their personal exome, which was used for improved transcriptome mapping. A detailed analysis of the genomic risk of the subjects is reported elsewhere (Rego et al., 2017); however, we did not find any known high-risk variants for diabetes or insulin resistance (e.g., MODY [maturity-onset diabetes of the young] mutations), thus allowing us to study insulin resistance in a population without elevated Mendelian genetic risk. As expected, the overall density of variants was consistent from participant to participant (Figure S1).

At baseline, modest differences were evident between IR and IS groups. In blood PBMCs, significant transcripts differentially expressed between IR and IS (Figure 2A and Table S2) comprised a number of common significant functional groups, including pathways associated with the cardiovascular system (false discovery rate [FDR] < 0.0001), angiogenesis



Pathway	Database	FDR
Arginine and proline metabolism	HMDB	0.000
Glycine, serine and threonine metabolism	HMDB	0.001
Valine, leucine and isoleucine biosynthesis	HMDB	0.010
Cyanoamino acid metabolism	HMDB	0.022
Glycerophospholipid metabolism	HMDB	0.030

Figure 2. Differences between IR and IS Participants at Baseline

(A) Heatmap showing differences between IR and IS in baseline molecular abundance for each omic type. Each analyte is normalized according to the average expression in IS and significant differences in the IR group are plotted (red, upregulated in IR; blue, downregulated in IR).

(B) Pathways exhibiting significant transcriptomic and proteomics differences between IR and IS. The top Gene Ontology categories are presented and top transcripts and proteins are plotted in a network diagram showing pathway connections.

(C) Differences in microbial abundance (%) between IR and IS by both 16S and shotgun metagenomic sequencing.

(D) Regression analysis detailing association of multiple metabolites with clinical steady-state-plasma glucose (SSPG). MS signal intensity is plotted versus SSPG (mg/dL) for the selected metabolites indolelactic acid and tetrahydrocortisol glucuronide. Inset are the R^2 and p values for the selected comparisons. HMDB, Human Metabolome Database.

(FDR < 0.04), and actin/myosin cell motility (FDR < 0.00001), as well as multiple pathways representing an immune/inflammatory response (FDR < 0.05) (Figure S2 and Table S2). To explore whether these expression differences translated to the proteome level, we performed the same comparison in LC-MS proteomic profiles of the same PBMC population. Baseline analysis revealed three proteins (TC2N, DMTN, and PKD1) that were different between IR and IS participants; the smaller number of differential proteins versus RNA is likely due to the size of the cohort and potentially higher variability of protein levels. Indeed, expanding our analysis to the pathway level, we observed significantly enriched pathways that closely mirrored those uncovered in the transcriptome as well as those that were unique to the proteome (fat cell differentiation, etc.) (Table S2). A combined ranked pathway analysis of transcriptome and proteome from PBMCs further confirmed the enrichment of these core pathways (Figure 2B), as well as uncovering additional enriched pathways that spanned RNA and proteins (platelet/blood coagulation, plasminogen activation). Despite the increased immune/inflammatory responses observed here in IR participants, we note that all participants self-identified as healthy (no common colds, etc.) at the time of sampling, so it was particularly striking that participants exhibited this phenotype at baseline.

Using both 16S and shotgun metagenomics of the stool microbiome, we observed significant differences between IR and IS participants in the abundance of the gram-negative proteobacterium *Oxalobacter formigenes* ($p < 0.006$; Figure 2C and Table S2). Interestingly, although this bacterium was present at relatively high levels in IS participants, it was not detected in any of the IR participants' samples. *O. formigenes* is particularly unusual in that it processes oxalate, and absence of this bacterium is associated with increased risk of kidney stones (Duncan et al., 2002), which was linked to diabetes and insulin resistance (Chung et al., 2011; Daudon et al., 2006) and can be sensitive to high-oxalate-containing foods such as almonds (Haaskjold et al.,

2015). While we only observed significant IR associations with one microbe at baseline, we hypothesized that this may be due to the small sample size of the cohort. As such, we expanded the analysis to include all time points, revealing a substantially different picture. Both 16S and shotgun metagenomics identified that bacteria of the *Alistipes* genus in the Rikenellaceae family were more abundant in IS than IR subjects ($p < 0.04$). *Alistipes* spp. were shown previously to associate with glucose regulation, diabetes, and obesity (Brown et al., 2011; Ridaura et al., 2013; Serino et al., 2012). Our results demonstrate that many important biological molecules, pathways, and microorganisms differ in IR and IS patients. Although some of these differences were known previously in diabetics or obese individuals relative to healthy people, our results demonstrate that these differences are already present in earlier stages/aspects of the disease.

Expanding the analysis to include all time points also revealed IR-associated signatures in other omic data types. In plasma, we observed an IR-associated proteomic signature spanning multiple targeted assays (Figure 2A and Table S2). Among these was the folate receptor FOLR1 (FDR < 0.004), the hormone prolactin (FDR < 0.007), and brain-derived neurotrophic factor (BDNF) (FDR < 0.03). BDNF has been shown to be one of the main signaling factors for appetite as well as heart function (Feng et al., 2015; Fulgenzi et al., 2015; Rosas-Vargas et al., 2011). The former is an interesting target as folate supplementation has been explored as a method to combat metabolic-disease-associated cardiovascular disease (van Etten et al., 2002) and prolactin levels have been shown to be inversely associated with diabetes (Wang et al., 2013). Along with targeted plasma proteomic assays, clinical blood panels showed ten analytes that were significantly different between IS and IR participants (Table S2, FDR < 0.2). Most of these clinical variables were associated with dysregulated lipid metabolism (i.e., triglycerides (TGLs), low-density lipoprotein [LDL]/high-density lipoprotein [HDL] ratio, etc.) and inflammation (high-sensitivity C-reactive protein [hs-CRP]), consistent with results from PBMCs and plasma (Figure 2A).

Untargeted metabolomic profiling of plasma also led to the identification of 122 metabolites that were different between IR and IS participants (FDR < 0.2) (Figure 2A and Table S2). While the SSPG cutoff for defining IR and IS has precedent clinically, we asked whether different thresholds for defining IR and IS led to the identification of different numbers of metabolites. We observed that changes to the SSPG cutoff (from 100 to 170) did not significantly change the number of differential metabolites in IR versus IS participants (Figure S2), suggesting that there is some consistency in metabolic differences across a relatively wide range of insulin resistance measurements.

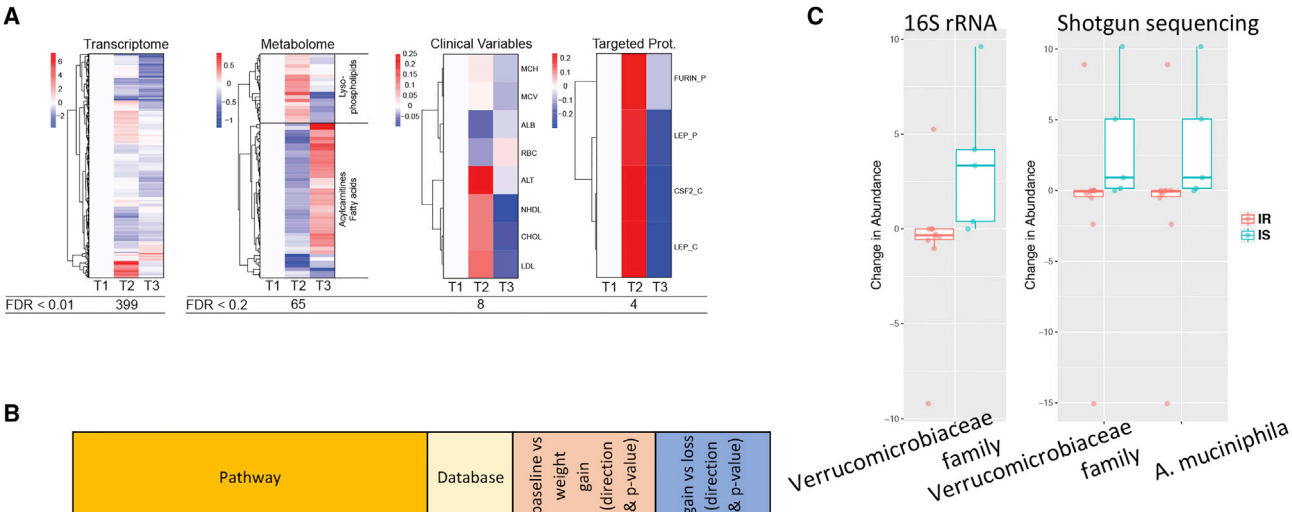
Pathway analysis revealed that amino acid as well as lipid metabolism were deregulated in IR individuals compared with healthy controls (FDR < 0.05) (Figure 2). These observations are consistent with the observation that compromised insulin action is associated with altered intermediary metabolism of fats and amino acids (Adams, 2011). In particular, we found that the branched-chain amino acid (BCAA) biosynthetic pathway was deregulated in insulin-resistant participants (FDR = 0.01), which is consistent with previous findings (Pedersen et al., 2016; Yoon, 2016). In addition to BCAAs, many more amino acids, such as sulfur-containing amino acids (e.g., methionine), were found in higher levels in the plasma of insulin-resistant individuals (Figure S2). Lipid metabolism was also altered with the increased level of many short- and medium-chain acylcarnitines, and the reduction of phospholipid and plasmalogen levels (Figure S2). Accumulation of acylcarnitines is commonly observed in cases of insulin resistance and T2DM (Schooneman et al., 2013).

For highly significant metabolites, we also tested how well correlated their plasma abundances were with clinical insulin resistance measures (e.g., SSPG), thus indicating their potential as a biomarker for insulin resistance. Indeed, tetrahydrocortisol glucuronide was strongly positively associated with baseline SSPG levels and indolelactic acid was negatively associated with SSPG levels (Figure 2D). Interestingly, indolelactic acid has recently been associated with insulin resistance (Pedersen et al., 2016). Tetrahydrocortisol glucuronide is a modified, more water-soluble version of tetrahydrocortisol, which allows elimination from the body and/or ease of transportation around the body. To the best of our knowledge, this is the first report of an association between insulin resistance and tetrahydrocortisol glucuronide, and, based on these data, both metabolites show promise as IR biomarkers.

We next attempted to use the metabolome to predict SSPG values. We performed binary classification using ensemble learning as well as quantitative prediction using delta SSPG values based on delta metabolomics features. Using random forest and AdaBoost classifications, we achieved an accuracy rate of 87.5% on our hold-out test set (see STAR Methods). Predictive performance measures in terms of precision, recall, and F1 score were 87% for the random forest method and 78% for AdaBoost (Figure S2). The high accuracy and precision show that the model was able to correctly predict the future SSPG trend by only looking at the difference between the current and future metabolomics features. From both methods, only one wrong prediction was made, which was a false-positive (see confusion matrix, Figure S2). For the regression model to compute Δ_{SSPG} at time (t^n), we used LASSO (least absolute shrinkage and selection operator) and elastic net regularized regression, with 10-fold cross-validation. We achieved a low root-mean-square error (RMSE) of 27.5110 for the elastic net model. Regression curves (Figure S2) show that the predicted $\Delta_{SSPG}(t^n)$ are very close to the actual values for most samples. In conclusion, metabolomics profiling has significant potential as a predictor for changes in insulin resistance.

Omics Profiles and Pathways Associated with Weight Gain

Following characterization of group differences, we next examined the biomolecules and pathways that were specifically responsive to the 30-day weight gain perturbation (Figure 3). Comparing omics profiles at peak weight with baseline profiles, we observed a number of significant changes across multiple omes. Specifically, at the transcriptome level, we observed 318 transcripts that were significantly differentially expressed between baseline and peak weight (FDR < 0.01; Table S3). As expected, this included an increase in expression for a number of genes associated with lipid metabolism, such as lipoprotein lipase (*LPL*; FDR < 5×10^{-5}). Pathway enrichment analysis



B

Pathway	Database	baseline vs weight gain (direction & p-value)		gain vs loss (direction & p-value)	
Actin filament-based process	GO	↑	3.1E-15	↓	1.4E-18
Dilated cardiomyopathy	KEGG	↑	3.4E-07	↓	1.2E-07
Response to lipid	GO	↑	2.4E-08	↓	7.8E-03
Response to lipopolysaccharide	GO	↑↓	1.5E-09	↓↑	6.1E-02
Inflammatory response	GO	↑↓	8.1E-08	↓↑	1.1E-03
Response to external biotic stimulus	GO	↑↓	2.0E-10	↓↑	3.9E-04
Response to cytokine	GO	↑↓	1.1E-06	↓↑	8.2E-05
Platelet activation	GO	↑↓	2.9E-07		
Positive regulation of ATPase activity	GO			↑↓	2.3E-02

Figure 3. Multi-omic Differences over the Course of a Weight Gain and Loss Perturbation Experiment

(A) Heatmap showing analytes that vary in abundance in response to the weight gain and loss perturbation.

