Title

1 2 3

4

6

7

8 9

Multiomic Body Mass Index signatures in blood reveal clinically relevant population heterogeneity and variable responses to a healthy lifestyle intervention

5 Authors

Kengo Watanabe¹, Tomasz Wilmanski¹, Christian Diener¹, John C. Earls^{1,2}, Anat Zimmer^{1,†}, Briana Lincoln¹, Jennifer J. Hadlock¹, Jennifer C. Lovejoy¹, Sean M. Gibbons^{1,3,4}, Andrew T. Magis¹, Leroy Hood^{1,5}, Nathan D. Price^{1,2}, and Noa Rappaport^{1,*}

10 Affiliations

- ¹Institute for Systems Biology, Seattle, WA 98109, USA.
- ²Thorne HealthTech, New York, NY 10019, USA.
- ³Department of Bioengineering, University of Washington, Seattle, WA 98195, USA.
 - ⁴eScience Institute, University of Washington, Seattle, WA 98195, USA.
- ⁵Phenome Health, Seattle, WA 98109, USA.
- [†]Present address: Division of Public Health Sciences, Fred Hutchinson Cancer Research Center,
- 17 Seattle, WA 98109, USA.
- 18 *Correspondence to Noa Rappaport (noa.rappaport@isbscience.org)
- 19

14

20 Abstract

21 Multiomic profiling can reveal population heterogeneity for both health and disease states. Obesity 22 drives a myriad of metabolic perturbations in individuals and is a risk factor for multiple chronic 23 diseases. Here, we report a global atlas of cross-sectional and longitudinal changes in 1,111 blood 24 analytes associated with variation in Body Mass Index (BMI), as well as the multiomic associations 25 with host polygenic risk scores and gut microbiome composition, from a cohort of 1,277 individuals enrolled in a wellness program. Machine learning model predictions of BMI from blood multiomics 26 27 captured heterogeneous phenotypic states of host metabolism and gut microbiome composition, better 28 than classically-measured BMI. Moreover, longitudinal analyses identified variable BMI trajectories 29 for different omics measures in response to a healthy lifestyle intervention; metabolomics-inferred BMI decreased to a greater extent than actual BMI, while proteomics-inferred BMI exhibited greater 30 resistance to change. Our analyses further revealed blood analyte-analyte associations that were 31 32 significantly modified by metabolomics-inferred BMI and partially reversed in the metabolically 33 obese population during the intervention. Taken together, our findings provide a blood atlas of the 34 molecular perturbations associated with changes in obesity status, serving as a valuable resource to robustly quantify metabolic health for predictive and preventive medicine. 35

36

37 Introduction

54

55

56

57

58 59

60

61

62

63

64

65

66

67

38 Obesity has been increasing in prevalence over the past four decades in adults, adolescents, and children around most of the world^{1,2}. Many studies have demonstrated that obesity is a major risk 39 40 factor for multiple chronic diseases such as type 2 diabetes mellitus (T2DM), metabolic syndrome (MetS), cardiovascular disease (CVD), and certain types of cancer³⁻⁶. In individuals with obesity, even 41 42 a 5% loss in body weight can improve metabolic and cardiovascular health⁷, and weight loss through 43 lifestyle interventions can reduce the risk for obesity-related chronic diseases⁸. Nevertheless, obesity and its physiological manifestations can vary widely across individuals, necessitating additional 44 45 research to better understand this prevalent health condition.

Most commonly, obesity is quantified using the anthropometric Body Mass Index (BMI), 46 defined as the body weight divided by body height squared [kg m^{-2}]. While BMI does not directly 47 48 measure body composition, BMI correlates well at the population level with direct measurements of 49 body fat percentage using computed tomography (CT), magnetic resonance imaging (MRI), or dual-50 energy X-ray absorptiometry (DXA) (partial Pearson's $r = 0.74 - 0.84)^9$. As an easily calculated and 51 commonly understood measure among researchers, clinicians, and the general public, BMI is widely 52 used for the primary diagnosis of obesity, and changes in BMI are often used to assess the efficacy of 53 lifestyle interventions.

At the same time, there are considerable limitations to BMI as a surrogate measure of health state; e.g., differences in body composition can lead to misclassification of people with a high muscleto-fat ratio (e.g., athletes) as the individual with obesity, and can undervalue metabolic improvements in health following exercise¹⁰. A meta-analysis showed that the common obesity diagnoses based on BMI cutoffs had high specificity but low sensitivity in identifying individuals with excess body fat¹¹. The misclassification is likely due, in part, to the differences in BMI thresholds for obesity across different ethnic populations¹², as well as the existence of a metabolically unhealthy, normal-weight (MUNW) group within the normal BMI class^{13,14}. Likewise, there are health-heterogeneous groups among the individuals with obesity: metabolically healthy obese (MHO) and metabolically unhealthy obese (MUO). While most individuals in the MHO group are not necessarily healthy but simply healthier than individuals in the MUO group¹⁵, the transition from MHO to MUO phenotype may be a preceding step to the development of obesity-related chronic diseases¹⁶. Moreover, this transition is potentially preventable through lifestyle interventions¹⁷. Altogether, BMI is unequivocally useful at the population level, but too crude to capture a variety of heterogeneous metabolic health states.

68 Recent omics studies have demonstrated how blood omic profiles contain information relevant to a wide range of human health conditions; e.g., blood proteomics captured 11 health 69 indicators such as the liver fat measured by ultrasound and the body composition measured by DXA¹⁸, 70 71 while blood metabolomics tended to reflect dietary intake, lifestyle patterns, and gut microbiome profiles^{19,20}. Intriguingly, a machine learning model that was trained to predict BMI using 49 BMI-72 associated blood metabolites captured obesity-related clinical measurements (e.g., insulin resistance, 73 74 visceral fat percentage) better than observed BMI or genetic predisposition for high BMI²¹. Moreover, 75 in a recent study on coronary artery disease, another blood metabolomics-based model of BMI efficiently reflected differences between individuals with or without acute coronary syndrome 76 (ACS)²². Thus, while a single targeted metric (e.g., body composition) or a specific biomarker (e.g., 77 78 leptin, adiponectin²³) provides useful information, multiomic blood profiling has the potential to 79 comprehensively bridge the multifaceted gaps between BMI and heterogeneous physiological states.

80 In this study, we report heterogeneous molecular signatures of obesity by leveraging a cohort 81 of 1,277 individuals with a wealth of phenotypic data, including human genomes and longitudinal 82 measurements of metabolomics, proteomics, clinical laboratory tests, gut microbiomes, physical 83 activity (i.e., wearables), and health/lifestyle questionnaires, and by employing machine learning to 84 predict BMI. Blood-based analytes across all studied omics platforms exhibit strong performance in 85 predicting measured BMI, explaining 48–78% of the variance in our main study cohort. We further 86 show that multiomic phenotyping captures more refined levels of heterogeneity in metabolic states accompanying obesity, which is not apparent when using measured BMI. Moreover, longitudinal 87

89 90 91

92

95

analyses demonstrate variable changes in blood analytes across the studied omics platforms during a healthy lifestyle intervention; i.e., plasma metabolomics exhibited a stronger response to the intervention than measured BMI, while plasma proteomics exhibited a weaker response within a oneyear span. Our findings highlight the utility and translational potential of blood multiomic profiling for investigating the complex molecular phenotypes underlying obesity and weight loss.

93

94 **Results**

Plasma multiomics captured 48–78% of the variance in BMI

To investigate the molecular phenotypic perturbations associated with obesity, we selected a study 96 cohort of 1,277 adults who participated in a scientific wellness program (Arivale)^{20,24–29} and whose 97 98 datasets included coupled measurements of plasma metabolomics, proteomics, and clinical laboratory 99 tests from the same blood draw (Fig. 1a; see Methods). This study design allowed us to directly 100 investigate the similarities and differences between omics platforms with regards to how they reflected the physiological health state of each individual across the BMI spectrum. This cohort was 101 characteristically female (64.3%), middle-aged (mean \pm s.d.: 46.6 \pm 10.8 years), and white (69.7%) 102 (Supplementary Fig. 1a-c, Supplementary Data 1). Based on the World Health Organization (WHO) 103 international standards for BMI cutoffs (underweight: <18.5 kg m⁻², normal: 18.5–25 kg m⁻², 104 overweight: $25-30 \text{ kg m}^{-2}$, obese: $\geq 30 \text{ kg m}^{-2}$)¹², the baseline BMI prevalence was similar among 105 normal, overweight, and obese classes, while only 0.8% of participants were in the underweight class 106 107 (underweight: 10 participants (0.8%), normal: 426 participants (33.4%), overweight: 391 participants (30.6%), obese: 450 participants (35.2%)). 108

109 Leveraging the baseline measurements of plasma molecular analytes (766 metabolites, 274 proteins, and 71 clinical laboratory tests; Supplementary Data 2), we trained machine learning models 110 to predict baseline BMI (i.e., not forecast a future outcome but calculate an out-of-sample outcome) 111 for each of the omics platforms (metabolomics, proteomics, and clinical labs) or in combination 112 113 (combined omics of all metabolomics, proteomics, and clinical labs): metabolomics-based, 114 proteomics-based, clinical labs (chemistries)-based, and combined omics-based BMI (MetBMI, ProtBMI, ChemBMI, and CombiBMI, respectively) models. To address multicollinearity among the 115 116 analytes (Supplementary Fig. 2a) and to obtain predictions for all participants, we applied a tenfold 117 iteration scheme of the least absolute shrinkage and selection operator (LASSO) algorithm with tenfold cross-validation (CV) (Fig. 1a; see Methods). This approach generated ten fitted sparse models 118 for each omics category (Supplementary Data 3) and one single testing (hold-out) set-derived 119 120 prediction from each omics category for each participant. The resulting models retained (i.e., assigned 121 non-zero β -coefficient to) 62 metabolites, 30 proteins, 20 clinical laboratory tests, and 132 analytes across all ten MetBMI, ProtBMI, ChemBMI, and CombiBMI models, respectively, which exhibited 122 low collinearity (Supplementary Fig. 2b, c) as expected from the LASSO algorithm³⁰. In contrast to a 123 model including obesity-related standard clinical measures (i.e., ordinary least squares (OLS) linear 124 regression model with sex, age, triglycerides, high-density lipoprotein (HDL)-cholesterol, low-density 125 lipoprotein (LDL)-cholesterol, glucose, insulin, and homeostatic model assessment for insulin 126 127 resistance (HOMA-IR) as regressors; StandBMI model), each omics-based model demonstrated significantly higher performance in BMI prediction, ranging from out-of-sample $R^2 = 0.48$ 128 (ChemBMI) to 0.70 (ProtBMI) compared to 0.37 (StandBMI) (Fig. 1b, c). The CombiBMI model 129 exhibited the best performance in BMI prediction (out-of-sample $R^2 = 0.78$; Fig. 1c), but the variances 130 explained were not completely additive, suggesting that, although there is a considerable overlap in 131 the signal detected by each omics platform, different omic measurements still contain non-redundant 132 information regarding BMI. Additionally, these results were consistent in sex-stratified models, with 133 134 the exception of male ChemBMI model that tended to exhibit higher performance than StandBMI 135 model without statistical significance (Supplementary Fig. 2d).

136To confirm the generalizability of our results, we investigated an external cohort of 1,834137adults from the TwinsUK registry³¹, whose datasets included serum metabolomics³² and the138aforementioned standard clinical measures (Fig. 1a; see Methods). This external cohort was

demographically distinct from the Arivale cohort (Supplementary Fig. 1d-f, Supplementary Data 1); 139 140 the TwinsUK cohort was overwhelmingly female (96.7%), senior (mean \pm s.d.: 61.4 \pm 9.0 years), and white (99.2%), and consisted of 15 (0.8%), 779 (42.5%), 706 (38.5%), and 334 (18.2%) participants in 141 142 the underweight, normal, overweight, and obese BMI classes, respectively. To manage the differences in the metabolomics panels, we regenerated MetBMI models in the Arivale cohort, while restricting 143 144 the metabolomic features to an overlapping set of 489 metabolites between the Arivale and TwinsUK 145 panels (called restricted model). Although 25 of the retained metabolites in the original MetBMI models were replaced with other metabolites due to their absences in the restricted panel, 35 of the 146 147 remaining 37 metabolites were consistently retained across the restricted MetBMI models 148 (Supplementary Fig. 3a). Moreover, β -coefficients for the retained metabolites and MetBMI predictions for the Arivale cohort were consistent between the original and restricted models 149 (Supplementary Fig. 3b, c). We then calculated BMI predictions for the TwinsUK cohort using the 150 StandBMI and restricted MetBMI models that were fitted to the Arivale datasets. The restricted 151 152 MetBMI model exhibited a lower absolute performance on the TwinsUK cohort compared to the Arivale cohort, but a significantly higher performance than StandBMI model (out-of-sample $R^2 = 0.30$ 153 (MetBMI), -0.13 (StandBMI); Fig. 1d, Supplementary Fig. 3d), confirming that blood metabolomics 154 generally captures BMI better than the standard clinical measures. 155

BMI has been reported to be associated with multiple anthropometric and clinical measures, 156 157 such as waist circumference (WC), blood pressure, sleep quality, and several polygenic risk scores (PRSs)^{3,4,15,27,33}. Thus, we examined the association between the omics-inferred BMI and each of the 158 available numeric physiological measures (see Methods; Supplementary Data 4). Among the 51 159 assessed features, measured BMI was significantly associated with 27 features (false discovery rate 160 (FDR) < 0.05 including daily physical activity measures from wearable devices, waist-to-height ratio 161 162 (WHtR), blood pressure, and BMI PRS (Fig. 1e). With minor differences in effect sizes, these BMIassociated features were concordantly associated with each omics-inferred BMI (Fig. 1e), indicating 163 164 that the omics-inferred BMIs primarily maintain the characteristics of classical BMI in terms of anthropometric, genetic, lifestyle, and physiological associations. 165

166 167

Omics-based BMI estimates captured the variation in BMI better than any single analyte

168 Because our LASSO linear regression model showed comparable performance to elastic net (EN) and ridge linear regression models and a non-linear random forest (RF) regression model (Supplementary 169 Fig. 4a, b) and because LASSO model β -coefficients are generally easier to be interpreted, we chose 170 to focus on the LASSO models. However, the LASSO algorithm randomly retains variables from 171 172 highly collinear groups, and sets β -coefficients of the other variables to zero. To confirm the robustness of the variable selection process, we iterated the LASSO modeling while removing the 173 strongest analyte (i.e., the analyte that had the highest absolute value for the mean of the ten β -174 175 coefficients) from the input omic dataset at the end of each iteration. If a variable is indispensable for a model, the performance should largely decrease after removing it. In all omics categories, a steep 176 177 decay in the out-of-sample R^2 was observed in the first 5–9 iterations (Supplementary Fig. 2e–h), suggesting that, at least, the top 5–9 variables that had the highest absolute β -coefficient values in the 178 original LASSO models were indispensable for predicting BMI. Interestingly, the overall slope of R^2 179 in MetBMI model decayed more gradually compared to ProtBMI and ChemBMI models 180 181 (Supplementary Fig. 2e-g), implying that metabolomics data contain more redundant information 182 about BMI than the other omics data. Although larger number of metabolites in the input dataset might be a plausible explanation, the proportion of the variables that were robustly retained across all 183 ten LASSO models (Supplementary Fig. 5) to the variables that were retained in at least one of the ten 184 LASSO models was lower in MetBMI model compared to ProtBMI and ChemBMI models (MetBMI: 185 186 62/209 metabolites \approx 30%, ProtBMI: 30/74 proteins \approx 41%, ChemBMI: 20/41 clinical laboratory tests \approx 49%), confirming the higher level of redundancy within metabolomics data. Nevertheless, 187 metabolites still constituted 58% of the 132 analytes that were retained across all ten CombiBMI 188 189 models (77 metabolites, 51 proteins, 4 clinical laboratory tests; Fig. 2a), suggesting that each of the omics categories possesses unique information about BMI. The strongest predictors in CombiBMI 190

191 model were primarily proteins; e.g., analytes having the mean absolute β -coefficient > 0.02 (i.e., affecting more than ~2% BMI in prediction per 1 s.d. of its change, according to the Taylor/Maclaurin 192 series: $e^{\beta} \approx 1 + \beta$ when $\beta << 1$) were leptin (LEP), adrenomedullin (ADM), and fatty acid-binding 193 194 protein 4 (FABP4) as the positive predictors and insulin-like growth factor-binding protein 1 (IGFBP1) and advanced glycosylation end-product specific receptor (AGER; also described as 195 196 receptor of AGE, RAGE) as the negative predictors. Note that these strongest proteins were consistent 197 in the EN models (Supplementary Fig. 4c-f) and had high importance in the ridge and RF models 198 (Supplementary Fig. 4g, h).

199 At the same time, the existence of these strong and consistently-retained predictors in the 200 omics-based BMI models implied that a single analyte might be a suitable biomarker to predict BMI. To address this possibility, we regressed BMI independently on each of the analytes that were retained 201 in at least one of the ten LASSO models (MetBMI: 209 metabolites, ProtBMI: 74 proteins, 202 203 ChemBMI: 41 clinical laboratory tests; Supplementary Data 5). Among the analytes that were 204 significantly associated with BMI (180 metabolites, 63 proteins, 30 clinical laboratory tests), only 205 LEP, FABP4, and interleukin 1 receptor antagonist (IL1RN) exhibited over 30% of the explained variance in BMI by themselves (Fig. 2b-d), with a maximum of 37.9% variance explained (LEP). In 206 207 contrast, MetBMI, ProtBMI, and ChemBMI models explained 68.9%, 70.6%, and 48.8% of the 208 variance in BMI, respectively. Moreover, even upon eliminating several strong predictor analytes such 209 as LEP and FABP4 from the omic datasets, the models still explained more variance in BMI than any single analyte (Supplementary Fig. 2e-h). These results indicate that the multiomic BMI prediction 210 models explain a larger portion of the variation in BMI than any single analyte, and highlight the 211 multivariate perturbation of blood analytes across all platforms with increasing BMI. 212

214Metabolic heterogeneity was responsible for the high rate of misclassification within the215standard BMI classes

213

While the omics-inferred BMIs showed the similar phenotypic associations as the measured BMI (Fig. 216 1e), we observed that the difference of the predicted BMI from the measured BMI (Δ BMI) was highly 217 correlated among the omics-based BMI models, ranging from Pearson's r = 0.64 (ChemBMI vs. 218 CombiBMI) to 0.83 (ProtBMI vs. CombiBMI) (Fig. 3a). In other words, the different omics 219 220 consistently detected deviation of the omics-inferred BMI from the measured BMI per individual, 221 implying that this deviation stemmed from a true biological signal of a perturbed physiological state 222 rather than from noise or modeling artifacts. Actually, when individuals in the normal and obese BMI classes (defined by the WHO international standards) were subdivided by a clinical definition of 223 metabolic health (i.e., defining metabolically unhealthy if having two or more MetS risks of the 224 225 National Cholesterol Education Program (NCEP) Adult Treatment Panel III (ATP III) guidelines; see Methods)^{34,35}, Δ BMI was significantly higher in MUNW and MUO groups compared to metabolically 226 healthy, normal-weight (MHNW) and MHO groups, respectively, for all omics categories (Fig. 3b), 227 suggesting that the deviations of model predictions are related to metabolic health. 228

Nevertheless, there has been no universally accepted definition of metabolic health^{14,15,34,35}. 229 230 Thus, given the high interpretability and intuitiveness of the omics-inferred BMI, we further explored 231 a potential application: using the omics-inferred BMI (instead of the measured BMI) for improved classification of both obesity and metabolic health with the WHO international standards. Each 232 233 participant was classified using each of the measured and omics-inferred BMIs based on the standard BMI cutoffs, and categorized into either Matched or Mismatched group when the measured BMI class 234 235 was matched or mismatched to each omics-inferred BMI class, respectively. The misclassification rate against the omics-inferred BMI class was $\sim 30\%$ across all omics categories and BMI classes (Fig. 3c), 236 consistent with the previously reported misclassification rates about the cardiometabolic health 237 classification^{36,37}. We then examined relationships between this omics-based misclassification within 238 239 normal or obese BMI class and the obesity-related clinical blood markers (Supplementary Data 6), 240 including triglycerides, HDL-cholesterol, LDL-cholesterol, high-sensitivity C-reactive protein (hs-CRP), glucose, insulin, HOMA-IR, glycated hemoglobin A1c (HbA1c), adiponectin, and vitamin 241

D^{3,15,23,38,39}. Because ChemBMI and CombiBMI models were not independent of these markers, only 242 243 the misclassification against MetBMI or ProtBMI class was examined in this analysis. The 244 Mismatched group of normal BMI class exhibited significantly higher values of the markers that are 245 positively associated with BMI (+BMI), such as triglycerides, hs-CRP, glucose, and HOMA-IR, and significantly lower values of the markers that are negatively associated with BMI (-BMI), such as 246 HDL-cholesterol and adiponectin, compared to the Matched group of normal BMI class (FDR ≤ 0.05 ; 247 Fig. 3d). These patterns suggest that the participant misclassified into the normal BMI class possesses 248 less healthy molecular profiles as similarly as the individual with overweight or obesity, 249 250 corresponding to the individual with MUNW phenotype. Conversely, the Mismatched group of obese BMI class exhibited significantly lower and higher values of the positively and negatively BMI-251 252 associated markers, respectively, compared to the Matched group of obese BMI class (FDR < 0.05; Fig. 3d), suggesting that the participant misclassified as obese BMI class has healthier blood 253 signatures, more similarly to the individual with overweight or normal-weight, corresponding to the 254 255 individual with MHO phenotype. Likewise, we re-examined the 27 BMI-associated numeric physiological features (Fig. 1e, Supplementary Data 6), and found the concordant pattern of 256 257 significant phenotypic differences between Matched and Mismatched groups in WHtR (+_{BMI}), heart 258 rate ($+_{BMI}$), blood pressure ($+_{BMI}$), and daily physical activity measures ($-_{BMI}$) (FDR < 0.05; Fig. 3e). Importantly, there was no difference in BMI PRS (+BMI) between Matched and Mismatched groups 259 (Fig. 3e), implying that lifestyle or environmental factors, rather than genetic risk, is likely involved in 260 the discordance between the measured and omics-inferred BMIs. Furthermore, we validated and 261 262 expanded these findings in the TwinsUK cohort: Δ MetBMI was significantly higher in the metabolically unhealthy group compared to the metabolically healthy group within the normal BMI 263 class (Supplementary Fig. 6a); the misclassification rate against MetBMI class was much higher 264 265 (>60%) in the normal BMI class but $\sim 30\%$ in the others (Supplementary Fig. 6b); the concordant 266 phenotypic differences between Matched and Mismatched groups were significantly observed in triglycerides ($+_{BMI}$), HDL-cholesterol ($-_{BMI}$), LDL-cholesterol ($+_{BMI}$), hs-CRP ($+_{BMI}$), and HOMA-IR 267 $(+_{BMI})$ (FDR < 0.05; Supplementary Fig. 6c). Remarkably, while DXA measurements were not 268 269 performed in the Arivale cohort, the percentage of total fat in whole body $(+_{BMI})$ and the ratio of fat in 270 android region to fat in gynoid region ($+_{BMI}$) were significantly higher in Mismatched group compared to Matched group within the normal BMI class of the TwinsUK cohort (FDR < 0.05; Supplementary 271 272 Fig. 6c). Taken together, these results suggest that the omics-based BMI models can identify 273 heterogeneous metabolic health states which are not captured by the measured BMI with the standard 274 BMI cutoffs. 275

276

Metabolomics-inferred BMI reflected gut microbiome profiles better than BMI

The gut microbiome has been shown to causally affect host obesity phenotypes in a mouse model⁴⁰. 277 and humans with obesity generally exhibit lower bacterial α -diversity (i.e., the species richness and/or 278 evenness of an ecological community)^{41,42}. However, certain meta-analyses of human case-control 279 studies suggest an inconsistent relationship between the gut microbiome and obesity^{43,44}. Given our 280 previous finding that the association between blood metabolites and bacterial diversity is dependent on 281 BMI²⁰ and the current finding that the omics-based BMI models capture heterogeneous metabolic 282 health states (Fig. 3), we hypothesized that MetBMI represents gut microbiome α -diversity better than 283 the measured BMI. For the 702 Arivale participants who had both stool-derived gut microbiome and 284 285 blood omic datasets (Fig. 4a; see Methods), we examined relationships between gut microbiome α -286 diversity (the number of observed species, Shannon's index, and Chaol index) and the omics-based 287 BMI misclassification. Matched and Mismatched groups against MetBMI class showed significant differences in all α -diversity metrics within both normal and obese BMI classes (Fig. 4b), with the 288 289 concordant pattern to the clinical markers and BMI-associated features ($-_{BMI}$; e.g., HDL-cholesterol; Fig. 3d, e), implying that the MetBMI class reflects bacterial diversity better than BMI class. 290 Interestingly, the misclassification against the other omics categories did not show these significant 291 292 differences for all α -diversity metrics and both BMI classes (Fig. 4b), consistent with our previous

293 observation that plasma metabolomics showed a much stronger correspondence to gut microbiome 294 structure than either proteomics or clinical labs²⁰.