(B) Pathways that are significantly different between baseline versus weight gain, and weight gain versus weight loss, respectively. GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

(C) IR- or IS-specific microbiome changes are shown for selected taxonomic units confirmed by both methods (16S and shotgun metagenomics).

showed that these genes comprised a number of common functional categories (Figure 3B), including a large number of inflammatory response genes and pathways at peak weight; this may indicate significant stress response despite the relatively modest weight gain (average 2.4 kg). Importantly, we also observed an increase in expression of genes associated with dilated cardiomyopathy, potentially reflecting a molecular explanation for the association between weight gain and heart failure (Dela Cruz and Matthay, 2009).

At the microbiome level, 16S sequencing revealed a significant increase in microbes of the Verrucomicrobiaceae family in response to weight gain; using shotgun metagenomics, we were able to further specify this response to the species *Akkermansia muciniphila* ($p < 0.03$; Figure 3C and Table S3). Interestingly, this response was only observed in the insulin-sensitive participants; this is of particular note in light of prior studies showing that *A. muciniphila* confers a protective effect against insulin resistance in response to weight gain in animal models (Everard et al., 2013; Roopchand et al., 2015). Overall, these results indicate that there are substantial biological pathways that change during weight gain that affect immune response, heart function, biochemistry, and microbiome.

Omics Profiles and Pathways Associated with Weight Loss

As short-term weight gain induced a number of significant blood-based and microbiome-based changes, we next determined whether these responses would persist or revert upon each participant's return to baseline weight. We compared omics profiles assayed after weight loss with those of the previous peak weight sample and again observed a number of significant changes across multiple omes (Figure 3). At the transcriptome level, 213 genes were significantly differentially expressed in weight loss versus peak weight (Figure 3A and Table S3). Pathway enrichment analyses strikingly revealed that the majority of the significant pathways observed after weight gain reversed after subsequent weight loss (Figure 3B); examples include the genes associated with lipid metabolism and inflammation. This reversal upon weight loss is of particular importance as it suggests that the negative effects of short-term weight gain can potentially be ameliorated by corrective action.

Mirroring the dynamic changes of RNA levels in PBMCs, most of the 65 plasma metabolites that changed upon weight gain reversed after weight loss (Figure 3A and Table S3; FDR < 0.2). Interestingly, most of these molecules function in lipid metabolism, including acylcarnitines and fatty acids, showing that a subtle change in weight profoundly affects lipid metabolic pathways. We specifically examined associations with BMI, and identified 133 metabolites that were significantly associated with changes in BMI (Table S3). Similar to the time-points model, most changing molecules belong to lipid metabolism and include many acylcarnitines, fatty acids, and lysophospholipids (Figure S3), possibly reflecting decreased catabolism due to an increased calorie intake. We also had the opportunity to validate

these results in a separate, unpublished weight gain cohort conducted in Sweden with blood LC-MS metabolomics generated using the Metabolon platform. We examined the subset of metabolites that were positively identified by both platforms and asked whether they behaved similarly across both cohorts. Of these, we found that 77% of weight-responsive metabolites were also enriched upon weight change in the second study (Table S4). Thus, despite differences in study populations and metabolomic analytical platforms, weight-responsive metabolites were reproducible across studies.

Targeted proteomic assays (Luminex and Proseek) revealed 27 proteins (FDR < 0.2) associated with BMI change, with the adipokine leptin ($p < 8 \times 10^{-5}$) positively associated with BMI change (Figure S3 and Table S3) with a substantial 30% increase in leptin levels per unit of BMI; this is consistent with previous cross-sectional studies finding increased leptin levels in obese subjects (Considine et al., 1996). Novel significant responses to changes in BMI included the immuno-modulating covertase furin ($p < 8 \times 10^{-5}$) (Pesu et al., 2008), as well as lipopolysaccharide (LPS)-induced tumor necrosis factor (LITAF) ($p < 4 \times 10^{-5}$). The latter is of particular note considering that LPS-responsive immune/inflammatory pathways also exhibited a significant response in the PBMC transcriptome. As LITAF- and LPS-responsive pathways are known to respond to endotoxins of microbial origin, this response may indicate crosstalk between the microbiome and the human host's immune system in response to weight gain/loss. Consistent with this interpretation was the association of white blood cell counts with delta BMI. Also associated with delta BMI were non-HDL-cholesterol ($p < 0.03$) and LDL-cholesterol ($p < 0.02$) measured as part of the clinical panel (Table S3). In summary, these results indicate that most changes identified during weight gain reverse upon weight loss.

Uncovering Trends in Response to Dietary Change

While the prior analysis provides insight into biomolecules that vary along with changing weight/BMI, it is possible that more complex patterns may be evident across the perturbation. Specifically, we were interested in (1) analytes that may have changed in response to weight gain but did not revert upon subsequent weight loss (and indicative of a long-term effect) or (2) molecules that may not have changed upon weight gain but were only responsive to weight loss. To better understand the breadth of possible responses, we performed longitudinal pattern recognition analysis across blood-based analytes (transcriptome, proteome, and metabolome) using fuzzy c-means clustering. From these data, we observed diverse response patterns to the perturbation (Figure 4A).

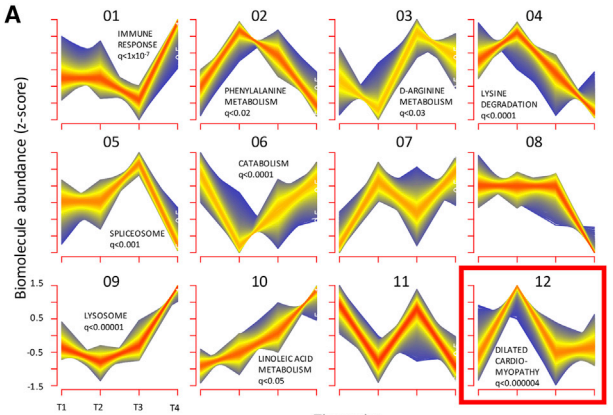
From these analyses, a number of key patterns were evident. For the pattern observed in cluster 12 (Figure 4A), which comprised an initial increase in biomolecular abundance followed by a decrease back to baseline that persisted through the quarterly follow-up visit, and most closely resembled the weight gain/loss trajectory each participant exhibited (Figure S1), we observed a highly significant enrichment for genes and proteins associated with risk for dilated cardiomyopathy (FDR < 5×10^{-6}, Kyoto Encyclopedia of Genes and Genomes [KEGG] enrichment via the Database for Annotation, Visualization and Integrated Discovery [DAVID] algorithm; Huang da et al., 2009a, 2009b). As shown by the KEGG pathway diagram (Figure 4B), this enrichment included a large number of members of this functional category spanning the cell surface receptors to the internal machinery. The overfeeding-induced changes in these analytes may indicate ill-effects of the short-term weight gain, such as increased coronary artery disease; however, it is important to note that this biomolecular signature returned to baseline upon subsequent weight loss. Other intriguing responses included sets of biomolecules that were anticorrelated with weight gain. A cluster genes/proteins and metabolites comprising catabolic pathways decreased upon weight gain and returned to baseline upon subsequent weight loss (Figure 4A, cluster 6). A downregulation of catabolism is logical in an overfed state as the body may switch to a largely anabolic state with an overabundance of energy intake from food.

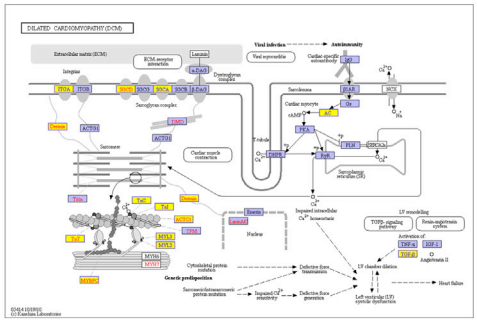
Of particular note was a cluster of biomolecules that increased upon weight gain but failed to return to baseline at the end of the weight loss period (Figure S4). Enrichment testing showed that cluster 2 comprised biomolecules associated with a variety of metabolic pathways, including folate metabolism ($p < 0.001$), phenylalanine metabolism ($p < 0.005$), and BCAA degradation ($p < 0.03$). Thus, these pathways have long-lasting effects after weight gain.

In order to identify biomolecular trends across the dataset in a more unbiased fashion, we performed weighted gene-co-expression analysis (Zhang and Horvath, 2005) across all conditions and time points (Figure S4). From this analysis, a series of core gene expression patterns emerged, and the co-expressed genes were functionally related based on pathway enrichment. We next asked whether these common expression patterns were significantly associated with any of the clinical variables tested and thus may cause clinical phenotypes. A number of these significant associations emerged (Figure 4C). For example, a gene module significantly enriched for mitochondrial genes involved in oxidative phosphorylation (FDR < 1×10^{-12}) was significantly associated with multiple clinical parameters, including glycated hemoglobin A1C (HbA1C), hs-CRP, and LDL-cholesterol (Figure 4D); this may suggest a link between mitochondrial activity and diabetes-/cardiovascular-related issues, thus showing that novel associations can be gleaned from the unsupervised associations.

Associations between Omic Measures and the Microbiome

In order to better understand how the microbiome may interact with the changes we observed in human molecular physiology over the course of the dietary perturbation, we explored (1) how microbes co-vary with each other across the entire dataset and (2) how microbes co-vary with metabolites in the host. Some microbes showed a high degree of covariance in both IR and IS subjects over the course of the perturbation experiment; for example, the proportion of *Bacteroides dorei* was positively correlated with *Alistipes putredinis* (Figure 5A, highlighted by squares). *B. dorei* has been shown to modulate immune responses in humans (Vatanen et al., 2016) and *A. putredinis* has been shown to be significantly responsive to dietary changes (David et al., 2014), thus painting a complex portrait of how diet-induced changes to the microbiome may relate to the immune changes we observe in the host. In addition to these





Cluster enrichment: Dilated cardiomyopathy (q=4x10^{-6})

C

Module	Major biological pathway	Clinical Parameter						
		A1C	HDL	LDL	Hs-CRP	IGM	PLT	TBIL
A	insulin signaling (q<0.006)		1.3E-02		1.6E-03	1.7E-03		3.3E-03
B	inflammatory response (q<0.05)		3.9E-03		3.0E-04	9.3E-04		5.7E-03
C	dilated cardiomyopathy (q<1x10^{-7})		6.1E-03	4.7E-03				4.2E-02
D	OXPHOS (q<1x10^{-12})	6.2E-03	1.1E-02	1.8E-04	1.4E-03	1.7E-02		3.7E-06
E	platelet activation (q<4x10^{-11})	3.6E-02			2.5E-04		1.7E-03	4.7E-02
F	adaptive immune response (q<0.06)	1.5E-02	5.1E-04					9.1E-03
G	immune response (q<1x10^{-21})	6.1E-02			2.0E-03			3.9E-04
H	histone demethylation (q<0.004)		2.6E-03	1.0E-02	1.6E-05	1.7E-04		1.0E-05

D

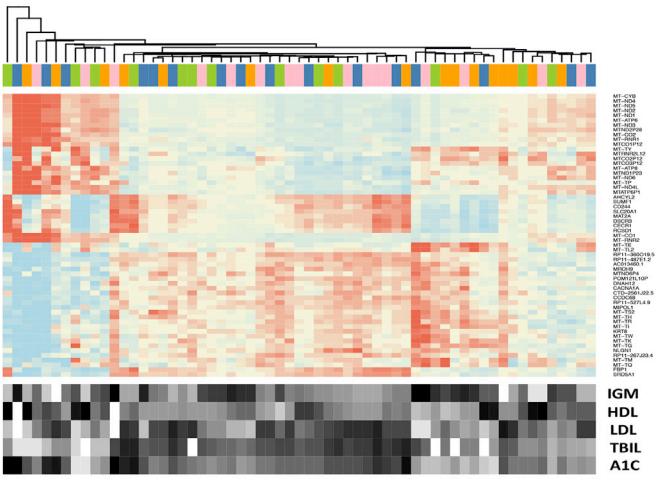


Figure 4. Multiparametric and Trend Analyses Reveal Novel Responses to Weight Gain and Loss

(A) Longitudinal pattern recognition using fuzzy c-means clustering across all host omes. Data from the transcriptome, proteome, cytokines, and metabolites were standardized to *Z* scores for each analyte and subjected to c-means clustering across all four time points. Each subplot shows a unique cluster and the trend for all analytes comprising the cluster. The red outline indicates the cluster that is featured in the pathway analysis in (B).

(B) KEGG pathway diagram for analytes implicated in dilated cardiomyopathy, a pathway that was significantly enriched in cluster 12 (FDR < 0.000004). Elements highlighted in yellow indicate the pathway analytes that comprise cluster 12.

(C) Table showing biological pathway enrichment and association with clinical blood panel analytes for key gene co-expression clusters. A1C, HbA1C; IGL, immunoglobulin lambda; OXPHOS, oxidative phosphorylation; PLT, platelet count; TBIL, total bilirubin.