We further examined the predictive power of gut microbiome profiles for MetBMI. For each 295 296 of the measured BMI and MetBMI classes, we generated models classifying individuals into normal 297 class versus obese class based on gut microbiome 16S rRNA gene amplicon sequencing data, using a 298 fivefold iteration scheme of the RF algorithm with fivefold CV (Fig. 4a; see Methods). Compared to 299 the classifier for the measured BMI class, the classifier for MetBMI class showed significantly larger area under curve (AUC) in the receiver operator characteristic (ROC) curve in the Arivale cohort 300 (AUC = 0.66 (BMI), 0.75 (MetBMI); Fig. 4c), with significantly higher sensitivity and precision (Fig. 301 4d). Moreover, by applying the same scheme to the stool-derived whole metagenomic shotgun 302 sequencing (WMGS) data of the 329 TwinsUK participants⁴⁵ (Fig. 4a; see Methods), we validated that 303 the gut microbiome-based obesity classifier for MetBMI class significantly outperformed the classifier 304 for the measured BMI class in the TwinsUK cohort (AUC = 0.57 (BMI), 0.75 (MetBMI); Fig. 4e, f). 305 306 Note that these classifiers were regenerated for the TwinsUK cohort (instead of using the classifiers that were fitted to the Arivale dataset; Fig. 4a) due to the difference in sequencing methods (amplicon 307 308 sequencing vs. WMGS), while considering that the TwinsUK participants' MetBMIs were predicted 309 from the Arivale-fitted MetBMI models (Fig. 1a). Altogether, these findings suggest that, although other factors (e.g., dietary intake¹⁹) may be involved, MetBMI has a stronger correspondence to gut 310 311 microbiome features than the standard BMI.

312

313Metabolic health of the metabolically obese group was substantially improved following a314healthy lifestyle intervention

315 In the Arivale program, healthy lifestyle coaching was provided to all participants, resulting in clinical improvement across multiple measures of health²⁵. This coaching intervention was personalized for 316 317 each participant to improve the participant's health based on the combination of clinical laboratory tests, genetic predispositions, and published scientific evidence, and administered via telephone by 318 registered dietitians, certified nutritionists, or registered nurses (see Methods and a previous report²⁵). 319 320 To investigate the longitudinal changes in omic profiles during the program, we defined a sub-cohort of 608 participants based on the available longitudinal measurements (Fig. 5a; see Methods). Given 321 322 the participant-dependent variability in both count and time point of data collections, we estimated the average trajectory of each measured or omics-inferred BMI in the Arivale sub-cohort using a linear 323 mixed model (LMM) with random effects for each participant (see Methods). Consistent with the 324 previous analysis^{25,46}, the mean BMI estimate for the overall cohort decreased during the program 325 (Fig. 5b). The decrease of MetBMI was larger than that of measured BMI, while the decrease of 326 ProtBMI was minimal and even smaller than that of measured BMI (Fig. 5b), suggesting that plasma 327 metabolomics is highly responsive to the lifestyle intervention in the short term, while proteomics 328 329 (measured from the same blood draw) is more resistant to change during the same intervention period. Subsequently, we generated LMMs with the baseline BMI class stratification, and confirmed that a 330 331 significant decrease in the mean BMI estimate was observed in the overweight and obese BMI classes, 332 but not in the normal BMI class (Fig. 5c). Concordantly, the mean estimates of ProtBMI and 333 ChemBMI exhibited negative changes over time in the overweight and obese BMI classes, but not in 334 the normal BMI class (Fig. 5c). In contrast, the mean estimate of MetBMI exhibited a significant 335 decrease across all BMI classes (Fig. 5c), suggesting that metabolomics data captures information 336 about the metabolic health response to the lifestyle intervention, beyond the baseline BMI class and 337 the changes in BMI and other omic profiles.

Given the existence of multiple metabolic health sub-states within the standard BMI classes (Fig. 3), we further investigated the difference between misclassification strata against the baseline MetBMI class. In the (baseline) normal BMI class, while the mean estimate of the measured BMI remained constant in both Matched and Mismatched groups, the mean MetBMI estimate exhibited larger reduction in Mismatched group than Matched group (Fig. 5d), suggesting that the participants with MUNW phenotype improved their metabolic health to a greater extent than the participants with

MHNW phenotype. Likewise, in the (baseline) obese BMI class, while the decrease in the mean 344 345 estimate of the measured BMI was not significantly different between Matched and Mismatched groups at one year after the enrollment, the decrease in the mean MetBMI estimate was larger in 346 Matched group than in Mismatched group (Fig. 5e), suggesting that the participants with MUO 347 phenotype improved their metabolic health to a greater extent than the participants with MHO 348 349 phenotype. Altogether, these results suggest that metabolic health was substantially improved during the program, in accordance with an individual's baseline metabolomic state, rather than with the 350 individual's baseline BMI class. 351

352

Plasma analyte correlation network in the metabolically obese group shifted toward a structure observed in metabolically healthier state following a healthy lifestyle intervention

We explored longitudinal changes in plasma analyte correlation networks, focusing on the 355 356 metabolically obese group. Based on the importance of the baseline metabolomic state (Fig. 5d, e), we first assessed relationships between each plasma analyte-analyte correlation and the baseline MetBMI 357 within the Arivale sub-cohort (Fig. 5a; 608 participants), using their interaction term in a generalized 358 linear model (GLM; see Methods) of each analyte-analyte pair. In this type of model, the statistical 359 test assesses whether the relationship between any two analytes is dependent on a third variable (in 360 this case, the baseline MetBMI). Among 608,856 pairwise relationships of plasma analytes, 100 361 362 analyte-analyte correlation pairs, comprising 82 metabolites, 33 proteins, and 16 clinical laboratory 363 tests, were significantly modified by the baseline MetBMI (FDR < 0.05; Supplementary Data 7). Subsequently, we assessed longitudinal changes of these 100 pairs within the metabolically obese 364 group (i.e., the baseline obese MetBMI class; 182 participants), using the interaction term (i.e., 365 366 interaction with days in the program) in a generalized estimating equation (GEE; see Methods) of each analyte-analyte pair. Among the 100 pairs, 27 analyte-analyte correlation pairs were significantly 367 modified by days in the program (FDR < 0.05; Fig. 6a, Supplementary Data 7). These 27 pairs were 368 mainly derived from metabolites (21 metabolites, 3 proteins, 3 clinical laboratory tests). One of these 369 time-varying pairs was homoarginine and phenyllactate (PLA). Homoarginine was recently found to 370 371 be a biomarker for CVD⁴⁷ and was a robustly retained positive predictor in MetBMI and CombiBMI models (Fig. 2a, Supplementary Fig. 5a). PLA is a gut microbiome-derived phenylalanine derivative 372 known to have antimicrobial activity and antioxidant activity^{48,49}. The positive correlation between 373 homoarginine and PLA was observed in the metabolically obese group at baseline (Fig. 6b) and 374 375 became weaker in this group during the course of the intervention (Fig. 6c), implying that metabolic dysregulation specific to the metabolically obese group was somewhat improved during the program. 376 377 Collectively, these findings indicate that metabolic improvement was not limited to changes in 378 specific blood analyte concentrations but also changes in the association structure among analytes. 379

380 Discussion

Obesity is a significant risk factor for many chronic diseases³⁻⁶. The heterogeneous nature of human 381 382 health conditions, with variable manifestation ranging from metabolic abnormalities to cardiovascular 383 symptoms, calls for deeper molecular characterizations in order to optimize wellness and reduce the current global epidemic of chronic diseases. In this study, we have demonstrated that obesity 384 385 profoundly perturbs human physiology, as reflected across all the studied omics modalities. The key 386 findings of this study are: (1) machine learning-based multiomic BMI estimates were better suited to identifying heterogeneous metabolic health than the classically-measured BMI, while maintaining a 387 high level of interpretability and intuitiveness attributed to the original metric (Fig. 1–3); (2) among 388 all omics studied, metabolomic reflection of obesity exhibited the strongest correspondence to gut 389 microbiome community structure (Fig. 4); (3) plasma metabolomics exhibited the strongest (and/or 390 391 earliest) response to lifestyle coaching, while plasma proteomics exhibited a weaker (and/or more 392 delayed) response than the measured BMI (Fig. 5b, c); (4) compared to the participants with 393 metabolically healthy phenotype (i.e., BMI class = MetBMI class), the participants with metabolically 394 unhealthy phenotype (i.e., BMI class < MetBMI class) exhibited a greater improvement in their

metabolic health (but not in weight loss per se) in response to the healthy lifestyle coaching (Fig. 5d,
e); (5) dozens of analyte–analyte associations were modified in the participants of the metabolically
obese group (i.e., obese MetBMI class), following the healthy lifestyle intervention (Fig. 6).

398 Although BMI is used as a measure of obesity, fat distribution in the body is an important 399 factor for understanding the heterogeneous nature of obesity. In particular, abdominal obesity, which is characterized by excessive visceral fat (rather than subcutaneous fat) around the abdominal region. 400 is known to be associated with chronic diseases such as MetS⁵⁰. Thus, we addressed abdominal 401 obesity by analyzing the anthropometric WHtR^{51,52}, which was highly correlated with BMI in the 402 Arivale sub-cohort (Pearson's r = 0.86; Supplementary Fig. 7a–c). We generated omics-based WHtR 403 404 models (Supplementary Fig. 7a, Supplementary Data 8), and obtained consistent findings to the omics-based BMI models (Supplementary Fig. 7d-m). Interestingly, the majority of the retained 405 406 analytes in each omics-based WHR model was also retained in its corresponding omics-based BMI 407 model with the similar feature importance (Supplementary Fig. 8a–d). In addition, Δ WHtR was highly 408 correlated with Δ BMI across all omics categories (Supplementary Fig. 8e). Moreover, although the WC measurements were not available for the defined TwinsUK cohort, direct fat measurements of the 409 android region by DXA were associated with MetBMI class in the TwinsUK cohort (Supplementary 410 411 Fig. 6c). Therefore, although BMI requires complementary information of the WC-related measurements for the diagnosis of abdominal obesity, the omics-based BMI model likely captures the 412 413 obesity characteristics including abdominal obesity.

Multiple observational studies have explored obesity biomarkers. The involvements of 414 insulin/insulin-like growth factor (IGF) axis and chronic low-grade inflammation have been discussed 415 in the context of obesity-related disease risks^{5,6}, backed up by robust associations of obesity with 416 IGFBP1/2 ($-_{BMI}$), adipokines such as LEP ($+_{BMI}$), adiponectin ($-_{BMI}$), FABP4 ($+_{BMI}$), and ADM 417 418 $(+_{BMI})$, and proinflammatory cytokines such as interleukin 6 (IL6; $+_{BMI})^{23,53}$. Consistent with these well-known associations, we observed positive BMI associations with LEP, FABP4, IL1RN, IL6, 419 420 ADM, and insulin and negative BMI associations with IGFBP1/2 and adiponectin (Fig. 2c, d). 421 Importantly, all these known biomarkers were incorporated into our omics-based BMI models, and 422 most of them were consistently retained as important features of these models (Fig. 2a; Supplementary 423 Fig. 5b, c). At the same time, we observed that RAGE explained a relatively small proportion of the variance in BMI (Fig. 2c), while being a strong negative predictive feature in all ten models of 424 ProtBMI and CombiBMI (Fig. 2a, Supplementary Fig. 5b). Soluble RAGE (sRAGE) has been 425 gradually highlighted in the contexts of T2DM and CVD⁵⁴, with several reports on the negative 426 association between sRAGE and BMI⁵⁵. Therefore, omics-inferred BMI may reflect not only obesity 427 428 status but also the early transition towards clinical manifestations of obesity-related chronic diseases.

429 Likewise, many epidemiological studies have revealed metabolomic biomarkers for obesity^{56,57}. In line with these previous findings, we have confirmed positive BMI associations with 430 431 mannose, uric acid (urate), and glutamate and negative BMI associations with asparagine and glycine (Fig. 2b). Furthermore, all of these metabolites were consistently incorporated into all ten models of 432 433 MetBMI and CombiBMI (Fig. 2a, Supplementary Fig. 5a). In addition, many lipids emerged as strong predictors in MetBMI and CombiBMI models; in particular, glycerophosphocholines (GPCs) were 434 435 negative predictors in these models, while sphingomyelins (SMs) were positive predictors (Fig. 2a, Supplementary Fig. 5a), even though both have a phosphocholine group in common. Although lipid 436 437 has traditionally been regarded as a factor that is positively associated with obesity, recent 438 metabolomics studies have revealed variable trends for different fatty acid species; e.g., plasma lysophosphatidylcholines (LPCs) are decreased in mice with obesity (high-fat diet model)⁵⁸, which 439 440 corresponded well with our results (e.g., LPC(18:1), described as 1-oleoyl-GPC(18:1) in Fig. 2b and Supplementary Fig. 5b). However, because there are many combinations of acyl residues in lipids and 441 442 many potential confounding factors with obesity, systematic understanding of the species-level lipid biomarkers for obesity remains challenging^{56,57}. Our approach, applying machine learning to 443 metabolomics data, addresses this challenge by automatically and systematically providing a 444 445 molecular signature of obesity, reflecting the versatile and complex metabolite species. Altogether,

omics-based BMI models can be regarded as multidimensional profiles of obesity, possessing detailed
 mechanistic information.

448 Recently, Cirulli and colleagues have reported a machine learning model for estimating BMI 449 from blood metabolomics, which captured obesity-related phenotypes²¹. Their main model explained 450 39.1% of the variance in BMI, while our MetBMI model explained 68.9% of the variance in BMI (Fig. 2b). Other than the difference in cohorts, the performance gap is likely a result of differences in 451 452 modeling strategies. Cirulli and colleagues stringently selected 49 metabolites, out of their metabolomics panel of 1,007 metabolites, based on a pre-screening for significant adjusted-453 associations with BMI, and subsequently applied a tenfold CV implementation of ridge or LASSO 454 455 method. In contrast, we used LASSO method for feature selection, applying it to our full 456 metabolomics panel of 766 metabolites. In addition to the increased number of metabolites included in 457 the model fitting, our higher performance may stem from the presence of metabolites which were 458 critical for BMI prediction in a multivariate model, but not strongly associated with BMI on their own. 459 Actually, similarly to the above example of RAGE in ProtBMI model, our MetBMI model contained multiple metabolites that were weakly associated with BMI but consistently retained across all ten 460 models (Fig. 2b, Supplementary Fig. 5a). At the same time, the majority of the 49 metabolites reported 461 462 by Cirulli and colleagues (14–20 metabolites among the 31–41 corresponding metabolites in our 463 metabolomics panel) were retained in at least one of the ten MetBMI models. Therefore, our strategy 464 of feature selection through machine learning, without a pre-filtering step, may be preferable for predicting BMI from metabolomics. 465

A recent study investigating multiomic changes in response to weight perturbations 466 demonstrated that some weight gain-associated blood signatures were reversed during subsequent 467 weight loss, while others persisted⁵⁹. Interestingly, we found that MetBMI was more responsive to the 468 healthy lifestyle intervention than the measured BMI or ChemBMI, while ProtBMI was more resistant 469 to the same intervention (Fig. 5b, c). Our analyses of the predictors in the omics-based BMI models 470 (Fig. 2; Supplementary Fig. 2e-h, 5) suggested that the distribution of feature importance among 471 472 metabolites was considerably wider, while only a small subset of measured proteins (~5 proteins) was 473 predominantly reflective of obesity profiles. Therefore, the effect of lifestyle coaching may consist of 474 small additive contributions in blood metabolites in the short term. However, a longer longitudinal 475 analysis is needed to infer the physiological meaning of these omics-dependent dynamics. For 476 instance, it is possible that ProtBMI shows a delayed response to weight loss (over a span greater than 477 a year measured presently; Fig. 5b, c), indicating blood metabolites and proteins may be early and late responders to a lifestyle intervention, respectively, such as in the case of the changes in blood glucose 478 compared to the changes in HbA1c when assessing glucose homeostasis⁶⁰. If the difference between 479 the measured and omics-inferred BMIs remains constant even after one year, we would conclude that 480 481 blood metabolites and proteins are more and less sensitive to weight loss than the measured BMI, 482 respectively. In either scenario, monitoring blood multiomics during weight loss programs could help 483 participants maintain their motivation to stay engaged with persistent lifestyle changes, because they 484 would receive rapid feedback on how lifestyle changes were impacting their health, even in the absence of weight loss. In addition, long-term maintenance of the improvement is an important 485 486 challenge for lifestyle interventions; although there is variability between prior reports, one study estimated that only ~20% of the individuals with overweight successfully maintain their weight loss in 487 post-intervention⁶¹. Despite this relatively low rate of long-term success, there is evidence that 488 lifestyle interventions had benefits in preventing diabetes incidence as far as 20 years post-489 intervention, even if weight was regained^{62,63}. The observed larger improvement of MetBMI 490 compared to the measured BMI could potentially contribute to this protective long-term effect, 491 492 persisting even when weight is regained. Further investigation is required, especially with regard to 493 the long-term dynamics of MetBMI and ProtBMI responses, which may provide a foothold in 494 developing scientific strategies aimed at long-term maintenance of metabolic health.

495Despite a number of highly promising findings, there were several limitations to our study.496For example, this study was not designed as a randomized control trial, and we cannot strictly evaluate497the effectiveness of the lifestyle intervention (e.g., bigger improvements in the obese group compared

498	to the normal-weight group may be due to the regression-toward-the-mean effect ⁴⁶). In addition, we
499	used time as the variable in longitudinal analyses under an assumption that the program enrollment
500	itself affected participant's BMI and omic profiles. However, if we had more detailed data on the
501	intervention (e.g., magnitude, participants' compliance), we would be able to improve the assessment
502	of its effect. The generalizability of our findings may be limited, because this study was an
503	observational study of largely Caucasian cohorts from the Pacific West of the U.S. and from the U.K.
504	and because validation with an external cohort relied on the female-dominated cohort (96.7%) and its
505	metabolomics data. Our measurements did not cover all biomolecules in blood; in particular,
506	proteomics was based on three targeted Olink panels. Thus, our findings on metabolomic and
507	proteomic states are restricted to the analytes that we could measure. Nevertheless, this study will
508	serve as a valuable resource for robustly characterizing metabolic health from the blood and
509	identifying actionable targets for health management.
510	

511 Methods

512 Study cohort

513 The main study cohort (Arivale cohort) was derived from 6.223 individuals who participated in a wellness program offered by a currently closed commercial company (Arivale Inc., Washington, 514 515 USA) between 2015–2019. An individual was eligible for enrollment if the individual was over 18 years old, not pregnant, and a resident of any U.S. state except New York; participants were primarily 516 recruited from Washington, California, and Oregon. The participants were not screened for any 517 518 particular disease. During the Arivale program, each participant was provided personalized lifestyle 519 coaching via telephone by registered dietitians, certified nutritionists, or registered nurses. This 520 coaching was designed to improve the participant's health based on the combination of clinical laboratory tests, genetic predispositions, and published scientific evidence; e.g., reduction of sodium 521 522 intake might be recommended to any participants with high blood pressure, but if they also had risk 523 alleles indicating enhanced susceptibility to dietary sodium, this risk would be emphasized (see a previous report²⁵ for more details). In the current study, to compare the associations between Body 524 Mass Index (BMI) and host phenotypes across different omics, we limited the original cohort to the 525 526 participants whose datasets contained (1) all main omic measurements (metabolomics, proteomics, 527 clinical laboratory tests) from the same first blood draw, (2) a BMI measurement within ± 1.5 month from the first blood draw, and (3) genetic information (for using as covariates). We also eliminated (1) 528 529 outlier participants whose baseline BMI was beyond ± 3 s.d. from the mean in the baseline BMI distribution and (2) participants whose any of omic datasets contained more than 10% missingness in 530 531 the filtered analytes (see the next section). The final Arivale cohort consisted of 1,277 (821 female and 532 456 male) participants (Fig. 1a), which exhibited consistent demographics (Supplementary Fig. 1a-c, Supplementary Data 1) with the study cohorts defined in the previous Arivale studies^{20,25-29}. For the 533 analyses of gut microbiome, sub-cohort was defined with the 702 (486 female and 216 male) 534 participants from the Arivale cohort, who collected a stool sample within ±1.5 month from the first 535 blood draw and did not use antibiotics in the last three months (Fig. 4a, Supplementary Data 1). For 536 537 longitudinal analyses, sub-cohort was defined with the 608 (410 female and 198 male) participants 538 from the Arivale cohort, whose datasets contained two or more time-series datasets for both BMI and 539 omics during 18 months after enrollment (Fig. 5a, Supplementary Data 1). For the analyses of waist-540 to-height ratio (WHtR), sub-cohort was defined with the 1,078 (689 female and 389 male) participants 541 from the Arivale cohort, whose datasets contained the baseline WHtR measurement within ± 1.5 542 month from the first blood draw and within ± 3 s.d. from the mean in the baseline WHtR distribution (Supplementary Fig. 7a, Supplementary Data 1). 543

The external cohort (TwinsUK cohort) was derived from 17,630 individuals who participated 544 in the TwinsUK Registry, a British national register of adult twins³¹. Twins were recruited as 545 546 volunteers by media campaigns without screening for any particular disease. The participants had two 547 or more clinical visits for biological sampling between 1992–2022. In the current study, to validate our findings in the Arivale cohort, we limited the original cohort to the participants whose datasets 548 contained all measurements for metabolomics³², BMI, and the obesity-related standard clinical 549 550 measures (i.e., defined by triglycerides, high-density lipoprotein (HDL)-cholesterol, low-density 551 lipoprotein (LDL)-cholesterol, glucose, insulin, and homeostatic model assessment for insulin resistance (HOMA-IR) throughout the current study) from the same visit. We also eliminated (1) 552 outlier participants whose BMI was beyond ± 3 s.d. from the mean in the overall BMI distribution and 553 554 (2) participants whose metabolomic dataset contained more than 10% missingness in the filtered 555 metabolites (see the next section). The final TwinsUK cohort consisted of 1,834 (1,774 female and 60 556 male) participants (Fig. 1a, Supplementary Fig. 1d–f, Supplementary Data 1). For the analyses of gut 557 microbiome, sub-cohort was defined with the 329 (307 female and 22 male) participants from the 558 TwinsUK cohort, who collected a stool sample within ± 1.5 month from the clinical visit and did not use antibiotics at that time (Fig. 4a, Supplementary Data 1). 559

560The current study was conducted with de-identified data of the participants who had561consented to the use of their anonymized data in research. All procedures were approved by the

562 Western Institutional Review Board (WIRB) with Institutional Review Board (IRB) (Study Number: 20170658 at Institute for Systems Biology and 1178906 at Arivale) and by the TwinsUK Resource 563 Executive Committee (TREC) (Project Number: E1192). 564

Data collections and data cleaning 566

Multiomics data for the Arivale participants included genomics and longitudinal measurements of 567 metabolomics, proteomics, clinical laboratory tests, gut microbiomes, wearable devices, and 568 569 health/lifestyle questionnaires. Peripheral venous blood draws for all measurements were performed 570 by trained phlebotomists at LabCorp (Laboratory Corporation of America Holdings, North Carolina, 571 USA) or Quest (Quest Diagnostics, New Jersey, USA) service centers. Saliva to measure analytes such as diurnal cortisol and dehydroepiandrosterone (DHEA) was sampled by participants at home 572 using a standardized kit (ZRT Laboratory, Oregon, USA). Likewise, stool samples for gut microbiome 573 574 measurements were obtained by participants at home using a standardized kit (DNA Genotek, Inc., 575 Ottawa, Canada).

576

565

577

578

579

580

581

582 583 584

585

586

587 588

589 590

591

592 593

594

595

596

597

598

599

600

601 602

603

607

608 609

- Genomics

DNA was extracted from each whole blood sample and underwent whole genome sequencing (1,257 participants) or single-nucleotide polymorphisms (SNP) microarray genotyping (20 participants). Genetic ancestry was calculated with principal components (PCs) using a set of ~100,000 ancestry-informative SNP markers, as described previously²⁵. Polygenic risk scores (PRSs) were constructed using publicly available summary statistics from published genomewide association studies (GWAS), as described previously²⁷.

- Blood-measured omics

Metabolomics data was generated by Metabolon, Inc. (North Carolina, USA), using ultrahigh-performance liquid chromatography-tandem mass spectrometry (UHPLC-MS/MS) for plasma derived from each whole blood sample. Proteomics data was generated using proximity extension assay (PEA) for plasma derived from each whole blood sample with several Olink Target panels (Olink Proteomics, Uppsala, Sweden), and only the measurements with the Cardiovascular II, Cardiovascular III, and Inflammation panels were used in the current study since the other panels were not necessarily applied to all samples. All clinical laboratory tests were performed by LabCorp or Quest in a Clinical Laboratory Improvement Amendments (CLIA)-certified lab, and only the measurements by LabCorp were selected in the current study to eliminate potential differences between vendors. In the current study, the batch-corrected datasets with in-house pipeline were used, and metabolomic dataset was \log_{e} -transformed. In addition, analytes missing in more than 10% of the baseline samples were removed from each omic dataset, and observations missing in more than 10% of the remaining analytes were further removed. The final filtered metabolomics, proteomics, and clinical labs consisted of 766 metabolites, 274 proteins, 71 clinical laboratory tests, respectively (Supplementary Data 2).