(D) Gene expression heatmap for transcripts comprising the yellow module from Figure S3. The expression for each gene is shown for all time points (T1, pink; T2, blue; T3, orange; T4, green) along with the relative levels for each of the enriched clinical parameters (A1C, LDL, HDL, immunoglobulin M [IGM], and bilirubin [TBIL]). Red indicates high expression while blue indicates low for genes; dark squares indicate high relative abundance for clinical parameters.

similarities, some microbiota show strikingly opposite correlations depending on IR and IS context. For example, *Eubacterium hallii* and *Parabacteroides* exhibit a strong positive correlation in IR participants across time points and a negative correlation in IS individuals. This is of particular interest, because *E. hallii* has been shown to metabolize glucose and is a major contributor of short-chain fatty acids in the gut (Engels et al., 2016), whereas *Parabacteroides* has previously been shown to be directly associated with weight gain (Lecomte et al., 2015). It is intriguing to speculate that the unique gut microenvironment in IR versus IS individuals contributes to these differing responses to host dietary intervention and weight changes. Additional examples are *Bacteroides vulgatus* and *Eubacterium eligens*, which exhibit a positive correlation in IR individuals and a negative correlation in IS individuals (Figure 5A, highlighted by circles). This is of particular note because *B. vulgatus* has been shown to be a mediator between BCAA metabolism and insulin resistance (Pedersen et al., 2016), whereas *E. eligens* has been shown to respond significantly to dietary fiber (Chung et al., 2016). Furthermore, most pairwise correlations were significant only in IR or IS individuals, again showing a potential difference where IR individuals may possess a unique microbiome that possibly contributes to the pathology or progression of their disease.

The contribution of microbial products to host metabolic signaling is one of the fundamental mechanisms underlying the host-microbiome interactions (Holmes et al., 2011). In order to better understand how these microbiome alterations are associated with host metabolism, we examined whether specific microbes correlated with host metabolites. Associations between microbial populations at the phylum level (with at least 1% abundance) and metabolite quantities revealed 26 cross-omics associations significantly different between IR and IS ($p < 0.1$ for the interaction term in the ANOVA model), of which eight showed significant associations only in IR or IS individuals (Table S3). As an example of these, we found that the antioxidant 3-indolepropionic acid was positively associated with Proteobacteria but specifically in the IR subjects. As a second example, N6-trimethyllysine was significantly positively associated with the phylum Proteobacteria exclusively in the IS participants (Figure 5B). It is worth noting that 3-indolepropionic acid is a microbial fermentation

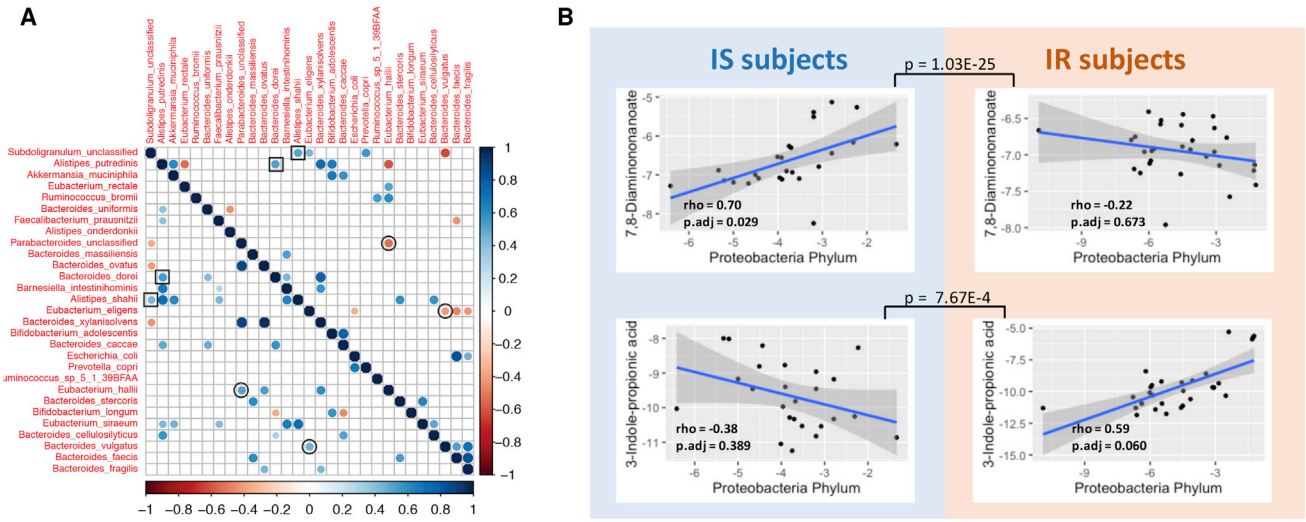


Figure 5. Associations of Analytes across IR and IS and across Omes

(A) Co-varying microbial species are plotted based on whether they are co- or inversely associated (blue or red, respectively), and whether this occurs in IR (upper quadrant) or IS (lower quadrant). Squares indicate *A. putredinis*. Circled values indicate relationships that are the opposite between IR and IS, while squared circles indicate the same trend across IR and IS.

(B) Co-variation of microbes and metabolites for IR and IS is plotted for selected associations. Inset are the Spearman's rho and adjusted p values after FDR correction for the selected associations. Also, adjusted p values by FDR are shown between IS and IR individuals for the interaction term of the linear model describing different trends in the two groups.

product from tryptophan, and it can be a potential drug target for the management of insulin resistance (Khan et al., 2014). In the case of N6-trimethyllysine, it is a precursor for L-carnitine biosynthesis, which improves glucose tolerance, increases total energy expenditure in obesity (Flanagan et al., 2010), and decreases cardiovascular risks (Koeth et al., 2013; Ussher et al., 2013). These results demonstrate that microbial differences exist between IR and IS participants and these are associated with differences in metabolism in the host.

Each Individual Has Unique Biomolecular Profiles, and Molecular Variation Can Come from Distinct Sources

Although much of the prior analyses focused on the similarities between the individuals across the dietary perturbation, with such a diverse set of multi-omic measurements we also had the unique opportunity to describe in greater detail the differences that exist among individuals both at steady state and through a longitudinal perturbation. The extensive measurements across different people and perturbations allowed us to use ANOVA to decompose the observed variance for each analyte into components originating from inter-personal differences, the experimental perturbation (weight gain, weight loss, IR versus IS) and other parameters (technical noise and unknown sources). This was examined for individual molecules as well as the general type of molecule (RNA, cytokine, microbiome) (Figures 6A and S5). Strikingly, we found that all molecular measurements were dominated by inter-personal variation, which accounted for more than 90% of the observed variance in some cases (e.g., cytokines). On the opposite side of the spectrum, proteomics and metabolomics measurements had a substantial unexplained component (30% and 35%, respectively), highlighting the presence of unaccounted factors (e.g., food, exercise, and other changing environmental factors) or a subject-specific reaction to the perturbation. We then examined how cytokine levels differed between individuals versus within multiple time points for an individual (Figure 6B). To exclude the effects of the dietary perturbation for this exercise, we performed this analysis in the set of participants that provided follow-up time points 3 months after the end of the perturbation study (from our prior analyses we observed that the majority of the effects of the weight perturbation have dissipated at this follow-up time point; Figure 4B). Comparing the variation in cytokine levels between multiple baselines in a single individual versus across individuals, we observed a striking difference: for almost all cytokines, the within-individual coefficient of variation was under 20%, whereas the variation across individuals was 40%–60%. This shows that our baseline cytokine profiles are unique to the individual, a point that has significant implications for one-size-fits-all clinical cytokine assays for the detection and/or monitoring of disease. We observed similar (but less dramatic) effects for the other omes (Figure S5).

The Power of Longitudinal Analyses

As personal variation proved to be one of the main sources of variation across these datasets, we next quantified to what degree the longitudinal study design (i.e., comparing each individual with himself/herself across time) increased our ability to identify biomolecular responses to weight gain and loss. To do this, we performed a power analysis using the metabolome data as an example (Figure 6C). By comparing our analysis, which accounts for the personal baseline, with a regular group comparison using a t test, we can compute the population size required to detect the same effect size with the same statistical power as found in our study. For the 127 significant pairwise

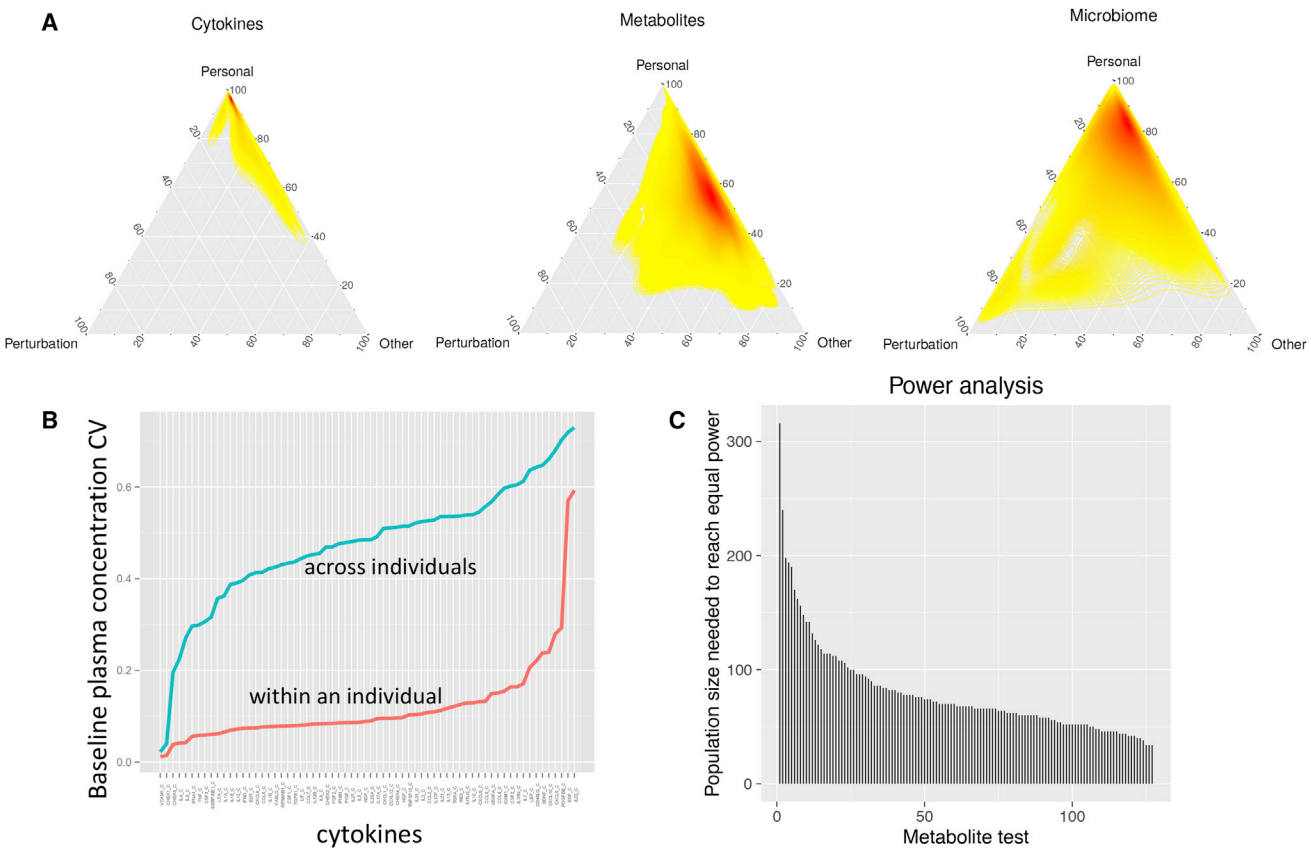


Figure 6. Personal Variation of Omics Data

(A) Variance decomposition analysis of selected omes (see Figure S5 for others). The variance across all time points was deconvolved into experiment-dependent variation (i.e., due to the perturbation), personal variation (within an individual), or other types of variation (technical or unknown sources). The heatmap color (yellow to red) indicates the density of analytes at each coordinate.

(B) Variation in cytokine/chemokine/adipokine abundance within participants versus across participants. The coefficient of variation (CV) for all measured Luminex immunoassays is plotted across steady-state time points (T1 and T4) within an individual (red) and across individuals (blue).

(C) Power comparison for longitudinal versus groupwise study designs using metabolomics data as an example.

differences ($p < 0.05$) across 73 metabolites, we find that a cross-sectional setup would, on average, require 79 participants to reach the same power, with some analytes requiring 300 participants or more in a cross-sectional setup. Therefore, by correcting for the personal baseline of each analyte we were able to improve power in every case, for some analytes quite dramatically. This is consistent with the fact that for 85.6% of all metabolites we found significant (q-value < 0.2) differences evident in baseline samples that also persisted throughout the perturbation. These findings highlight the fact that each individual is biochemically unique, and this stresses the need for personalized analysis in medicine.