- Gut microbiome

Gut microbiome data was generated based on 16S amplicon sequencing of the V3+V4 region using a MiSeq sequencer (Illumina, Inc., California, USA) for DNA extracted from each stool 604 sample, as previously described²⁸. Briefly, the FASTQ files were processed using the mbtools 605 606 workflow (https://github.com/Gibbons-Lab/mbtools) to remove noise, infer amplicon sequence variants (ASVs), and remove chimeras. Taxonomy assignment was performed using the SILVA ribosomal RNA gene database (version 132)⁶⁴. In the current study, the final collapsed ASV table across the samples consisted of 394, 341, 85, 45, 26, and 16 taxa for 610 species, genus, family, order, class, and phylum, respectively. Gut microbiome α -diversity

611 612	was calculated at the ASV level using Shannon's index calculated by $H = -\sum_{i=1}^{S} p_i \ln p_i$, where p_i is the proportion of a community <i>i</i> represented by ASVs, or using Chaol diversity
613	score calculated by $S_{\text{Chao1}} = S_{\text{obs}} + \frac{n_1^2}{2n_2}$, where S_{obs} is the number of observed ASVs, n_1 is the
614	number of singletons (ASVs captured once), and n_2 is the number of doubletons (ASVs
615	captured twice).
616	
617	– Anthropometrics, saliva-measured analytes, and daily physical activity measures
618	Anthropometrics including weight, height, and waist circumference (WC) and blood pressure
619	were measured at the time of blood draw and also reported by participants, which generated
620	diverse timing and number of observations depending on each participant. BMI and WHtR
621	were simultaneously calculated from the measured anthropometrics with the weight divided
622	by squared height [kg m ⁻²] and the WC divided by height [unitless], respectively.
623	Measurements of saliva samples were performed in the testing laboratory of ZRT Laboratory.
624	Daily physical activity measures such as heart rate, moving distance, step count, burned
625	calories, floors climbed, and sleep quality were tracked using the Fitbit wearable device
626	(Fitbit, Inc., California, USA). To manage variations between days, monthly averaged data
627	was used for these daily measures. In the current study, the baseline measurement for these
628	longitudinal measures was defined with the closest observation to the first blood draw per
629	participant and data type, and each dataset was eliminated from analyses when its baseline
630	measurement was beyond ± 1.5 month from the first blood draw.
631	Data resource for the TwinsUK participants included longitudinal measurements of metabolomics,
632	clinical laboratory tests, dual-energy X-ray absorptiometry (DXA), and health/lifestyle
633	questionnaires ³¹ . The necessary datasets for the current study were provided by Department of Twin
634	Research & Genetic Epidemiology (King's College London). In the current study, after each provided
635	dataset was cleaned as follows, the earliest visit among the visits from which all of metabolomics,
636	BMI, and the standard clinical measures had been measured was defined as the baseline visit for each
637	participant. As exception, the later visit among them was prioritized as the baseline visit, if the
638	participant had gut microbiome data within ± 1.5 month from the visit. Only the baseline visit
639	measurements were analyzed.
640	- Blood-measured metabolomics
641	Metabolomics data was originally generated by Metabolon, Inc., using UHPLC-MS/MS for
642	each serum sample ³² . In the current study, the provided median-normalized dataset was $\log_{e^{-1}}$
643	transformed. In addition, metabolites missing in more than 10% of the overall samples were
644	removed from metabolomic dataset, and observations missing in more than 10% of the
645	remaining metabolites were further removed. The final filtered metabolomics consisted of
646	683 metabolites.
647	
648	– BMI
649	In the current study, the BMI values that had been already calculated and included in the
650	provided metabolomics data file were used.
651	
652	 Standard clinical measures and other phenotypic measures
653	In the current study, because the provided phenotypic datasets contained multiple
654	measurements for a phenotype even from a single visit of a participant (e.g., due to project
655	difference, repeated measurements), multiple measurements were flattened into a single
<pre>/ - /</pre>	difference, repeated incusatements), maniple incusatements were nationed into a single
656	measurement for a phenotype per each participant's visit by taking the mean value. During
656 657	measurement for a phenotype per each participant's visit by taking the mean value. During this flattening step, difference in unit was properly adjusted, and the value indicating below

661

662

663

664

665 666

667

668

669

670

671

672

673 674 insulin, and fasting condition with the formula: HOMA-IR = fasting glucose [mmol L^{-1}] × fasting insulin [mIU L^{-1}] × 22.5⁻¹.

– Gut microbiome

Gut microbiome data was originally generated based on whole metagenomic shotgun sequencing (WMGS) using a HiSeq 2500 sequencer (Illumina, Inc.) for DNA extracted from each stool sample⁴⁵. In the current study, the raw sequencing data was obtained from the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) project (PRJEB32731), and applied to a processing pipeline (<u>https://github.com/Gibbons-Lab/pipelines</u>). Briefly, the obtained FASTQ files were processed using the fastp (version 0.23.2) tool⁶⁵ to filter and trim the reads, and taxonomic abundance was obtained using the Kraken 2 (version 2.1.2) and Bracken (version 2.6.0) tools⁶⁶ with the Kraken 2 default database (based on NCBI RefSeq). The final collapsed taxonomic table across the samples consisted of 4,669, 1,225, 354, 167, 76, and 35 taxa for species, genus, family, order, class, and phylum, respectively.

675 Blood omics-based BMI and WHtR models

676 For each Arivale baseline omic dataset, missing values were first imputed with a random forest (RF) algorithm using Python missingpy (version 0.2.0) library (corresponding to R MissForrest package⁶⁷). 677 For sex-stratified models (Supplementary Fig. 2d), the datasets after imputation were divided into sex-678 stratified datasets. Subsequently, the values in each omic dataset were standardized with Z-score using 679 680 the mean and s.d. per analyte. Then, ten iterations of least absolute shrinkage and selection operator 681 (LASSO) modeling with tenfold cross-validation (CV) (Fig. 1a, Supplementary Fig. 7a) were performed for the (unstandardized) loge-transformed BMI or WHtR and each processed omic dataset, 682 using LassoCV application programming interface (API) of Python scikit-learn (version 1.0.1) library. 683 684 Training and testing (hold-out) sets were generated by splitting participants into ten sets with one set as a testing (hold-out) set and the remaining nine sets as a training set, and iterating all combinations 685 over those ten sets; i.e., overfitting was controlled using tenfold iteration with ten testing (hold-out) 686 sets, and hyperparameter was decided using tenfold CV with internal training and validation sets from 687 688 each training set. Consequently, this procedure generated ten fitted sparse models for each omics 689 category (Supplementary Data 3) and one single testing (hold-out) set-derived prediction from each omics category for each participant. The same modeling scheme while replacing LASSO with elastic 690 net (EN), ridge, or RF was performed using Python scikit-learn *ElasticNetCV*, *RidgeCV*, or 691 RandomForestRegressor-implemented GridSearchCV API, respectively. In this RF-modeling, the 692 number of trees in the forest and the number of features were set as the hyperparameters to be decided 693 694 through CV. For the standard measures-based models, the above modeling scheme was applied to 695 ordinary least squares (OLS) linear regression with sex, age, triglycerides, HDL-cholesterol, LDL-696 cholesterol, glucose, insulin, and HOMA-IR as regressors, using Python scikit-learn LinearRegression API. Of note, ten split sets were fixed among the omics categories and the modeling methods, and no 697 significant difference in BMI, WHtR, sex, age, and ancestry PC1-5 among those ten sets was 698 confirmed, using Pearson's γ^2 test for categorical variable and Analysis of Variance (ANOVA) for 699 numeric variable while adjusting multiple testing with the Benjamini-Hochberg method across the 700 tested variables (Supplementary Data 1). 701

For the TwinsUK cohort, metabolomic dataset was applied to RF imputation and then each dataset of metabolomics and the standard clinical measures was applied to Z-score standardization, as well as the Arivale datasets. Utilizing the ten LASSO or OLS linear regression models that were fitted by the Arivale dataset, one single prediction was calculated from each processed dataset for each participant by taking the mean of ten predicted values. For metabolomics, ten metabolomics-based BMI (MetBMI) models were regenerated while restricting the input Arivale metabolomics to the common 489 metabolites in the Arivale and TwinsUK panels (Supplementary Fig. 3). 709For the LASSO-modeling iteration analysis (Supplementary Fig. 2e-h, 7f-i), ten LASSO710models were repeatedly generated with the above modeling scheme. At the end of each iteration, the711variable that was retained across ten models and that had the highest absolute value for the mean of712ten β -coefficients was removed from the input omic dataset.

713 For longitudinal predictions of the Arivale sub-cohort, one single prediction at a time point 714 was calculated from each processed time-series omic dataset for each participant, utilizing the baseline 715 LASSO model for which the participant was included in the baseline testing (hold-out) set. This was because (1) the baseline measurements were minimally affected by the personalized lifestyle 716 coaching, (2) both count and time point of data collections were different among the participants, and 717 718 (3) potential data leakage might be derived from the relationships between the baseline and following 719 measurements for the same participant. For processing, each time-series omic dataset was applied to 720 two-step RF imputation, where the baseline missingness was first imputed based on the baseline data 721 structure and the remaining missingness was next imputed based on the overall data structure, and 722 subsequently applied to Z-score standardization using the mean and s.d. in the baseline distribution.

Model performance was conservatively evaluated by the out-of-sample R^2 that was calculated from each corresponding hold-out testing set in the Arivale cohort or from the external testing set in the TwinsUK cohort. Pearson's *r* between the measured and predicted values was calculated from the overall participants of the Arivale or TwinsUK cohort. Difference of the predicted value from the measured value (Δ Measure; i.e., Δ BMI or Δ WHtR) was calculated with (the predicted value – the measured value) × (the measured value)⁻¹ × 100 (i.e., the unit of Δ Measure was [% Measure]). In the RF model, the importance of a feature was calculated as the normalized total reduction of the mean squared error that was brought by the feature.

732 Health classification

733Each participant was classified using each of the measured and omics-inferred BMIs based on the734World Health Organization (WHO) international standards for BMI cutoffs (underweight: <18.5 kg</td>735 m^{-2} , normal: 18.5–25 kg m^{-2} , overweight: 25–30 kg m^{-2} , obese: ≥ 30 kg m^{-2})¹². For the736misclassification of BMI class against the omics-inferred BMI class, each participant was categorized737into either Matched or Mismatched group when the measured BMI class was matched or mismatched738to each omics-inferred BMI class, respectively.

739 For a clinically-defined metabolic health classification, the participants having two or more 740 metabolic syndrome (MetS) risks of the National Cholesterol Education Program (NCEP) Adult Treatment Panel III (ATP III) guidelines were judged as the metabolically unhealthy group, while the 741 other participants were judged as the metabolically healthy group^{34,35}. Concretely, the MetS risk 742 components were (1) systolic blood pressure \geq 130 mm Hg, diastolic blood pressure \geq 85 mm Hg, or 743 using antihypertensive medication, (2) fasting triglyceride level \geq 150 mg dL⁻¹, (3) fasting HDL-744 cholesterol level $<50 \text{ mg dL}^{-1}$ for female and $<40 \text{ mg dL}^{-1}$ for male or using lipid-lowering 745 medication, and (4) fasting glucose level $\geq 100 \text{ mg dL}^{-1}$ or using antidiabetic medication. Only the 746 participants who had all these information were assessed in the corresponding analyses (Fig. 3b; 747 748 Supplementary Fig. 6a, 7m).

749 750

723

724

725 726

727

728

729

730 731

Gut microbiome-based models for classifying obesity

751 For the Arivale gut microbiome dataset, the whole ASV table (907 taxa from species to phylum) was preprocessed (i.e., positively shifted by one, loge-transformed, and standardized with Z-score using the 752 753 mean and s.d. per taxon) and then applied to dimensionality reduction using PCA API of Python 754 scikit-learn (version 1.0.1) library; the projected values onto the first 50 PCs (0.4–5.1% variance 755 explained) were supplied as the input gut microbiome features. Two types of classifiers were trained on these gut microbiome features: one predicting whether an individual is obese BMI class and the 756 757 other predicting whether an individual is obese MetBMI class. Both models were independently constructed through a fivefold iteration scheme of RF with fivefold CV (Fig. 4a), using Python scikit-758

759 learn RandomForestClassifier-implemented GridSearchCV API. In this RF-modeling, the number of 760 trees in the forest and the number of features were set as the hyperparameters to be decided through CV. Training and testing (hold-out) sets were generated by splitting the participants of the normal and 761 762 obese classes into five sets with one set as a testing (hold-out) set and the remaining four sets as a training set, and iterating all combinations over those five sets; i.e., overfitting was controlled using 763 764 fivefold iteration with five testing (hold-out) sets, and hyperparameters were decided using fivefold 765 CV with internal training and validation sets from each training set. Consequently, this procedure generated five fitted classifiers for each BMI or MetBMI class and one single testing (hold-out) set-766 767 derived prediction from each classifier type for each participant. Note that this prediction included two types: either normal or obese class by a vote of the trees (i.e., binary prediction) and the mean 768 769 probability of obese class among the trees.

For the TwinsUK gut microbiome dataset, the whole taxonomic table (6,526 taxa from species to phylum) was preprocessed and then applied to dimensionality reduction, as well as the Arivale dataset; the projected values onto the first 50 PCs (0.2–40.1% variance explained) were supplied as the input gut microbiome features. Then, the five obesity classifiers for each BMI or MetBMI class were generated as well as the above Arivale procedure, and one single testing (hold-out) set-derived prediction from each classifier type was calculated for each participant (Fig. 4a).

Model performance of each classifier was conservatively evaluated using each corresponding hold-out testing set. Area under curve (AUC) in the receiver operator characteristic (ROC) curve and the average precision were calculated using the probability predictions, while sensitivity and specificity were calculated from confusion matrix using the binary predictions. The overall ROC curve and its AUC was calculated from all the participant's probability predictions, using R pROC (version 1.18.0) package⁶⁸.

783

770

771 772

773

774

775

776

777

778

779 780

781

782

800

Longitudinal changes in the measured and omics-inferred BMIs

A linear mixed model (LMM) was generated for each log_e-transformed measured or omics-inferred 784 BMI in the Arivale sub-cohort, following the previous approach²⁵. As fixed effects regarding time, 785 786 linear regression splines with knots at 0, 6, 12, and 18 months were applied to days in program to fit time as a continuous variable rather than a categorical variable, because both count and time point of 787 data collections were different among the participants. In addition to the linear regression splines of 788 time as fixed effects, the LMM included sex, baseline age, ancestry PC1-5, and meteorological 789 790 seasons as fixed effects (to adjust potential confounding effects) and random intercepts and random 791 slopes of days in the program as random effects for each participant. Additionally, the same LMM for 792 each measured or omics-inferred BMI was independently generated from each baseline BMI class-793 stratified group. Of note, this stratified LMM was not generated from the underweight group because 794 its sample size was too small for convergence. For comparing difference between the misclassification 795 strata against the baseline MetBMI class, the above LMM while adding additional fixed effects, the categorical baseline misclassification of BMI class against MetBMI class (i.e., binary for Matched vs. 796 797 Mismatched) and its interaction terms with the linear regression splines of time, was generated for 798 each measured BMI or MetBMI from each baseline BMI class-stratified group. All LMMs were 799 modeled using *MixedLM* API of Python statsmodels (version 0.13.0) library.

801 Plasma analyte correlation network analysis

802Prior to the analysis, outlier values which were beyond ± 3 s.d. from the mean in the Arivale sub-803cohort baseline distribution were eliminated from the dataset per analyte, and seven clinical laboratory804tests which became almost invariant across the participants were eliminated from analyses, allowing805convergence in the following modeling. Per each analyte, values were converted with a transformation806pipeline producing the lowest skewness (e.g., no transformation, the logarithm transformation for right807skewed distribution, the square root transformation with mirroring for left skewed distribution) and808standardized with Z-score using the mean and s.d.

809 810 811 812 813 814 815 816 817 818	Against 608,856 pairwise combinations of the analytes (766 metabolites, 274 proteomics, 64 clinical laboratory tests), generalized linear models (GLMs) for the baseline measurements of the Arivale sub-cohort (Fig. 5a; 608 participants) were independently generated with the Gaussian distribution and identity link function using <i>glm</i> API of Python statsmodels (version 0.13.0) library. Each GLM consisted of an analyte as dependent variable, another analyte and the baseline MetBMI as independent variables with their interaction term, and sex, baseline age, and ancestry PC1–5 as covariates. The analyte–analyte correlation pair that was significantly modified by the baseline MetBMI was obtained based on the β -coefficient (two-sided <i>t</i> -test) of the interaction term between the independent variables in GLM, while adjusting multiple testing with the Benjamini–Hochberg method (false discovery rate (FDR) < 0.05).
819 820 821 822 823 824 825 826 827 828 828 829	Against the significant 100 pairs from the GLM analysis (82 metabolites, 33 proteins, and 16 clinical laboratory tests; Supplementary Data 7), generalized estimating equations (GEEs) for the longitudinal measurements of the metabolically obese group (i.e., the baseline obese MetBMI class; 182 participants) were independently generated with the exchangeable covariance structure using Python statsmodels <i>GEE</i> API. Each GEE consisted of an analyte as dependent variable, another analyte and days in the program as independent variables with their interaction term, and sex, baseline age, ancestry PC1–5, and meteorological seasons as covariates. The analyte–analyte correlation pair that was significantly modified by days in the program was obtained based on the β -coefficient (two-sided <i>t</i> -test) of the interaction term between the independent variables in GEE, while adjusting multiple testing with the Benjamini–Hochberg method (FDR < 0.05).
830	Statistical analysis
831 832 833 834 835 836 837 838 839 840	All data preprocessing and statistical analyses were performed using Python NumPy (version 1.18.1 or 1.21.3), pandas (version 1.0.3 or 1.3.4), SciPy (version 1.4.1 or 1.7.1) and statsmodels (version 0.11.1 or 0.13.0) libraries, except for using R pROC (version 1.18.0) package ⁶⁸ for DeLong's test ⁶⁹ . All statistical tests were performed using a two-sided hypothesis. In all cases of multiple testing, <i>P</i> -value was adjusted with the Benjamini–Hochberg method. Of note, because some hypotheses were not completely independent (e.g., between combined omics and each individual omics; between glucose, insulin, and HOMA-IR), this simple <i>P</i> -value adjustment was regarded as a conservative approach. Significance was based on $P < 0.05$ for single testing and FDR < 0.05 for multiple testing. Test summaries (e.g., sample size, degrees of freedom, test statistic, exact <i>P</i> -value) are found in Supplementary Data 4, 5, 6, 9, and 10.
841 842 843 844 845 846	Correlations (Fig. 1b, 3a; Supplementary Fig. 3b–d, 4b, 4f, 7c, 7d, 7l, 8d, 8e) were independently assessed using Pearson's correlation test (Python SciPy <i>pearsonr</i> API), with the <i>P</i> -value adjustment if multiple testing. Comparisons of model performance (Fig. 1c, 1d, 4d, 4f; Supplementary Fig. 2d, 4a, 7e) were independently assessed using Welch's <i>t</i> -test (Python statsmodels <i>ttest_ind</i> API), with the <i>P</i> -value adjustment if multiple testing. Comparison of overall ROC curves (Fig. 4c, 4e) was assessed using unpaired DeLong's test ⁶⁹ .
847 848 849 850 851 852 853 854 855 856 857 858 858 859	In all regression analyses, only the baseline datasets were used, and, unless otherwise specified, all numeric variables were centered and scaled in advance. For the Arivale datasets of anthropometrics, saliva-measured analytes, daily physical activity measures, and PRSs, (1) outlier values which were beyond ± 3 s.d. from the mean in the cohort distribution were eliminated from the dataset per variable, (2) variables which became almost invariant across the participants were eliminated from the datasets, (3) values were converted with a transformation pipeline producing the lowest skewness (e.g., no transformation, the logarithm transformation for right skewed distribution, the square root transformation with mirroring for left skewed distribution), and (4) the transformed values were standardized with <i>Z</i> -score using the mean and s.d.; these preprocessed 51 variables were used as the numeric physiological features (Supplementary Data 4). Likewise, the Arivale datasets of the obesity-related clinical blood markers (i.e., selected clinical labs; Supplementary Data 6) and the TwinsUK datasets of the obesity-related phenotypic measures (Supplementary Data 6) were preprocessed. For gut microbiome α -diversity metrics, the number of observed ASVs and Chaol

860	index were converted with square root transformation while Shannon's index was converted with
861	square transformation, and then these transformed values were standardized with Z-score using the
862	mean and s.d. Relationships of the numeric physiological features with the measured or omics-inferred
863	BMI (Fig. 1e) were independently assessed using each OLS linear regression model with the
864	(unstandardized) loge-transformed measured or omics-inferred BMI as dependent variable, a feature as
865	independent variable, and sex, age, and ancestry PC1-5 as covariates, while adjusting multiple testing
866	across the 255 (51 features × 5 BMI types) regressions. Relationships between Measure (i.e., BMI or
867	WHtR) and the analytes that were retained in at least one of ten LASSO models (Fig. 2b-d,
868	Supplementary Fig. 7k) were independently assessed using each OLS linear regression model with the
869	(unstandardized) loge-transformed Measure as dependent variable, an analyte as independent variable,
870	and sex, age, and ancestry PC1-5 as covariates, while adjusting multiple testing across the 210 (Fig.
871	2b), 75 (Fig. 2c), 42 (Fig. 2d), or 289 (Supplementary Fig. 7k) regressions. In this regression analysis,
872	a model including the omics-inferred Measure as independent variable was also assessed as reference.
873	Differences in Δ Measure (i.e., Δ BMI or Δ WHtR) between clinically-defined metabolic health
874	conditions (Fig. 3b; Supplementary Fig. 6a, 7m) were independently assessed using each OLS linear
875	regression model with Δ Measure as dependent variable, metabolic condition (i.e., Healthy vs.
876	Unhealthy) as categorical independent variable, and Measure, sex, age, and ancestry PC1-5 as
877	covariates, while adjusting multiple testing across the eight (two BMI classes × four omics categories;
878	Fig. 3b, Supplementary Fig. 7m) or four (two BMI classes × two cohorts; Supplementary Fig. 6a)
879	regressions. Differences in the obesity-related clinical blood markers, the BMI-associated numeric
880	physiological features, or the gut microbiome α -diversity metrics between the misclassification strata
881	against the omics-inferred BMI class (Fig. 3d, 3e, 4b; Supplementary Fig. 6c) were independently
882	assessed using each OLS linear regression model with a marker, feature, or metric as dependent
883	variable, misclassification (i.e., Matched vs. Mismatched) as categorical independent variable, and
884	BMI, sex, age, and ancestry PC1-5 as covariates, while adjusting multiple testing across the 40 (2
885	BMI classes \times 2 omics categories \times 10 markers; Fig. 3d), 216 (2 BMI classes \times 4 omics categories \times
886	27 features; Fig. 3e), 24 (2 BMI classes × 4 omics categories × 3 metrics; Fig. 4b), or 24 (2 BMI
887	classes \times 12 measures; Supplementary Fig. 6c) regressions. In the above regression analyses for the
888	TwinsUK cohort, ancestry PCs were eliminated from the covariates due to data availability.
889	

890 Data visualization

891 Results were visualized using Python matplotlib (version 3.4.3) and seaborn (version 0.11.2) libraries, 892 except for the plasma analyte correlation network. Data were summarized as the mean with 95% 893 confidence interval (CI) or the boxplot (median: center line; 95% CI around median: notch; $[Q_1, Q_3]$: box limits; $[x_{\min}, x_{\max}]$: whiskers, where Q_1 and Q_3 are the 1st and 3rd quartile values, and x_{\min} and x_{\max} 894 are the minimum and maximum values in $[Q_1 - 1.5 \times IQR, Q_3 + 1.5 \times IQR]$ (IQR: the interquartile 895 896 range, $O_3 - O_1$, respectively), as indicated in each figure legend. For presentation purpose, CI was 897 simultaneously calculated during visualization using Python seaborn barplot or boxplot API with 898 default setting (1,000 times bootstrapping or a Gaussian-based asymptotic approximation, respectively). The OLS linear regression line with 95% CI was simultaneously generated during 899 900 visualization using Python seaborn *regplot* API with default setting (1,000 times bootstrapping). The plasma analyte correlation network was visualized with a circos plot using R circlize (version 0.4.15) 901 package⁷⁰. 902

904 Data availability

903

905The de-identified Arivale datasets that were used in this study can be accessed by qualified researchers906for research purposes. Requests should be sent to data-access@isbscience.org, and the data will be907available after submission and approval of a research plan. The de-identified TwinsUK datasets that908were used in this study were provided by Department of Twin Research & Genetic Epidemiology909(King's College London) after the approval of our Data Access Application (Project Number: E1192).