DISCUSSION

Obesity and T2DM are progressive disorders in which our key to a deep understanding of the etiology likely will come not from single observations of a limited number of analytes but from deep analyses and longitudinal profiling. Here, we have made millions of measurements of humans and their microbiomes across a longitudinal perturbation and identified some striking patterns within and between subjects. First, by taking an integrated multi-omic approach (Figure 1), we find marked differences between IR and IS individuals. In the fasted blood of IR individuals, we observed differential regulation of inflammatory/immune response pathways (Figures 2A and 2B). This is consistent with prior literature (Festa et al., 2000; Mardi et al., 2005); however, earlier studies used only a few analytical markers, limiting the breadth of information that can be gleaned. Using deep multi-omic profiling, we show that this response includes dysregulation of a large number of factors that specifically function in an antimicrobial response (*LTF*, *CAMP*, as well as various defensins), and was reflected in each of the different omics profiles (PBMC transcriptome, proteome, and circulating cytokines).

In addition, many circulating metabolites were present at different levels in IR and IS participants' amino acid (BCAAs, sulfur-containing amino acids) and lipid metabolism (acylcarnitines and ether-linked phosphatidylcholines) (Figure 2A). BCAAs have been most consistently described in the literature as associated with insulin resistance. In addition, indolelactic acid presented a very strong negative correlation with SSPG levels

(Figure 2D). This molecule is of particular interest as it is produced specifically by the microbiome in humans (Patten et al., 2013; Wikoff et al., 2009). Moreover, we show that metabolomics profiles can be used as an accurate predictor for changes in insulin resistance, and it will be of significant interest to apply this approach across significantly larger cohorts. It is possible that a metabolomic signature or a subset of metabolites could be used as a clinical assessment of the insulin sensitivity level in place of the expensive and time-consuming procedure that measures SSPG levels.

Concomitant with these observations is the differential relative abundance of several bacteria in the gut microbiome of IR compared with IS individuals, including gram-negative proteobacterium *O. formigenes* and several Firmicutes species (e.g., the Lachnospiraceae *Blautia* and *Dorea* at >1% abundance level) (Figure 2C and Table S2). Increases in the relative proportion of Firmicutes have been shown to be associated with obesity in numerous contexts (Ley et al., 2006; Turnbaugh et al., 2006, 2009), and in mice there is evidence that low-grade inflammation associated with weight gain is at least partially due to the microbiome (Cani et al., 2008). While this association has been largely attributed to the metabolite LPS produced by gram-negative bacteria, it is interesting to note that increases in gram-positive Firmicutes correlate with increased inflammation in this study, raising the possibility of other non-LPS triggers of a low-level systemic immune response in overweight/obese humans. Interestingly, a prior study showed that Firmicutes abundance was reduced in patients with T2DM relative to healthy controls (Larsen et al., 2010), so it is possible that, upon progression from a prediabetic state to outright diabetes, a remodeling of the microbiome occurs, again emphasizing the need for detailed longitudinal omic monitoring of patients at risk for T2DM in order to better understand the events that precede the development of frank T2DM.

We also found extensive molecular changes after weight gain and weight loss. Notably the inflammation response was one of the major pathways induced upon weight gain; similar results have been found previously for fat cells (Nishimura et al., 2009). This dysregulation is evident at several different levels, including transcriptome, proteome, and cytokines (Figures 3 and 4). These results suggest that a systemic inflammatory pathway is activated in response to short-term weight gain, which is surprising given the modest weight gain induced here (~2–3 kg); however, it is important to note that this response was largely reversed upon subsequent weight loss (Figure 3).

Of particular interest from the global pathway analysis was the discovery that the dilated cardiomyopathy (DCM) pathway is activated upon weight gain (Figure 4). Since heart conditions are associated with increased weight, these results provide a potential biochemical explanation for this response; indeed, many of the associated biochemical pathways observed herein, including inflammation and oxidative stress, have been previously implicated as causative for DCM (Tiwari and Ndisang, 2014). Interestingly, we also observed activation of the platelet plug formation pathway, which was also altered during diabetes onset (Chen et al., 2012). This pathway may be an indication of altered metabolism and possibly play a role in signaling and/or preparation for blood clotting response during loss of metabolic control.

It is important to note that not all of the responses we observed were consistent across IR and IS participants. In particular, for the microbiome, we observed that the microbe *A. muciniphila* was weight gain responsive only in insulin-sensitive participants (Figure 3C). The abundance of this particular microbe in IR individuals did not change across perturbations and was barely or not detectable in most IR individuals. While prior studies have also shown that gut colonization by *A. muciniphila* is associated with protection from metabolic disease in mice and humans (Dao et al., 2016; Everard et al., 2013; Roopchand et al., 2015), here we show in humans that *A. muciniphila* is responsive to even a modest weight gain and clearly differentiates between IR and IS participants. In addition, we observed a subset of microbes that are synchronized in response to the dietary perturbation but that are different between IR and IS individuals (Figure 5A). Furthermore, we discovered longitudinal associations between microbes and host metabolites that are significantly different between the two groups (Figures 5B and 5C). Taken together, microbial composition and their metabolic activities may be one explanation as to why some individuals respond metabolically very poorly to weight gain while others do not; as such, we hypothesize that ensuring the presence of certain classes of microbes in the gut could be a key target for diabetes prevention.

Our study also allowed us to investigate the relation between inter-individual variation and intra-individual variation over time and during a perturbation. Our ANOVA components showed that, for all omics measurements, the inter-individual variation dominated the intra-individual variation and was the main explanatory factor for different analyte levels (inter-individual variation ranged from 93% for cytokines to 54% for transcripts). Our data highlight how personalized analysis can provide additional insight and improve statistical power several fold. Even within different types of analytes (metabolites, cytokines, etc.), we observed that subsets exhibit highly personal variability while others showed common trends in response to the perturbation. In addition, the study participants have agreed to participate in long-term follow-up with regular sampling, thus it will eventually be possible to track individual trajectories of specific biomolecules over multi-year periods, further adding to our understanding of personal versus universal changes in biomolecular abundance.

While patterns emerged that implicate common pathways found in the PBMC transcriptome, plasma proteome and metabolome, and gut microbiome, it is important to note that, despite performing many high-throughput biochemical assays across multiple biospecimen types from the same patient, we are likely a long way from a complete omic representation of complex human biochemical systems. For example, the PBMC cell types profiled here represent only a fraction of the myriad cell types and tissues that actively secrete free proteins and metabolites into the plasma. Many of these tissues are inaccessible in a non-invasive fashion, limiting the breadth of omic profiling in humans versus similar model organism studies. To this end, mechanistic models derived from mouse or other studies could be leveraged as a template for mapping quantitative human omic data, potentially leading to more accurate predictive signatures for T2DM or other human diseases.

In addition to demonstrating the dynamics of extensive molecular changes during weight gain and loss, this study provides a

unique resource for the scientific community. Nearly all of the data are publicly available, enabling exploration of inter-omic relationships and alterations across a longitudinal perturbation. With the many levels of linked biological information available here, from genome, transcriptome, proteome, metabolome, and microbiome, all open access, we feel that this study may provide a valuable resource for the development and validation of bioinformatic tools and pipelines integrating disparate data types.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODELS AND SUBJECT DETAILS
  - Participant Recruitment and IRB Consent
- METHOD DETAILS
  - Sample Preparation - Blood Samples
  - Sample Preparation - Stool
  - Exome Sequencing
  - RNA-Seq
  - Microbiome Sequencing -16S
  - Microbiome Sequencing – Metagenome Shotgun
  - Untargeted Metabolomics by LC-MS
  - Additional Metabolomics Assays
  - Untargeted Proteomics by LC-MS
  - Plasma Protein Profiling Using Proseek Multiplex
  - Luminex Assays
- STATISTICAL ANALYSES
  - Power and Variance Calculations
  - Random Forest and AdaBoost Classification
  - RNA-Seq Data Processing and Analysis
  - Microbiome
  - Metabolomics
  - Untargeted Proteomics
  - Multivariate Data Analysis
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and six tables and can be found with this article online at https://doi.org/10.1016/j.cels.2017.12.013.

ACKNOWLEDGMENTS

Our work was supported by grants from the NIH Common Fund Human Microbiome Project (HMP) (1U54DE02378901) (M.P.S., G.W., and T.M.), American Diabetes Association (grants 1-14-TS-28 and 1-11-CT-35) (T.M.), National Heart, Lung, and Blood Institute (1T32HL098049) (B.D.P.), National Institute of Diabetes and Digestive and Kidney Diseases (1F32DK100072) (B.D.P.), as well as gifts from anonymous donors. H.L.R. was funded by SNSF (P2EZP3_162268) and is supported by EMBO (ALTF 854-2015). G.J.G.U is supported by Vetenskapsradet, Swedish Research Council (D0046401). The authors would like to thank the Stanford Human Immune Monitoring Center for performing immunoassays, the Stanford Center for Genomics and Personalized Medicine for sequencing services (grant 1S10OD020141-01), the Stanford Genetics Bioinformatics Service Center for computational and informatics support, the staff in the Stanford Clinical and Translational Research Unit, Lita M. Proctor, Jon LoTiempo, and Salvador Secchi for HMP leadership and logistics, as well as the dedicated research participants for their invaluable contributions.

AUTHOR CONTRIBUTIONS

Conceptualization, M.P.S., G.M.W., and T.L.M.; Methodology, M.P.S., G.M.W., T.L.M., E.S., B.D.P., W.Z., K.C., H.R., G.J.G.U., S.P., U.S., and A.M.; Investigation, B.D.P., W.Z., K.C., H.R., G.J.G.U., T.M., B.M.H., S. Leopold, C.Y.Y., K.K., D.P., C. Craig, E.C., S.R., S. Lee, C.Z., J.W., M.R.S., L.L., A.M., U.S., S.P., and T.L.M.; Formal Analysis, H.R., W.Z., K.C., B.D.P., T.M., C.Y.Y., B.M.H., E.B., I.B., C. Chen, D.L.R., S. Lee, D. Spakowicz, S.R., S. Leopold, C.Z., M.R.S., L.L., A.M., and S.P.; Writing – Original Draft, B.D.P., W.Z., K.C., H.R., T.M., and C.Y.Y.; Writing – Review & Editing, M.P.S., G.M.W., T.L.M., E.S., S.P., A.M., B.D.P., W.Z., K.C., H.R., G.J.G.U., C.Y.Y.; Visualization, C.A., H.R., W.Z., K.C., and B.D.P.; Funding Acquisition, M.P.S., G.M.W., and T.L.M.; Data Curation, D. Salins; Supervision, M.P.S., G.M.W., T.L.M.,E.S., S.P., U.S., A.M., D.L.R., and M.G.

DECLARATION OF INTERESTS

M.P.S. is a founder and member of the science advisory board of Personalis and SensOmix and a science advisory board member of Genapsys and AxioMX.

Received: May 24, 2017
Revised: October 9, 2017
Accepted: December 14, 2017
Published: January 17, 2018

REFERENCES

Abubucker, S., Segata, N., Goll, J., Schubert, A.M., Izard, J., Cantarel, B.L., Rodriguez-Mueller, B., Zucker, J., Thiagarajan, M., Henrissat, B., et al. (2012). Metabolic reconstruction for metagenomic data and its application to the human microbiome. PLoS Comput. Biol. *8*, e1002358.

Adams, S.H. (2011). Emerging perspectives on essential amino acid metabolism in obesity and the insulin-resistant state. Adv. Nutr. *2*, 445–456.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Brown, C.T., Davis-Richardson, A.G., Giongo, A., Gano, K.A., Crabb, D.B., Mukherjee, N., Casella, G., Drew, J.C., Ilonen, J., Knip, M., et al. (2011). Gut microbiome metagenomics analysis suggests a functional model for the development of autoimmunity for type 1 diabetes. PLoS One *6*, e25792.

Cani, P.D., Bibiloni, R., Knauf, C., Waget, A., Neyrinck, A.M., Delzenne, N.M., and Burcelin, R. (2008). Changes in gut microbiota control metabolic endotoxemia-induced inflammation in high-fat diet-induced obesity and diabetes in mice. Diabetes *57*, 1470–1481.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. Nat. Methods *7*, 335–336.

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

Chung, S.D., Chen, Y.K., and Lin, H.C. (2011). Increased risk of diabetes in patients with urinary calculi: a 5-year followup study. J. Urol. *186*, 1888–1893.

Chung, W.S., Walker, A.W., Louis, P., Parkhill, J., Vermeiren, J., Bosscher, D., Duncan, S.H., and Flint, H.J. (2016). Modulation of the human gut microbiota by dietary fibres occurs at the species level. BMC Biol. *14*, 3.

Considine, R.V., Sinha, M.K., Heiman, M.L., Kriauciunas, A., Stephens, T.W., Nyce, M.R., Ohannesian, J.P., Marco, C.C., McKee, L.J., Bauer, T.L., et al. (1996). Serum immunoreactive-leptin concentrations in normal-weight and obese humans. N. Engl. J. Med. *334*, 292–295.