910	Requests should be referred to their website (<u>http://twinsuk.ac.uk/resources-for-researchers/access-</u>
911	our-data/).
912	
913	Code availability
914	Code used in this study is freely available on GitHub (https://github.com/PriceLab/Multiomics-BMI).
915	

916 **References**

- NCD Risk Factor Collaboration (NCD-RisC). Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19·2 million participants. *Lancet (London, England)* 387, 1377–1396 (2016).
- 920
 921
 921
 921
 921
 922
 923
 924
 925
 925
 926
 926
 927
 927
 928
 929
 929
 929
 920
 920
 920
 921
 921
 921
 922
 923
 923
 924
 924
 925
 925
 926
 926
 926
 927
 927
 928
 929
 929
 929
 920
 920
 920
 921
 921
 921
 922
 923
 923
 924
 924
 925
 925
 926
 926
 926
 927
 927
 928
 928
 928
 929
 929
 929
 929
 929
 920
 920
 920
 921
 921
 921
 921
 922
 923
 923
 924
 924
 925
 925
 926
 926
 926
 927
 928
 928
 928
 929
 929
 929
 929
 929
 929
 920
 920
 920
 921
 921
 921
 921
 921
 921
 921
 921
 922
 923
 921
 923
 921
 923
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
 921
- 924 3. Kopelman, P. G. Obesity as a medical problem. *Nature* **404**, 635–43 (2000).
- 925 4. Haslam, D. W. & James, W. P. T. Obesity. Lancet (London, England) 366, 1197–209 (2005).
- 5. Kahn, S. E., Hull, R. L. & Utzschneider, K. M. Mechanisms linking obesity to insulin resistance and type 2 diabetes. *Nature* 444, 840–6 (2006).
- Van Gaal, L. F., Mertens, I. L. & De Block, C. E. Mechanisms linking obesity with cardiovascular disease.
 Nature 444, 875–80 (2006).
- Magkos, F. *et al.* Effects of Moderate and Subsequent Progressive Weight Loss on Metabolic Function and
 Adipose Tissue Biology in Humans with Obesity. *Cell Metab.* 23, 591–601 (2016).
- B. Hamman, R. F. *et al.* Effect of weight loss with lifestyle intervention on risk of diabetes. *Diabetes Care* 29, 2102–7 (2006).
- 934 9. Sun, Q. *et al.* Comparison of dual-energy x-ray absorptiometric and anthropometric measures of adiposity in 935 relation to adiposity-related biologic factors. *Am. J. Epidemiol.* **172**, 1442–54 (2010).
- 936 10. Prentice, A. M. & Jebb, S. A. Beyond body mass index. *Obes. Rev.* 2, 141–7 (2001).
- 937 11. Okorodudu, D. O. *et al.* Diagnostic performance of body mass index to identify obesity as defined by body
 938 adiposity: A systematic review and meta-analysis. *Int. J. Obes.* 34, 791–799 (2010).
- WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *Lancet (London, England)* 363, 157–63 (2004).
- Ruderman, N., Chisholm, D., Pi-Sunyer, X. & Schneider, S. The metabolically obese, normal-weight individual revisited. *Diabetes* 47, 699–713 (1998).
- Ding, C., Chan, Z. & Magkos, F. Lean, but not healthy: the 'metabolically obese, normal-weight'
 phenotype. *Curr. Opin. Clin. Nutr. Metab. Care* 19, 408–417 (2016).
- Smith, G. I., Mittendorfer, B. & Klein, S. Metabolically healthy obesity: facts and fantasies. J. Clin. Invest. **129**, 3978–3989 (2019).
- Appleton, S. L. *et al.* Diabetes and cardiovascular disease outcomes in the metabolically healthy obese phenotype: a cohort study. *Diabetes Care* 36, 2388–94 (2013).
- 17. Schröder, H. *et al.* Determinants of the transition from a cardiometabolic normal to abnormal overweight/obese phenotype in a Spanish population. *Eur. J. Nutr.* **53**, 1345–53 (2014).
- Williams, S. A. *et al.* Plasma protein patterns as comprehensive indicators of health. *Nat. Med.* 25, 1851–1857 (2019).
- Bar, N. *et al.* A reference map of potential determinants for the human serum metabolome. *Nature* 588, 135–140 (2020).
- Wilmanski, T. *et al.* Blood metabolome predicts gut microbiome α-diversity in humans. *Nat. Biotechnol.* 37, 1217–1228 (2019).
- 21. Cirulli, E. T. *et al.* Profound Perturbation of the Metabolome in Obesity Is Associated with Health Risk. *Cell*Metab. 29, 488-500.e2 (2019).
- 22. Talmor-Barkan, Y. *et al.* Metabolomic and microbiome profiling reveals personalized risk factors for coronary artery disease. *Nat. Med.* **28**, 295–302 (2022).
- Nimptsch, K., Konigorski, S. & Pischon, T. Diagnosis of obesity and use of obesity biomarkers in science and clinical medicine. *Metabolism.* 92, 61–70 (2019).
- Price, N. D. *et al.* A wellness study of 108 individuals using personal, dense, dynamic data clouds. *Nat. Biotechnol.* 35, 747–756 (2017).
- 25. Zubair, N. *et al.* Genetic Predisposition Impacts Clinical Changes in a Lifestyle Coaching Program. *Sci. Rep.*966 9, 6805 (2019).
- 26. Earls, J. C. *et al.* Multi-Omic Biological Age Estimation and Its Correlation With Wellness and Disease
 Phenotypes: A Longitudinal Study of 3,558 Individuals. *J. Gerontol. A. Biol. Sci. Med. Sci.* 74, S52–S60
 (2019).
- Wainberg, M. *et al.* Multiomic blood correlates of genetic risk identify presymptomatic disease alterations. *Proc. Natl. Acad. Sci. U. S. A.* 117, 21813–21820 (2020).

- Wilmanski, T. *et al.* Gut microbiome pattern reflects healthy ageing and predicts survival in humans. *Nat. Metab.* 3, 274–286 (2021).
- 29. Zimmer, A. *et al.* The geometry of clinical labs and wellness states from deeply phenotyped humans. *Nat.* 75 *Commun.* 12, 3578 (2021).
- 30. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. J. R. Stat. Soc. Ser. B 58, 267–288 (1996).
- Moayyeri, A., Hammond, C. J., Valdes, A. M. & Spector, T. D. Cohort Profile: TwinsUK and healthy ageing twin study. *Int. J. Epidemiol.* 42, 76–85 (2013).
- 32. Long, T. *et al.* Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. *Nat. Genet.* 49, 568–578 (2017).
- 33. Xu, X. *et al.* Habitual sleep duration and sleep duration variation are independently associated with body mass index. *Int. J. Obes. (Lond).* 42, 794–800 (2018).
- 34. Stefan, N., Schick, F. & Häring, H.-U. Causes, Characteristics, and Consequences of Metabolically
 984 Unhealthy Normal Weight in Humans. *Cell Metab.* 26, 292–300 (2017).
- 985 35. Blüher, M. Metabolically Healthy Obesity. *Endocr. Rev.* 41, 405–420 (2020).
- Shah, N. R. & Braverman, E. R. Measuring adiposity in patients: the utility of body mass index (BMI), percent body fat, and leptin. *PLoS One* 7, e33308 (2012).
- Tomiyama, A. J., Hunger, J. M., Nguyen-Cuu, J. & Wells, C. Misclassification of cardiometabolic health when using body mass index categories in NHANES 2005-2012. *Int. J. Obes. (Lond).* 40, 883–6 (2016).
- 99038.Bennett, C. M., Guo, M. & Dharmage, S. C. HbA(1c) as a screening tool for detection of Type 2 diabetes: a
systematic review. *Diabet. Med.* 24, 333–43 (2007).
- 992 39. Pereira-Santos, M., Costa, P. R. F., Assis, A. M. O., Santos, C. A. S. T. & Santos, D. B. Obesity and vitamin D deficiency: a systematic review and meta-analysis. *Obes. Rev.* 16, 341–9 (2015).
- 40. Ridaura, V. K. *et al.* Gut microbiota from twins discordant for obesity modulate metabolism in mice.
 Science 341, 1241214 (2013).
- 41. Turnbaugh, P. J. et al. A core gut microbiome in obese and lean twins. Nature 457, 480–484 (2009).
- 42. Le Chatelier, E. *et al.* Richness of human gut microbiome correlates with metabolic markers. *Nature* 500, 541–546 (2013).
- 43. Walters, W. A., Xu, Z. & Knight, R. Meta-analyses of human gut microbes associated with obesity and IBD. *FEBS Lett.* 588, 4223–4233 (2014).
- 100144.Duvallet, C., Gibbons, S. M., Gurry, T., Irizarry, R. A. & Alm, E. J. Meta-analysis of gut microbiome1002studies identifies disease-specific and shared responses. *Nat. Commun.* 8, 1784 (2017).
- 45. Visconti, A. *et al.* Interplay between the human gut microbiome and host metabolism. *Nat. Commun.* 10, 4505 (2019).
- 100546.Diener, C. *et al.* Baseline Gut Metagenomic Functional Gene Signature Associated with Variable Weight1006Loss Responses following a Healthy Lifestyle Intervention in Humans. *mSystems* 6, e0096421 (2021).
- 1007 47. Karetnikova, E. S. *et al.* Is Homoarginine a Protective Cardiovascular Risk Factor? *Arterioscler. Thromb.*1008 *Vasc. Biol.* 39, 869–875 (2019).
- 100948.Dieuleveux, V., Lemarinier, S. & Guéguen, M. Antimicrobial spectrum and target site of D-3-phenyllactic1010acid. Int. J. Food Microbiol. 40, 177–83 (1998).
- 101149.Beloborodova, N. *et al.* Effect of phenolic acids of microbial origin on production of reactive oxygen1012species in mitochondria and neutrophils. J. Biomed. Sci. 19, 89 (2012).
- 1013 50. Després, J.-P. & Lemieux, I. Abdominal obesity and metabolic syndrome. *Nature* 444, 881–7 (2006).
- 101451.Ashwell, M., Gunn, P. & Gibson, S. Waist-to-height ratio is a better screening tool than waist circumference1015and BMI for adult cardiometabolic risk factors: systematic review and meta-analysis. Obes. Rev. 13, 275–861016(2012).
- 1017 52. Swainson, M. G., Batterham, A. M., Tsakirides, C., Rutherford, Z. H. & Hind, K. Prediction of whole-body
 1018 fat percentage and visceral adipose tissue mass from five anthropometric variables. *PLoS One* 12, e0177175
 1019 (2017).
- 1020 53. Li, Y. *et al.* Adrenomedullin is a novel adipokine: adrenomedullin in adipocytes and adipose tissues.
 1021 *Peptides* 28, 1129–43 (2007).
- 102254.Egaña-Gorroño, L. *et al.* Receptor for Advanced Glycation End Products (RAGE) and Mechanisms and1023Therapeutic Opportunities in Diabetes and Cardiovascular Disease: Insights From Human Subjects and1024Animal Models. Front. Cardiovasc. Med. 7, 37 (2020).
- 102555.Norata, G. D. *et al.* Circulating soluble receptor for advanced glycation end products is inversely associated1026with body mass index and waist/hip ratio in the general population. Nutr. Metab. Cardiovasc. Dis. 19, 129–102734 (2009).

- 102856.Rauschert, S., Uhl, O., Koletzko, B. & Hellmuth, C. Metabolomic biomarkers for obesity in humans: A1029short review. Ann. Nutr. Metab. 64, 314–324 (2014).
- 1030 57. Rangel-Huerta, O. D., Pastor-Villaescusa, B. & Gil, A. Are we close to defining a metabolomic signature of 1031 human obesity? A systematic review of metabolomics studies. *Metabolomics* 15, 93 (2019).
- Barber, M. N. *et al.* Plasma lysophosphatidylcholine levels are reduced in obesity and type 2 diabetes. *PLoS One* 7, e41456 (2012).
- 1034 59. Piening, B. D. *et al.* Integrative Personal Omics Profiles during Periods of Weight Gain and Loss. *Cell Syst.*1035 6, 157-170.e8 (2018).
- Koenig, R. J. *et al.* Correlation of glucose regulation and hemoglobin AIc in diabetes mellitus. *N. Engl. J. Med.* 295, 417–20 (1976).
- 1038 61. Wing, R. R. & Phelan, S. Long-term weight loss maintenance. Am. J. Clin. Nutr. 82, 2228-225S (2005).
- 1039 62. Li, G. *et al.* The long-term effect of lifestyle interventions to prevent diabetes in the China Da Qing Diabetes
 1040 Prevention Study: a 20-year follow-up study. *Lancet (London, England)* 371, 1783–9 (2008).
- 104163.Diabetes Prevention Program Research Group *et al.* 10-year follow-up of diabetes incidence and weight loss1042in the Diabetes Prevention Program Outcomes Study. Lancet (London, England) 374, 1677–86 (2009).
- 1043 64. Yilmaz, P. *et al.* The SILVA and 'All-species Living Tree Project (LTP)' taxonomic frameworks. *Nucleic* 1044 *Acids Res.* 42, D643-8 (2014).
- 1045 65. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890 (2018).
- 1047 66. Lu, J. *et al.* Metagenome analysis using the Kraken software suite. *Nat. Protoc.* 1–25 (2022).
 1048 doi:10.1038/s41596-022-00738-y
- 1049 67. Stekhoven, D. J. & Bühlmann, P. Missforest-Non-parametric missing value imputation for mixed-type data.
 1050 *Bioinformatics* 28, 112–118 (2012).
- 1051 68. Robin, X. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC* 1052 *Bioinformatics* 12, 77 (2011).
- 1053 69. DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. Comparing the areas under two or more correlated 1054 receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44, 837–45 (1988).
- 1055 70. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. Circlize implements and enhances circular visualization
 1056 in R. *Bioinformatics* 30, 2811–2812 (2014).