Contrepois, K., Jiang, L., and Snyder, M. (2015). Optimized analytical procedures for the untargeted metabolomic profiling of human urine and plasma by combining hydrophilic interaction (HILIC) and reverse-phase liquid chromatography (RPLC)-mass spectrometry. Mol. Cell. Proteomics *14*, 1684–1695.

Dao, M.C., Everard, A., Aron-Wisnewsky, J., Sokolovska, N., Prifti, E., Verger, E.O., Kayser, B.D., Levenez, F., Chilloux, J., Hoyles, L., et al. (2016). *Akkermansia muciniphila* and improved metabolic health during a dietary intervention in obesity: relationship with gut microbiome richness and ecology. Gut *65*, 426–436.

Daudon, M., Traxer, O., Conort, P., Lacour, B., and Jungers, P. (2006). Type 2 diabetes increases the risk for uric acid stones. J. Am. Soc. Nephrol. *17*, 2026–2033.

David, L.A., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E., Wolfe, B.E., Ling, A.V., Devlin, A.S., Varma, Y., Fischbach, M.A., et al. (2014). Diet rapidly and reproducibly alters the human gut microbiome. Nature *505*, 559–563.

Dela Cruz, C.S., and Matthay, R.A. (2009). Role of obesity in cardiomyopathy and pulmonary hypertension. Clin. Chest Med. *30*, 509–523.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2012). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Duncan, S.H., Richardson, A.J., Kaul, P., Holmes, R.P., Allison, M.J., and Stewart, C.S. (2002). *Oxalobacter formigenes* and its potential role in human health. Appl. Environ. Microbiol. *68*, 3841–3847.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics *26*, 2460–2461.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. Bioinformatics *27*, 2194–2200.

Engels, C., Ruscheweyh, H.J., Beerenwinkel, N., Lacroix, C., and Schwab, C. (2016). The common gut microbe *Eubacterium hallii* also contributes to intestinal propionate formation. Front. Microbiol. *7*, 713.

Everard, A., Belzer, C., Geurts, L., Ouwerkerk, J.P., Druart, C., Bindels, L.B., Guiot, Y., Derrien, M., Muccioli, G.G., Delzenne, N.M., et al. (2013). Cross-talk between *Akkermansia muciniphila* and intestinal epithelium controls diet-induced obesity. Proc. Natl. Acad. Sci. USA *110*, 9066–9071.

Feng, N., Huke, S., Zhu, G., Tocchetti, C.G., Shi, S., Aiba, T., Kaludercic, N., Hoover, D.B., Beck, S.E., Mankowski, J.L., et al. (2015). Constitutive BDNF/TrkB signaling is required for normal cardiac contraction and relaxation. Proc. Natl. Acad. Sci. USA *112*, 1880–1885.

Festa, A., D'Agostino, R., Jr., Howard, G., Mykkanen, L., Tracy, R.P., and Haffner, S.M. (2000). Chronic subclinical inflammation as part of the insulin resistance syndrome: the Insulin Resistance Atherosclerosis Study (IRAS). Circulation *102*, 42–47.

Finucane, M.M., Stevens, G.A., Cowan, M.J., Danaei, G., Lin, J.K., Paciorek, C.J., Singh, G.M., Gutierrez, H.R., Lu, Y., Bahalim, A.N., et al. (2011). National, regional, and global trends in body-mass index since 1980: systematic analysis of health examination surveys and epidemiological studies with 960 country-years and 9.1 million participants. Lancet *377*, 557–567.

Flanagan, J.L., Simmons, P.A., Vehige, J., Willcox, M.D., and Garrett, Q. (2010). Role of carnitine in disease. Nutr. Metab. (Lond) *7*, 30.

Flegal, K.M., Carroll, M.D., Ogden, C.L., and Curtin, L.R. (2010). Prevalence and trends in obesity among US adults, 1999-2008. JAMA *303*, 235–241.

Fulgenzi, G., Tomassoni-Ardori, F., Babini, L., Becker, J., Barrick, C., Puverel, S., and Tessarollo, L. (2015). BDNF modulates heart contraction force and long-term homeostasis through truncated TrkB.T1 receptor activation. J. Cell Biol. *210*, 1003–1012.

Futschik, M.E., and Carlisle, B. (2005). Noise-robust soft clustering of gene expression time-course data. J. Bioinform. Comput. Biol. *3*, 965–988.

Greenfield, M.S., Doberne, L., Kraemer, F., Tobey, T., and Reaven, G. (1981). Assessment of insulin resistance with the insulin suppression test and the euglycemic clamp. Diabetes *30*, 387–392.

Haaskjold, Y.L., Drotningsvik, A., Leh, S., Marti, H.P., and Svarstad, E. (2015). Renal failure due to excessive intake of almonds in the absence of *Oxalobacter formigenes*. Am. J. Med. *128*, e29–30.

Harris, J.A., and Benedict, F.G. (1918). A biometric study of human basal metabolism. Proc. Natl. Acad. Sci. USA *4*, 370–373.

Holmes, E., Li, J.V., Athanasiou, T., Ashrafian, H., and Nicholson, J.K. (2011). Understanding the role of gut microbiome-host metabolic signal disruption in health and disease. Trends Microbiol. *19*, 349–359.

Hood, L., Lovejoy, J.C., and Price, N.D. (2015). Integrating big data and actionable health coaching to optimize wellness. BMC Med. *13*, 4.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. *37*, 1–13.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Integrative HMP (iHMP) Research Network Consortium. (2014). The Integrative Human Microbiome Project: dynamic analysis of microbiome-host omics profiles during periods of human health and disease. Cell Host Microbe *16*, 276–289.

Janssen, A.W., and Kersten, S. (2017). Potential mediators linking gut bacteria to metabolic health: a critical view. J. Physiol. *595*, 477–487.

Kahn, S.E., Hull, R.L., and Utzschneider, K.M. (2006). Mechanisms linking obesity to insulin resistance and type 2 diabetes. Nature *444*, 840–846.

Khan, M.T., Nieuwdorp, M., and Backhed, F. (2014). Microbial modulation of insulin sensitivity. Cell Metab. *20*, 753–760.

Koeth, R.A., Wang, Z., Levison, B.S., Buffa, J.A., Org, E., Sheehy, B.T., Britt, E.B., Fu, X., Wu, Y., Li, L., et al. (2013). Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. Nat. Med. *19*, 576–585.

Kumar, L., and E Futschik, M. (2007). Mfuzz: a software package for soft clustering of microarray data. Bioinformation *2*, 5–7.

Lam, H.Y., Pan, C., Clark, M.J., Lacroute, P., Chen, R., Haraksingh, R., O'Huallachain, M., Gerstein, M.B., Kidd, J.M., Bustamante, C.D., et al. (2012). Detecting and annotating genetic variations using the HugeSeq pipeline. Nat. Biotechnol. *30*, 226–229.

Larsen, N., Vogensen, F.K., van den Berg, F.W., Nielsen, D.S., Andreasen, A.S., Pedersen, B.K., Al-Soud, W.A., Sorensen, S.J., Hansen, L.H., and Jakobsen, M. (2010). Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. PLoS One *5*, e9085.

Lecomte, V., Kaakoush, N.O., Maloney, C.A., Raipuria, M., Huinao, K.D., Mitchell, H.M., and Morris, M.J. (2015). Changes in gut microbiota in rats fed a high fat diet correlate with obesity-associated metabolic parameters. PLoS One *10*, e0126931.

Lee, S., Zhang, C., Kilicarslan, M., Piening, B.D., Bjornson, E., Hallstrom, B.M., Groen, A.K., Ferrannini, E., Laakso, M., Snyder, M., et al. (2016). Integrated network analysis reveals an association between plasma mannose levels and insulin resistance. Cell Metab. *24*, 172–184.

Ley, R.E., Turnbaugh, P.J., Klein, S., and Gordon, J.I. (2006). Microbial ecology: human gut microbes associated with obesity. Nature *444*, 1022–1023.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

Mardi, T., Toker, S., Melamed, S., Shirom, A., Zeltser, D., Shapira, I., Berliner, S., and Rogowski, O. (2005). Increased erythropoiesis and subclinical inflammation as part of the metabolic syndrome. Diabetes Res. Clin. Pract. *69*, 249–255.

Matthews, D.R., Hosker, J.P., Rudenski, A.S., Naylor, B.A., Treacher, D.F., and Turner, R.C. (1985). Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia *28*, 412–419.

McLaughlin, T., Abbasi, F., Lamendola, C., Liang, L., Reaven, G., Schaaf, P., and Reaven, P. (2002). Differentiation between obesity and insulin resistance in the association with C-reactive protein. Circulation *106*, 2908–2912.

McLaughlin, T., Abbasi, F., Lamendola, C., and Reaven, G. (2007). Heterogeneity in the prevalence of risk factors for cardiovascular disease and type 2 diabetes mellitus in obese individuals: effect of differences in insulin sensitivity. Arch. Intern. Med. *167*, 642–648.

McLaughlin, T., Lamendola, C., Coghlan, N., Liu, T.C., Lerner, K., Sherman, A., and Cushman, S.W. (2014). Subcutaneous adipose cell size and distribution: relationship to insulin resistance and body fat. Obesity *22*, 673–680.

McLaughlin, T., Stuhlinger, M., Lamendola, C., Abbasi, F., Bialek, J., Reaven, G.M., and Tsao, P.S. (2006). Plasma asymmetric dimethylarginine concentrations are elevated in obese insulin-resistant women and fall with weight loss. J. Clin. Endocrinol. Metab. *91*, 1896–1900.

McLaughlin, T., Craig, C., Liu, L.F., Perelman, D., Allister, C., Spielman, D., and Cushman, S.W. (2016). Adipose cell size and regional fat deposition as predictors of metabolic response to overfeeding in insulin-resistant and insulin-sensitive humans. Diabetes *65*, 1245–1254.

Nishimura, S., Manabe, I., Nagasaki, M., Eto, K., Yamashita, H., Ohsugi, M., Otsu, M., Hara, K., Ueki, K., Sugiura, S., et al. (2009). CD8+ effector T cells contribute to macrophage recruitment and adipose tissue inflammation in obesity. Nat. Med. *15*, 914–920.

Patten, C.L., Blakney, A.J., and Coulson, T.J. (2013). Activity, distribution and function of indole-3-acetic acid biosynthetic pathways in bacteria. Crit. Rev. Microbiol. *39*, 395–415.

Patwardhan, A., Harris, J., Leng, N., Bartha, G., Church, D.M., Luo, S., Haudenschild, C., Pratt, M., Zook, J., Salit, M., et al. (2015). Achieving high-sensitivity for clinical applications using augmented exome sequencing. Genome Med. *7*, 71.

Pedersen, H.K., Gudmundsdottir, V., Nielsen, H.B., Hyotylainen, T., Nielsen, T., Jensen, B.A., Forslund, K., Hildebrand, F., Prifti, E., Falony, G., et al. (2016). Human gut microbes impact host serum metabolome and insulin sensitivity. Nature *535*, 376–381.

Pei, D., Jones, C.N., Bhargava, R., Chen, Y.D., and Reaven, G.M. (1994). Evaluation of octreotide to assess insulin-mediated glucose disposal by the insulin suppression test. Diabetologia *37*, 843–845.

Pesu, M., Watford, W.T., Wei, L., Xu, L., Fuss, I., Strober, W., Andersson, J., Shevach, E.M., Quezado, M., Bouladoux, N., et al. (2008). T-cell-expressed proprotein convertase furin is essential for maintenance of peripheral immune tolerance. Nature *455*, 246–250.

Poretsky, R., Rodriguez-R, L.M., Luo, C., Tsementzi, D., and Konstantinidis, K.T. (2014). Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. PLoS One *9*, e93827.

Price, N.D., Magis, A.T., Earls, J.C., Glusman, G., Levy, R., Lausted, C., McDonald, D.T., Kusebauch, U., Moss, C.L., Zhou, Y., et al. (2017). A wellness study of 108 individuals using personal, dense, dynamic data clouds. Nat. Biotechnol. *35*, 747–756.

Rego, S., Dagan-Rosenfeld, O., Zhou, W., Sailani, M.R., Limcaoco, P., Colbert, E., Avina, M., Wheeler, J., Craig, C., Salins, D., et al. (2017). High frequency actionable pathogenic exome mutations in an average-risk cohort. bioRxiv. https://doi.org/10.1101/151225.

Ridaura, V.K., Faith, J.J., Rey, F.E., Cheng, J., Duncan, A.E., Kau, A.L., Griffin, N.W., Lombard, V., Henrissat, B., Bain, J.R., et al. (2013). Gut microbiota from twins discordant for obesity modulate metabolism in mice. Science *341*, 1241214.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Roopchand, D.E., Carmody, R.N., Kuhn, P., Moskal, K., Rojas-Silva, P., Turnbaugh, P.J., and Raskin, I. (2015). Dietary polyphenols promote growth of the gut bacterium *Akkermansia muciniphila* and attenuate high-fat diet-induced metabolic syndrome. Diabetes *64*, 2847–2858.

Rosas-Vargas, H., Martinez-Ezquerro, J.D., and Bienvenu, T. (2011). Brain-derived neurotrophic factor, food intake regulation, and obesity. Arch. Med. Res. *42*, 482–494.

Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. Bioinformatics *27*, 863–864.

Schooneman, M.G., Vaz, F.M., Houten, S.M., and Soeters, M.R. (2013). Acylcarnitines: reflecting or inflicting insulin resistance? Diabetes *62*, 1–8.

Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., and Huttenhower, C. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. Nat. Methods *9*, 811–814.

Serino, M., Luche, E., Gres, S., Baylac, A., Bergé, M., Cenac, C., Waget, A., Klopp, P., Iacovoni, J., Klopp, C., et al. (2012). Metabolic adaptation to a high-fat diet is associated with a change in the gut microbiota. Gut *61*, 543–553.

Shah, N., Tang, H., Doak, T.G., and Ye, Y. (2011). Comparing bacterial communities inferred from 16S rRNA gene sequencing and shotgun metagenomics. Pac. Symp. Biocomput. *2011*, 165–176.

Shen, S.W., Reaven, G.M., and Farquhar, J.W. (1970). Comparison of impedance to insulin-mediated glucose uptake in normal subjects and in subjects with latent diabetes. J. Clin. Invest. *49*, 2151–2160.

Tiwari, S., and Ndisang, J.F. (2014). The role of obesity in cardiomyopathy and nephropathy. Curr. Pharm. Des. *20*, 1409–1417.

Truong, D.T., Franzosa, E.A., Tickle, T.L., Scholz, M., Weingart, G., Pasolli, E., Tett, A., Huttenhower, C., and Segata, N. (2015). MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nat. Methods *12*, 902–903.

Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E., Sogin, M.L., Jones, W.J., Roe, B.A., Affourtit, J.P., et al. (2009). A core gut microbiome in obese and lean twins. Nature *457*, 480–484.

Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R., and Gordon, J.I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. Nature *444*, 1027–1031.

Ussher, J.R., Lopaschuk, G.D., and Arduini, A. (2013). Gut microbiota metabolism of L-carnitine and cardiovascular risk. Atherosclerosis *231*, 456–461.

van Etten, R.W., de Koning, E.J., Verhaar, M.C., Gaillard, C.A., and Rabelink, T.J. (2002). Impaired NO-dependent vasodilation in patients with type II (non-insulin-dependent) diabetes mellitus is restored by acute administration of folate. Diabetologia *45*, 1004–1010.

Vatanen, T., Kostic, A.D., d'Hennezel, E., Siljander, H., Franzosa, E.A., Yassour, M., Kolde, R., Vlamakis, H., Arthur, T.D., Hamalainen, A.M., et al. (2016). Variation in microbiome LPS immunogenicity contributes to autoimmunity in humans. Cell *165*, 842–853.

Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. *73*, 5261–5267.

Wang, T., Lu, J., Xu, Y., Li, M., Sun, J., Zhang, J., Xu, B., Xu, M., Chen, Y., Bi, Y., et al. (2013). Circulating prolactin associates with diabetes and impaired glucose regulation: a population-based study. Diabetes Care *36*, 1974–1980.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. *10*, 57–63.

Wikoff, W.R., Anfora, A.T., Liu, J., Schultz, P.G., Lesley, S.A., Peters, E.C., and Siuzdak, G. (2009). Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. Proc. Natl. Acad. Sci. USA *106*, 3698–3703.

Williams, E.G., Wu, Y., Jha, P., Dubuis, S., Blattmann, P., Argmann, C.A., Houten, S.M., Amariuta, T., Wolski, W., Zamboni, N., et al. (2016). Systems proteomics of liver mitochondria function. Science *352*, aad0189.

Xia, J., and Wishart, D.S. (2016). Using MetaboAnalyst 3.0 for comprehensive metabolomics data analysis. Curr. Protoc. Bioinformatics *55*, 14 10 11–14 10 91.

Yeni-Komshian, H., Carantoni, M., Abbasi, F., and Reaven, G.M. (2000). Relationship between several surrogate estimates of insulin resistance and quantification of insulin-mediated glucose disposal in 490 healthy nondiabetic volunteers. Diabetes Care *23*, 171–175.

Yoon, M.S. (2016). The emerging role of branched-chain amino acids in insulin resistance and metabolism. Nutrients *8*, https://doi.org/10.3390/nu8070405.

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. Stat. Appl. Genet. Mol. Biol. *4*, Article17.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical Commercial Assays		
Gentra Puregene Kit	Qiagen	158445
AllPrep DNA/RNA/Protein Kit	Qiagen	80004
Powersoil Extraction Kit	Qiagen	12888
TruSeq Stranded Total RNA Library Prep Kit	Illumina	RS-122-2301
Ampure XP Beads	Beckman Coulter	A63881
Tandem Mass Tag Isobaric Tags (10 plex)	ThermoFisher	90010
Proseek Multiplex CVD I, Inflammation, and Oncology Panels	Olink Biosciences	varies
Stanford HIMC 63-plex immunoassay	Luminex/Stanford	N/A
Metabolon metabolomics	Metabolon	N/A
Deposited Data		
RNA-seq raw data	This paper	hmp2.org
Exome raw data	This paper	hmp2.org
Metabolomics raw data	This paper	hmp2.org
Proteomics raw data	This paper	hmp2.org and PRIDE
Microbiome Raw data	This paper	hmp2.org

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact Michael P. Snyder (mpsnyder@stanford.edu).

EXPERIMENTAL MODELS AND SUBJECT DETAILS

Participant Recruitment and IRB Consent

Participants provided informed written consent for the study under a research study protocol approved by the Stanford University Institutional Review Board. All participants were studied after an overnight fast at the Stanford Clinical and Translational Research Unit (CTRU).

Participants were recruited via placement of advertisements in local newspapers and radio stations seeking "healthy volunteers" for a study on body weight changes and effect on metabolism. Screening in the CTRU entailed history and physical, anthropometric measurements, and fasting blood tests for exclusions including presence of diabetes defined as fasting blood glucose > 126 mg/dL, anemia defined as hematocrit < 30, renal disease defined as creatinine > 1.5, history of any cardiovascular, malignancy, chronic inflammatory, psychiatric disease, and history of any bariatric surgery or liposuction.

Eligible consented subjects underwent quantification of insulin mediated glucose uptake via the modified insulin suppression test as previously described and validated (Greenfield et al., 1981; Pei et al., 1994; Shen et al., 1970). Briefly, following an overnight fast, subjects were infused for 180 minutes with octreotide (0.27 $\mu g/m^2$ min), insulin (25 mU/m^2 min), and glucose (240 mg/m^2 min). Blood was drawn at 10-minute intervals from 150 to 180 minutes of the infusion to measure plasma glucose (oximetric method) and insulin (radioimmunoassay) concentrations: the mean of these four values comprised the steady-state plasma glucose (SSPG) and insulin concentrations for each individual. At steady state, insulin concentrations (65 uU/mL) are similar in all subjects and the SSPG provides a direct measure of the relative ability of insulin to dispose of a glucose load: the higher the SSPG concentration, the more insulin-resistant the individual. While the SSPG is distributed continuously, for the purpose of this study, we defined IS as SSPG<120 mg/dL and IR as SSPG>150 (Yeni-Komshian et al., 2000), largely to provide separation between the two groups. Individuals with SSPG between 120 and 150 were excluded.

Subjects were then placed on a controlled weight gain diet for 30 days by adding an average of 880 calories/day in addition to their usual daily intake as previously described (McLaughlin et al., 2016). Exact caloric excess, administered by research dietitian in the form of snacks/beverages, with fixed macronutrient composition of 50% carbohydrate , 35% fat (<7% saturated), 15% protein, was

calculated individually using the Harris Benedict Equation (Harris and Benedict, 1918) for each subject to attain weight gain of 0.8 kg per week (goal of 3.2 kg total). Subjects were not allowed to change their physical activity or change medications during the study. Weekly visits with study dietitian for weight checks, return of food diary, dispensation of snacks, and caloric adjustment if needed, ensured compliance. After weight gain for 28 days, subjects underwent 7 days of weight maintenance with eucaloric diet after which they were resampled (T2) to measure effect of increased body weight rather than caloric excess *per se*. subjects then underwent supervised weight loss for 6-9 weeks (average 60 days) such that they returned to their baseline weight. A three-week range was given to ensure all subjects could return to baseline weight. One subject was unable to return to baseline weight due to lack of interest or concern about the higher weight, and several highly motivated individuals lost weight to below baseline.

For validation data obtained from a separate cohort conducted in Sweden, subjects were recruited by newspaper advertisements or by having participated in other studies in the laboratory. Inclusion criteria were general good health and no chronic medication. After initial careful phenotyping, including euglycemic clamps to measure degree of insulin sensitivity, 15 male individuals were placed on a hypercaloric diet aimed to increase their body weights by around 8% over a period of 6-8 wks. Phenotyping was then repeated after the weight gain and the volunteers were then helped by nutritionists and exercise plans to recover initial body weights.

Inclusion criteria were male sex, healthy and in general good health. Participants underwent clinical, radiological and metabolomic evaluation before and after a controlled weight gain of about 8% through a hypercaloric diet based on diet recall of each individual by a nutritionist. The 8% weight increase required about 6-8 weeks on the hypercaloric diet. Unbiased metabolomics was performed before and after the weight gain (Metabolon Inc.).

Lifestyle factors were evaluated through a questionnaire filled out at the laboratory. Body weight, height, waist and hip circumferences were recorded and BMI was calculated. The proportions of body fat and lean body mass were determined using bioelectrical impedance (single frequency, 50 kHz; Animeter, HTS, Odense, Denmark). Blood pressure was measured in a sitting position after a five minutes rest with a mercury sphygmomanometer.

To evaluate glucose tolerance status, fasting blood samples were drawn after 12 hours of fasting and were followed by an oral glucose tolerance test (OGTT) (75 g glucose orally). Samples for measurement of plasma glucose and serum insulin were drawn after 0, 30, 60 and 120 minutes.

At a separate examination and after 12 hours of fasting, an intravenous glucose tolerance test (IVGTT) was performed to determine the first and second phases of insulin secretion. A bolus of glucose (300 mg/kg in a 50% solution) was given within 30 seconds into the antecubital vein. Samples for the measurement of plasma glucose and insulin (arterialised venous blood) were drawn at -5, 0, 2, 4, 6, 8, 10, 20, 30, 40, 50 and 60 minutes. The acute and the late insulin responses, i.e. incremental area under the insulin curve, (AIR, 0-10 minutes; LIR, 10-60 minutes) were calculated using the trapezoidal method.

Fasting plasma insulin and fasting plasma glucose from the OGTT were used to calculate a HOMA-IR index using the formula HOMA-IR = (fasting plasma glucose x fasting plasma insulin) / 22.5 (Matthews et al., 1985). The HOMA-IR value was used to assess insulin sensitivity in this study.

Plasma glucose was measured using standard laboratory methods (Department of Chemistry, Sahlgrenska University Hospital, Gothenburg, Sweden). Plasma insulin was measured at the University of Tübingen, Germany, by micro-particle enzyme immunoassay (Abbott Laboratories, Tokyo, Japan).

This latter study was approved by the local Ethical Committees at the Sahlgrenska Academy at the University of Gothenburg and was performed in agreement with the Declaration of Helsinki.

METHOD DETAILS

Sample Preparation - Blood Samples

Blood was drawn from overnight-fasted participants at the indicated timepoints at the Stanford Clinical Translational Research Unit (CTRU). An aliquot of blood was incubated at room temperature to coagulate; clots were subsequently pelleted and the serum supernatant was pipetted off and immediately frozen at -80C. Blood from separate EDTA-tubes was immediately layered onto Ficoll media and spun via gradient centrifugation. The top layer plasma was pipetted off, aliquoted and immediately frozen at -80C. The PBMC layer was removed, counted via cell counter and aliquots of PBMCs were further pelleted and flash-frozen. For the subsequent multi-omic analyses, PBMCs were thawed on ice, and subsequently lysed and processed to DNA, RNA and protein fractions via Allprep Spin Columns (Qiagen) according to the manufacturer's instructions and using the Qiashredder lysis option. Plasma analysis was performed on individual aliquots to prevent freeze-thaw cycles.

Sample Preparation - Stool

Stool sampling was conducted according to the Human Microbiome Project – Core Microbiome Sampling Protocol A (hmpdacc.org). Once samples were received in the lab, they were subsequently stored in -80C until further processing.

Exome Sequencing

Briefly, DNA was isolated from blood via Gentra Puregene Kits (Qiagen) according to the manufacturer's instructions. Exome sequencing was performed in a CLIA- and CAP-accredited facility utilizing the ACE Clinical Exome Test (Personalis) (Patwardhan et al., 2015). Variant calling was performed using an in-house developed automated pipeline (HugeSeq) (Lam et al., 2012).

RNA-Seq

The transcriptome was evaluated by RNA sequencing (RNA-seq) (Wang et al., 2009) from bulk PBMCs. Ribosomal RNA was first removed and total RNA from each sample was converted into a cDNA sequencing library (using Illumina TruSeq Stranded Total RNA Gold kits according to the manufacturer's instructions). Each resultant library was quantified via Agilent Bioanalyzer as well as Qubit Fluorometric quantification (ThermoFisher) using a dsDNA high sensitivity kit. Quantified, barcoded libraries were normalized and mixed at equimolar concentrations into a multiplexed sequencing library. The pooled library was quantified and loaded onto a single lane of an Illumina flowcell and sequenced on a HiSeq 2500.

Microbiome Sequencing -16S

DNA extractions were performed following Human Microbiome Project – Core Microbiome Sampling Protocol A (HMP Protocol # 07-001, v12.0). Metagenomic DNA was isolated in a clean hood using the MOBIO PowerSoil DNA Extraction kit, with added proteinase K, followed by lysozyme and staphylolysin treatment. These digestion steps greatly improved the extraction of many difficult- to-lyse Gram-positive species.

For 16S (Bacterial) rRNA gene amplification, the V1 through V3 hyper-variable regions (V1-V3) of 16S were amplified from the metagenomic DNA using primers 27F and 534R (27F:5'-AGAGTTTGATCCTGGCTCAG-3' and 534R: 5'- ATTACCGCGGCTGCTGG-3'). The oligonucleotides containing the 16S primer sequences also contain an adaptor sequence for the Illumina sequencing platform. A barcode sequence unique to each sample is embedded within each of the forward and reverse oligonucleotides used to create the amplicons (dual tags). The uniquely barcoded amplicons from multiple samples were pooled and sequenced on the Illumina MiSeq sequencing platform using a V3 2x300 sequencing protocol.

Microbiome Sequencing – Metagenome Shotgun

DNA extracted from stools were also subject to whole genome metagenomic shotgun sequencing (mWGS). The libraries were prepared following a standard protocol from Illumina, and at least 1Gb of 150 bp pair-end (PE) reads per sample were sequenced on an Illumina HiSeq or MiSeq instrument.

Untargeted Metabolomics by LC-MS

Plasma samples were prepared and analyzed as previously described (Contrepois et al., 2015). Briefly, 400 μl of a solvent mixture of 1:1:1 acetone:acetonitrile:methanol was added to 100 μl of plasma, mixed for 15 min at 4°C and incubated for 2h at -20°C to allow protein precipitation. The solvent mixture contained seven internal standards to confirm extraction efficiency and evaluate LC-MS instrument performance. The supernatant was collected after centrifugation at 10,000 rpm for 10 min at 4°C and evaporated to dryness under nitrogen. The dry extracts were then reconstituted with 100 μl of a mixture of 1:1 methanol:water before analysis.

The metabolite extracts were analyzed in HILIC ESI(+) MS, HILIC ESI(-) MS, RPLC ESI(+) MS, RPLC ESI(-) MS using a Thermo Ultimate 3000 RSLC system coupled with a Thermo Q Exactive plus mass spectrometer. The Q Exactive plus was equipped with a HESI-II probe and operated in full MS scan mode. MS/MS data were acquired on quality control samples (QCs = equimolar mixture of all the samples comprised in the study). The source conditions were as follows: Spray Voltage = 3.4 kV (both ESI pos. and neg.), Vaporizer = 310°C, Capillary temp. = 375°C, S-Lens RF level = 55.0, Sheath Gas = 45 for HILIC and 55 for RPLC, Auxiliary gas = 8 for HILIC and 15 for RPLC, Sweep Gas = 0. The acquisition settings were as follow: AGC (MS) = 3e6, AGC (MS2) = 1e5, Injection Time (MS) = 200 ms, Injection Time (MS2) = 50 ms, Mass Range = 70-1000 Da, Resolution MS = 70,000 (FWHM at m/z 200), Resolution MS2 = 35,000 (FWHM at m/z 200), Top-10 experiments, Isolation Window = 1.0 Da, Dynamic Exclusion = 14 for HILIC and 8s for RPLC, Normalized Collision Energy (NCE) = 25 and 35 for HILIC, 25 and 50 for RPLC. Between each batch, the source and the transfer capillary were cleaned and the mass spectrometer calibrated using an infusion of Pierce LTQ Velos ESI Positive Ion Calibration Solution or Pierce ESI Negative Ion Calibration Solution.

HILIC experiments were performed using a ZIC-HILIC column 2.1 x 100 mm, 3.5 μm, 200Å (Merck Millipore) 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) (Contrepois et al., 2015). Metabolites were eluted from the columns at 0.5 mL/min using a 1–99% phase A gradient over 15 min. Before each injection, the column was equilibrated for 5 min with 1% phase A. Twelve QCs were injected at the beginning of the batch to equilibrate and condition the LC-MS system. The oven temperature was set at 40 °C, and the injection volume was 5 μL. RPLC experiments were performed using a Zorbax SBaq column 2.1 x 50 mm, 1.7 μm, 100Å (Agilent Technologies) and mobile phase solvents consisting of 0.06% acetic acid in water (A) and 0.06% acetic acid in methanol (B). Metabolites were eluted from the columns at 0.6 mL/min using a 1–99% phase B gradient over 9 min. Before each injection, the column was equilibrated for 5 min with 1% phase B. Five QCs were injected at the beginning of the batch to equilibrate and condition the LC-MS system. The oven temperature was set at 60°C, and the injection volume was 5 μL. Metabolite identities for significant metabolites were confirmed by validation standards if available (standards for tetrahydrocortisol glucuronide were not available).

Additional Metabolomics Assays

For the independent cohort conducted in Sweden. non-targeted metabolomic profiling analysis was performed by Metabolon Inc. Samples were prepared using the automated MicroLab STAR® system from Hamilton Company. A recovery standard was added

prior to the first step in the extraction process for quality control purposes. Sample preparation was conducted using aqueous methanol extraction process to remove the protein fraction while allowing maximum recovery of small molecules. The resulting extract was divided into four fractions: one for analysis by UPLC/MS/MS (positive mode), one for UPLC/MS/MS (negative mode), one for GC/MS, and one for backup. Samples were dried under nitrogen using a TurboVap® (Zymark) and resolubilized for the appropriate instrument either UPLC/MS/MS or GC/MS.

Untargeted Proteomics by LC-MS

Ethanol-precipitated protein pellets from PBMC samples were resuspended in 110 uL of 100 mM ammonium bicarbonate (ABC) and 0.1% Octyl β-D-glucopyranoside (OG) and subjected to a sonicator probe for efficient resuspension. The Thermo Scientific Pierce Micro BCA Protein Assay Kit was then used to quantify protein levels. 1 mg of each resuspended protein sample then underwent denaturation in 900 uL of 8M Urea, 100mM ABC and 0.1% OG. The samples were once again sonicated with a sonicator probe. Each sample then underwent chemical reduction with 10 uL of 1M dithiolthreitol (DTT) and incubated at room temperature for 2 hours. 20 uL of 1M iodoacetamide (IAA) was then added in each sample for the alkylation step. The alkylation reactions were left to incubate in the dark at room temperature for 1 hour.

After fractionation, the unbound fraction from each sample was concentrated and buffer exchanged into 100 uL of 100 mM tetraethylammonium bromide (TEAB) using the Amicon Ultra-15 followed by the Amicon Ultra-4 centrifugal filter unit. Each sample was digested with 40 ng of Promega Sequencing Grade Trypsin in 100 mM TEAB solution. Samples were then labeled with 10-plex Thermo Scientific™ Tandem Mass Tag™ (TMT) Reagents, using instructions provided by the manufacturer. All the samples were divided into groups of nine and pooled together with a master reference sample to create 10-plexed samples.

Each TMT 10-plex sample underwent shotgun liquid chromatography-tandem mass spectrometry (LC-MS/MS) with the LTQ Orbitrap Elite™ Hybrid Ion Trap-Orbitrap Mass Spectrometer coupled with a Dionex RSLC 3000 Nano-HPLC. 15 μL of each sample was loaded onto a C18 trap column at 5 μL/min for 10 minutes. Peptides were then separated by a 25 cm C18 analytical column (Picofrit 75 μm ID, New Objective, packed in-house with MagicC18 AQ resin). Tryptic peptides were separated using a multi-step gradient at a flow rate of 0.6 μL/minute in which Buffer B (0.1% FA in acetonitrile) was increased from 0% (100% Buffer A, 0.1% FA in water) to 85% over 120 minutes. The column was re-equilibrated for 20 minutes at 98% Buffer A. Blank runs were performed between each sample. Samples were then ionized by electrospray ionization set to 2.25 kV with a capillary temperature of 200 °C. An initial MS1 scan over an m/z range of 400-1800 was performed, followed by 10 data-dependent higher energy collision-induced dissociation fragmentation (35 eV) events on the 10 most intense +2 or +3 ions from the MS1 spectrum over an acquisition time of 140 minutes.

Plasma Protein Profiling Using Proseek Multiplex

Proteins were quantified from plasma at all time points using multiplex proximity extension assays (Proseek Multiplex, Olink Biosciences) according to the manufacturer's instructions. In this study three panels, cardiovascular disease (CVD I 96x96), inflammation (Inflammation I 96x96) and oncology (Oncology I 96x96) were applied, and a total of 276 proteins were measured in the plasma samples. Briefly, in each well of 96-well plate 3 μL incubation solution containing with two incubation probes, protein target-specific antibodies conjugated with distinctive single-strand oligonucleotides, was mixed with 1 μL plasma sample. The mixture was incubated overnight at 4 °C , and then added with 96 μL extension solution containing extension enzyme and PCR reagents. The plate was then placed in a thermal cycler for the extension (50 °C, 20 min) and preamplification (95 °C 30 min, 17 cycles of 95 °C 30 sec, 54 °C 1 min and 60 °C 1 min). A 96.96 dynamic array IFC (Fluidigm) was prepared and primed according to the manufacturer's instructions, 2.8 μL of the extension mix was mixed with 7.2 μL detection solution in a new 96-well plate. Next, 5 μL of the mix was loaded to the primed 96.96 Dynamic Array IFC (the right inlets), and 5 μL of each the 96 primer pairs was loaded to the other side of the 96.96 Dynamic Array IFC. The program for protein expression was run on a Fluidigm Biomark using the provided Proseek program (Olink BioSciences).

Luminex Assays

Levels of circulating cytokines in the blood were measured using a 63-plex Luminex antibody-conjugated bead capture assay (Affymetrix) that has been extensively characterized and benchmarked by the Stanford Human Immune Monitoring Center (HIMC). This assay was performed by the Human Immune Monitoring Center at Stanford University. Human 63-plex kits were purchased from eBiosciences/Affymetrix and used according to the manufacturer's recommendations with modifications as described below. Briefly, beads were added to a 96 well plate and washed using a Biotek ELx405 washer. Samples were added to the plate containing the mixed antibody-linked beads and incubated at room temperature for 1 hour followed by overnight incubation at 4 °C with shaking. Cold and room temperature incubation steps were performed on an orbital shaker at 500-600 rpm. Following the overnight incubation plates were washed using a Biotek ELx405 washer and then biotinylated detection antibody added for 75 minutes at room temperature with shaking. The plate was washed as above and streptavidin-PE was added. After incubation for 30 minutes at room temperature a wash was performed as above and reading buffer was added to the wells. Each sample was measured in duplicate. Plates were read using a Luminex 200 instrument with a lower bound of 50 beads per sample per cytokine. Custom assay control beads by Radix Biosolutions are added to all wells.

STATISTICAL ANALYSES

Power and Variance Calculations

To estimate power and sample sizes, we assumed an experiment on a set of patients with two conditions, before and after. From this, we wished to investigate the increase in power when we used paired tests vs unpaired tests. Our assumption is that there is inter-subject variance (personalized variance) and an effect with a size and variance. Thus, the formalized tests are as follows:

1 Unpaired test

In the case of a t-test of unequal variance we have the following test statistic:

$$t = \frac{\overline{X_1} - \overline{X_2}}{s_D} \quad \text{(Equation 1)}$$

where

$$s_D^2 = \frac{s_1^2 + s_2^2}{N}$$

as calculated in a standard Welch's test and N being the number of samples in each group (assuming each group has equal number of samples).

2 Paired test

For a paired t test, we compute the following test statistic:

$$t = \frac{\overline{X}_\Delta}{s_P}$$

where

$$s_P^2 = \frac{s_\Delta^2}{n}$$

with n being the number of paired samples necessary and $s\Delta$ the sample standard deviation of the difference due to the effect.

3 Power calculation

To compute the statistical power of the paired test, we compute the value of the cumulative distribution function under the alternative hypothesis HA for both sides:

$$power = P(t_n > T_{\alpha,n} | H_A) + P(t_n < -T_{\alpha,n} | H_A)$$

where tn is the computed test statistics with a population size of n and $T\alpha,n$ is the critical value from the Student t-distribution with n degrees of freedom for a given significance level α. For the power analysis, the power for the paired analysis is computed using the sample size from our analysis. Then, for increasing values of n, the test statistic for the unpaired test is computed until the power of the unpaired tests exceeds the power of the paired test. This value of n is then returned and reported as minimally necessary population size to achieve the same statistical power. Note that n for the unpaired test represents the number of subjects in one group, in the current study we had 3 groups (timepoints T1, T2 and T3) so the value is multiplied by 3.

Random Forest and AdaBoost Classification

For these analyses, we proposed an automated prediction model based on delta metabolomics feature array which incorporates at least 2-time subsequent timepoints of SSPG measurement. The delta features (Δ) were computed by taking the 1st order derivatives of metabolomics predictor values of the consecutive time-points for every patient: $\Delta_i(t^n) = metabolomics_i(t^n) - metabolomics_i(t^{n-1})$, where $metabolomics_i(t^n)$ is the ith metabolomics predictor value at time n and $metabolomics_i(t^{n-1})$ is value at the earlier timepoint. Using this technique, we created a *"delta cohort"* by computing the delta metabolomics feature array on the 23 patients, which gives us a total of 38 sample timepoints with 2472 features for each timepoint. We applied Z-score normalization to standardize the feature array to have a mean of zero and standard deviation of +/- 1.

On top of the *delta cohort*, we adopted two different classes of supervised machine learning method to predict the temporal change in the SSPG value:

(i) Binary classification using ensemble learning – We labeled the samples in the *delta cohort* into two groups based on SSPG increase or decrease compared to the earlier time point: (**Group i**) $SSPG\ increase \rightarrow SSPG(t^n) \geq SSPG(t^{n-1})$; (**Group ii**) $SSPG\ decrease \rightarrow SSPG(t^x) < SSPG(t^{x-1})$. In order to propose generalizable model to discriminate the *SSPG increase* and *SSPG decrease* by analyzing the delta metabolomic feature array, we trained two popular ensemble learning models – Random

Forest and AdaBoost. Finally, training and testing of the models were conducted using hold-out validation where 8 sample points (20% of the total cohort) were randomly chosen for testing. We utilized python scikit learn framework to implement the machine learning models. To handle high dimensionality of the metabolomics feature vector, the Random Forest was trained with 50 base estimators, 100 maximum tree depths, and entropy split criterion. AdaBoost was trained using 100 base estimators and learning rate of 0.1.

(ii) Δ_{SSPG} computation at time (t^n) using regression – We computed Δ^n_{SSPG} as: $\Delta^n_{SSPG} = SSPG(t^n) - SSPG(t^{n-1})$. We created a regression-based prediction model for computing the Δ^n_{SSPG} by analyzing delta metabolomics features. The model operates based on ElasticNet regularized regression that combines L_1 penalties of Lasso and L_2 penalties of Ridge to overcome the limitations of incorporating high dimensional features for relatively small number of samples.

RNA-Seq Data Processing and Analysis

For RNA-seq data analysis, raw transcripts were processed for adapter removal and low-quality base trimming using the Trimmomatic algorithm (Bolger et al., 2014). Depending on the analysis, reads were either mapped to personal genomes constructed from exome vcfs or the hg19 human reference genome using the STAR aligner (Dobin et al., 2012). Read counts after trimming and quality filtering and the fraction of features that were successfully mapped per sample are listed in Table S5. Counts were assessed from STAR-aligned sam files for all ENSEMBL transcripts using the featureCounts function in the Subread package (Liao et al., 2014). Raw feature counts were normalized via the edgeR package and differentially expressed genes were calculated via negative-binomial exact test with false-discovery rate correction (Robinson et al., 2010).

Unsupervised co-expression module discovery was performed using weighted gene co-expression network analysis (WGCNA) (Zhang and Horvath, 2005). The scale-free topology overlap matrix was computed using the "signed" parameter and using an empirically-defined soft threshold power of 12, and co-expressing modules were defined from this network. For each identified module of co-expression biomolecules, representative eigengenes were calculated (WGCNA function 'moduleEigengenes') and correlations between module eigengenes and clinical parameters were calculated (corresponding p-values were adjusted for MHT using R function p.adjust (using the Benjamini-Hochberg setting).

Microbiome

The 16S rRNA gene is about 1.5kb, and includes nine variable regions that provide much of the sequence distinction between different taxa. Variable regions one through three are generally sufficient to identify taxa down to the genus level, and sometimes to the species level. Illumina's software handles initial processing of all the raw sequencing data. One mismatch in primer and zero mismatch in barcodes were applied to assign read pairs to the appropriate sample within a pool of samples. Barcode and primers were removed from the reads. Reads were further processed by removing the sequences with low quality (average qual <35) and ambiguous bases (N's). Chimeric amplicons were removed using UChime (Edgar et al., 2011), and amplicon sequences were clustered and Operational Taxonomic Units (OTU) picking by Usearch (Edgar, 2010) against GreenGenes database (May 2013 version) and final taxonomic assignment were performed using RDP-classifier (Wang et al., 2007). All details were executed using QIIME (Caporaso et al., 2010) with custom scripts. Alignment results and read counts for microbiome samples are listed in Table S6.

For metagenomic data analysis, downstream processing of the mWGS reads included a) identification and masking human reads (using NCBI's BMTagger, ftp://ftp.ncbi.nlm.nih.gov/pub/agarwala/bmtagger); b) removal of duplicated reads which are artifacts of sequencing process, c) trimming low quality bases and d) low- complexity screening (b-d were done through PRINSEQ (Schmieder and Edwards, 2011)). Reads trimmed to less than 60bp were removed and the remaining high quality reads were analyzed using MetaPhlAn2 (Segata et al., 2012; Truong et al., 2015) for strain-level taxonomic classification and HUMAnN2 (Abubucker et al., 2012) (http://huttenhower.sph.harvard.edu/humann2, v6.0) for functional reconstruction on the gene and pathway levels.

Abundance tables (taxonomy, genes and pathways) obtained above were further analyzed in R (version 3.0.1) by custom scripts. Specifically, pairwise comparisons on the abundances between IR and IS individuals and between timepoints were performed using Wilcoxon rank-sum test, and Spearman's rank correlation coefficients were calculated for pairwise associations. Also, in the correlation analysis, a linear rank based regression model was employed to determine if trends were different between IR and IS groups (Rfit R package). As both 16S rRNA and whole genome shotgun sequencing (mWGS) methods have individual limitations (Poretsky et al., 2014; Shah et al., 2011), results were shown only for pairwise comparisons that were detected and significant by both methods. Among the intra- and inter-omic associations, mWGS data were used for species level analysis, and 16S for analysis on the phylum and genus levels to obtain the most accurate presentations.

Metabolomics

Samples for metabolomics were prepared and analyzed randomly. Data were analyzed using an in-house data analysis pipeline written in R (version 3.0.1). Metabolite features (characterized by a unique mass/charge ratio and retention time) were extracted, aligned and quantified with the "XCMS" package (version 1.39.4) after conversion of .RAW files to .mzXML using the ProteoWizard MS convert tool. Grouping and annotation were performed with the "CAMERA" package (version 1.16.0). Features from blanks were discarded. The signal drift with time was corrected by applying LOESS (Local Regression) normalization were each feature was

independently corrected by fitting a LOESS curve to the MS signal measured in QCs injected repetitively along the batch. After normalization, samples from the same individual tend to cluster together (Figure S6). Metabolic features were putatively identified by matching the accurate masses (± 5 ppm) against a local database containing +60,000 entries (compilation of various public databases such as HMDB, FoodB, DrugBank). Discriminant metabolites were validated manually by comparing the retention time and/or fragmentation spectrum to a local or public spectral libraries. Pathway analysis was performed using the metabolites with a HMDB accession number in the web tool Metaboanalyst (Xia and Wishart, 2016).

Untargeted Proteomics

Acquired data was converted and searched from the .raw files into peptide spectral matches (PSMs) using the Proteome Discoverer™ Software against the 2015 Human Reference Proteome and known contaminants from UniProt. The identifiers from Proteome Discoverer were mapped to Official Gene Symbols. Non-unique and modified peptides were filtered out before selection of top intensity unique PSMs in the reference channel for each 10-plex run. Ratios were then obtained by dividing the 9 sample channels with the reference channel and median centered. The intensities were then log2 transformed and then rolled into protein groups by taking the median log2 ratio of unique peptide groups in each run.

Multivariate Data Analysis

Data matrices from all omics (transcriptomics, MS-based proteomics, metabolomics, microbiome 16S data and WGS, Proseek Olink and Luminex cytokines) were merged into a common format. All data (excluding RNA-seq for reasons discussed below) was log-transformed and then a linear model was fitted for each individual analyte where the significance of the factors in the model were evaluated using ANOVA. For transcriptome, we used edgeR, a Bioconductor package specifically designed for modeling count data such as RNA-seq, and for improved variance estimation for overdispersed count data. We fit generalized linear models (GLMs) and performed ANOVA-like tests for main effects and interactions using the glmFit and glmLRT functions from the edgeR package (Robinson et al., 2010). The most basic model was used to identify analytes that differ between the two groups (insulin resistant and insulin sensitive):

log2Int ~ sample_group + sample_timepoint + sample_group:sample_timepoint

Note that this model compares *average* differences between groups. In order to represent the experimental conditions accurately, a factor for the timepoint and an interaction term between the groups and timepoints was introduced which allows the IR and IS group to be different at each timepoint, thus allowing us to capture when the two groups react differently to the treatment. Note that this model was only used to estimate group differences, except for transcriptome, where we used it to estimate timepoint differences as well.

In order to estimate timepoint differences accurately, we used a model that allows each subject to have a different baseline level for each analyte:

log2Int ~ patient_id + sample_timepoint + sample_group:sample_timepoint

By correcting for a personalized baseline, the model allows us to capture common differences between timepoints with higher power as people may start out at different baseline levels but move all in the same direction during perturbation - which would be hard to capture using averages only. It contains a term for the patient_id which estimates an average analyte level for each patient. It contains a term for the sample_timepoint which means the model estimates a different offset for each timepoint and it contains an interaction term which means that the model allows the IR and IS group have different relative analyte levels at each timepoint, thus allowing us to capture when the two groups react differently to the treatment. Note that by estimating a different offset for each patient, it is not possible any more to detect differences between the IR and IS groups which is why we used the first model for this purpose.

Finally, we speculated that using BMI as a continuous predictor would potentially be more informative than taking sampling timepoints which do not capture the amount of weight gained or lost by individuals. We indeed see that (Figure S xxx delta_BMI.pdf) weight gain was not consistent across all individuals, even though all individuals gained weight and subsequently lost weight. By relating the analyte levels directly to the amount of weight gained, we wanted to obtain quantitative insights into changes associated with weight gain and loss:

log2Int ~ patient_id + delta_BMI + sample_group:delta_BMI

For transcriptome data, we used the model:

log2Int ~ sample_group + delta_BMI + sample_group:delta_BMI

After model fitting and ANOVA analysis, the resulting p-values were corrected using the Benjamini-Hochberg method and a cutoff of 0.2 was used for all data. The reduction in variance by each factor was used for the variance decomposition analysis.

For longitudinal pattern recognition analysis across the weight-gain/weight-loss perturbation we used the mfuzz soft-clustering algorithm (Futschik and Carlisle, 2005; Kumar and Futschik, 2007). The mean expression/abundance was calculated for all biomolecules comprising the transcriptome, metabolome and proteome at each timepoint, and these values were subsequently standardized to z-scores (mfuzz::standardise). Soft clustering was run with fuzzifier parameter set to m=2.5. Cluster members at a low-stringency acore filter of 0.2 were used in subsequent enrichment analyses. Pathway enrichment analysis was conducted using multiple tools: DAVID for transcripts and proteins (Huang da et al., 2009b), and Metaboanalyst 3.0 for metabolites and transcripts (Xia and Wishart, 2016).

DATA AND SOFTWARE AVAILABILITY

Raw data included in this study are hosted on the NIH Human Microbiome 2 project site (http://www.hmp2.org). Data pertaining to this study are classified under the study ID "T2D" and datafiles can be retrieved by querying by subject ID and timepoint (Visit 1-4). Subject IDs included in this study are: ZJTKAE3, ZK112BX, ZK4CK8Y, ZKFV71L, ZKVR426, ZL63I8R, ZL9BTWF, ZLGD9M0, ZLPRB8E, ZLPZS0H, ZLTUJTN, ZM7JY3G, ZMBH10Z, ZMBVNFM, ZVGW5FI, ZVM4N7A, ZVTCAK9, ZW61YGW, ZWCZHHY, ZWFDEY0, ZWHMV5E, ZY1ZKJY, ZY7IW45. As LC-MS raw data consist of multiple samples isobarically tagged, they have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD007859.