1058 Acknowledgements

1059 We thank Sergey A. Kornilov, Gustavo Glusman, and Max Robinson (Institute for Systems Biology; ISB) for providing comments to this study. We thank Victoria Vazquez and Andrew Anastasiou 1060 1061 (King's College London) for their support in obtaining and utilizing the TwinsUK data access. We are grateful to all Arivale and TwinsUK participants who consented to using their deidentified data for 1062 1063 research purposes. This work was supported by the M.J. Murdock Charitable Trust (Reference No. 2014096:MNL:11/20/2014, awarded to N.D.P. and L.H.), the National Institutes of Health (NIH) 1064 grants awarded by the National Institute on Aging (NIA) (U19AG023122 and 5U01AG061359), and a 1065 1066 generous gift from K. Carole Ellison (to K.W., T.W., and A.Z.). K.W. was supported by The Uehara Memorial Foundation (Overseas Postdoctoral Fellowships). C.D. and S.M.G. were supported by the 1067 Washington Research Foundation Distinguished Investigator Award and startup funds from ISB. 1068 1069 TwinsUK is funded by the Wellcome Trust, Medical Research Council, Versus Arthritis, European 1070 Union Horizon 2020, Chronic Disease Research Foundation (CDRF), Zoe Ltd, the National Institute for Health and Care Research (NIHR) Clinical Research Network (CRN) and Biomedical Research 1071 1072 Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College 1073 London. 1074

1075 Author Contribution

K.W., T.W., L.H., N.D.P., and N.R. conceptualized the study. K.W., T.W., A.Z., N.D.P., and N.R.
participated in the study design. K.W., T.W., C.D., B.L., and N.R. performed data analysis and figure
generation. C.D., J.C.E., J.J.H., J.C.L., S.M.G., A.T.M., and L.H. assisted in results interpretation.
J.C.L. and A.T.M. managed the logistics of data collection and integration. K.W., T.W., and N.R.
were the primary authors of the paper, with contributions from all other authors. All authors read and
approved the final manuscript.

1083 Competing Interests

1084J.J.H. has received grants from Pfizer and Novartis for research unrelated to this study. All other1085authors declare no competing interests.



1087

1089

Figure 1. Plasma multiomics captured 48–78% of the variance in BMI.

1090a Overview of study cohorts and the omics-based Body Mass Index (BMI) model generation. LASSO:1091least absolute shrinkage and selection operator, CV: cross-validation. b Correlation between the1092measured and predicted BMIs. The solid line is the ordinary least squares (OLS) linear regression line1093with 95% confidence interval (CI), and the dotted line is measured BMI = predicted BMI. Standard1094measures: OLS linear regression model with sex, age, triglycerides, high-density lipoprotein (HDL)-1095cholesterol, low-density lipoprotein (LDL)-cholesterol, glucose, insulin, and homeostatic model

1096	assessment for insulin resistance (HOMA-IR) as regressors; P: adjusted P-value of two-sided
1097	Pearson's correlation test with the Benjamini–Hochberg method across the five categories. $n = 1,277$
1098	participants. c, d Model performance of each fitted BMI model. Out-of-sample R^2 was calculated from
1099	each corresponding hold-out testing set (c, Arivale in d) or from the external testing set (TwinsUK in
1100	d). Metabolomics (full): LASSO model trained by all 766 metabolites of the Arivale dataset,
1101	Metabolomics (restricted): LASSO model trained by the common 489 metabolites in the Arivale and
1102	TwinsUK datasets (see Supplementary Fig. 3). Note that Standard measures and Metabolomics (full)
1103	of Arivale in d are the same with corresponding ones in c . Data: mean with 95% CI, $n = 10$ models.
1104	***Adjusted $P < 0.001$ in two-sided Welch's <i>t</i> -test with the Benjamini–Hochberg method across the
1105	four (c) or three (d) comparisons. e Association between omics-inferred BMI and physiological
1106	feature. For each of the 51 numeric physiological features (Supplementary Data 4), β -coefficient was
1107	estimated using OLS linear regression model with the measured or omics-inferred BMI as dependent
1108	variable and sex, age, and ancestry principal components (PCs) as covariates. Presented are the 30
1109	features that were significantly associated with at least one of the BMI types after multiple testing
1110	adjustment with the Benjamini-Hochberg method across the 255 (51 features × 5 BMI types)
1111	regressions. BMI: measured BMI, MetBMI: metabolomics-inferred BMI, ProtBMI: proteomics-
1112	inferred BMI, ChemBMI: clinical chemistries-inferred BMI, CombiBMI: combined omics-inferred
1113	BMI, PRS: polygenic risk score, n: the number of assessed participants. Data: estimate with 95% CI.
1114	*Adjusted $P < 0.05$, **adjusted $P < 0.01$, ***adjusted $P < 0.001$.
1115	



1117Figure 2. Omics-based BMI estimates captured the variance in BMI better than any single1118analyte.

1119**a** The variables that were retained across all ten combined omics-based Body Mass Index1120(CombiBMI) models (132 analytes: 77 metabolites, 51 proteins, and 4 clinical laboratory tests). β -1121coefficient was obtained from the fitted CombiBMI model with least absolute shrinkage and selection1122operator (LASSO) regression. Each background color corresponds to the analyte category. Data:

1123	median (center line), $[Q_1, Q_3]$ (box limits), $[x_{\min}, x_{\max}]$ (whiskers), where Q_1 and Q_3 are the 1st and 3rd
1124	quartile values, and x_{min} and x_{max} are the minimum and maximum values in $[Q_1 - 1.5 \times IQR, Q_3 + 1.5]$
1125	× IQR] (IQR: the interquartile range, $Q_3 - Q_1$), respectively; $n = 10$ models. b - d Univariate explained
1126	variance in BMI by each metabolite (b), protein (c), or clinical laboratory test (d). BMI was
1127	independently regressed on each of the analytes that were retained in at least one of the ten LASSO
1128	models (209 metabolites, 74 proteins, 41 clinical laboratory tests; Supplementary Data 5), using
1129	ordinary least squares (OLS) linear regression with sex, age, and ancestry principal components (PCs)
1130	as covariates. Multiple testing was adjusted with the Benjamini-Hochberg method across the 210 (b),
1131	75 (c), or 42 (d) regressions, including each omics-based BMI (MetBMI: metabolomics-based BMI,
1132	ProtBMI: proteomics-based BMI, ChemBMI: clinical chemistries-based BMI) model as reference.
1133	Among the analytes that were significantly associated with BMI (180 metabolites, 63 proteins, 30
1134	clinical laboratory tests), only the top 30 significant analytes are presented with their univariate
1135	variances.
1136	



1137 1138

1139

Figure 3. Metabolic heterogeneity was responsible for the high rate of misclassification within the standard BMI classes.

1140	a Difference of the omics-inferred Body Mass Index (BMI) from the measured BMI (Δ BMI).
1141	MetBMI: metabolomics-inferred BMI, ProtBMI: proteomics-inferred BMI, ChemBMI: clinical
1142	chemistries-inferred BMI, CombiBMI: combined omics-inferred BMI, P: adjusted P-value of two-
1143	sided Pearson's correlation test with the Benjamini–Hochberg method across the six combinations, n:
1144	the number of participants in each BMI class (total $n = 1,277$ participants). The line in histogram panel
1145	indicates the kernel density estimate. b Difference in Δ BMI between clinically-defined metabolic
1146	health conditions within the normal or obese BMI class. Significance was assessed using ordinary
1147	least squares (OLS) linear regression with BMI, sex, age, and ancestry principal components (PCs) as
1148	covariates, while adjusting multiple testing with the Benjamini-Hochberg method across the eight
1149	(two BMI classes × four omics categories) regressions. c Misclassification rate of overall cohort or
1150	each BMI class against the omics-inferred BMI class. Range of the previously reported
1151	misclassification rate ^{36,37} is highlighted with orange background. Note that the underweight BMI class
1152	is not presented due to small sample size, but its misclassification rate was 80% against CombiBMI
1153	class and 100% against the others. d, e Difference in the obesity-related clinical blood marker (d) or
1154	BMI-associated physiological feature (e) between Matched and Mismatched groups within the normal
1155	or obese BMI class. Significance was assessed using OLS linear regression with BMI, sex, age, and
1156	ancestry PCs as covariates, while adjusting multiple testing with the Benjamini-Hochberg method
1157	across the 40 (d, 2 BMI classes \times 2 omics categories \times 10 markers) or 216 (e, 2 BMI classes \times 4 omics
1158	categories \times 27 features) regressions. Four of the 27 features that were significantly associated with
1159	BMI (Fig. 1c) are representatively presented in e, and the other results are found in Supplementary
1160	Data 6. HDL: high-density lipoprotein, LDL: low-density lipoprotein, CRP: C-reactive protein,
1161	HOMA-IR: homeostatic model assessment for insulin resistance, HbA1c: glycated hemoglobin A1c,
1162	25(OH)D: 25-hydroxyvitamin D. b, d, e Data: median (center line), 95% confidence interval (CI)
1163	around median (notch), $[Q_1, Q_3]$ (box limits), $[x_{\min}, x_{\max}]$ (whiskers), where Q_1 and Q_3 are the 1st and
1164	3rd quartile values, and x_{min} and x_{max} are the minimum and maximum values in $[Q_1 - 1.5 \times IQR, Q_3 +$
1165	1.5 × IQR] (IQR: the interquartile range, $Q_3 - Q_1$), respectively; $n = 373$ (b, Healthy in Normal), 49
1166	(b, Unhealthy in Normal), 208 (b, Healthy in Obese), 241 (b, Unhealthy in Obese) participants (see
1167	Supplementary Data 6 for each sample size in d and e). *Adjusted $P < 0.05$, **adjusted $P < 0.01$,
1168	***adjusted $P < 0.001$.



a Overview of study cohorts and the gut microbiome-based obesity classifier generation. BMI: Body Mass Index, MetBMI: metabolomics-inferred BMI, RF: random forest, CV: cross-validation. **b** Difference in gut microbiome α -diversity between Matched and Mismatched groups within the normal or obese BMI class. Significance was assessed using ordinary least squares (OLS) linear regression with BMI, sex, age, and ancestry principal components (PCs) as covariates, while adjusting multiple testing with the Benjamini–Hochberg method across the 24 (2 BMI classes × 4 omics categories × 3 metrics) regressions. ProtBMI: proteomics-inferred BMI, ChemBMI: clinical chemistries-inferred

1179	BMI, CombiBMI: combined omics-inferred BMI, ASV: amplicon sequence variant. Data: median
1180	(center line), 95% confidence interval (CI) around median (notch), $[Q_1, Q_3]$ (box limits), $[x_{\min}, x_{\max}]$
1181	(whiskers), where Q_1 and Q_3 are the 1st and 3rd quartile values, and x_{min} and x_{max} are the minimum
1182	and maximum values in $[Q_1 - 1.5 \times IQR, Q_3 + 1.5 \times IQR]$ (IQR: the interquartile range, $Q_3 - Q_1$),
1183	respectively. $n = 240$ (Normal), 260 (Obese) participants (see Supplementary Data 6 for each sample
1184	size). *Adjusted $P < 0.05$, **adjusted $P < 0.01$. c, e Receiver operator characteristic (ROC) curve of
1185	the gut microbiome-based model classifying participants to the normal vs. obese class in the Arivale
1186	(c) or TwinsUK (e) cohort. Each ROC curve was generated from the overall participants: $n = 500$ (c,
1187	BMI class), 427 (c, MetBMI class), 209 (e, BMI class), 145 (e, MetBMI class) participants. The red
1188	dashed line indicates a random classification line. AUC: area under curve. ** $P < 0.01$ in two-sided
1189	unpaired DeLong's test. d, f Comparison of model performance between the BMI and MetBMI
1190	classifiers in the Arivale (d) or TwinsUK (f) cohort. Out-of-sample metric value was calculated from
1191	each corresponding hold-out testing set. Data: mean with 95% CI, $n = 5$ models. * $P < 0.05$, ** $P <$
1192	0.01 in two-sided Welch's <i>t</i> -test.
1193	

Page 32 of 36



1195Figure 5. Metabolic health of the metabolically obese group was substantially improved1196following a healthy lifestyle intervention.

1197a Overview of the longitudinal analysis using omics-inferred Body Mass Index (BMI). BMI:1198measured BMI, MetBMI: metabolomics-inferred BMI, ProtBMI: proteomics-inferred BMI,1199ChemBMI: clinical chemistries-inferred BMI, LMM: linear mixed model. **b**, **c** Longitudinal change in1200the omics-inferred BMI within the overall cohort (**b**) or within each baseline BMI class (**c**). Average1201trajectory of each measured or omics-inferred BMI was independently estimated using LMM with1202random effects for each participant (see Methods) in the overall cohort (**b**) or in each baseline BMI1203class-stratified group (**c**). **d**, **e** Longitudinal change in MetBMI of the misclassified participants within

1204	the normal (d) or obese (e) BMI class. Average trajectory of each BMI or MetBMI was independently
1205	estimated using the above LMM with the baseline misclassification of BMI class against MetBMI
1206	class as additional fixed effects (see Methods) in each baseline BMI class-stratified group. b-e The
1207	dashed line corresponds to the baseline value of each estimate. Data: mean with 95% confidence
1208	interval (CI); <i>n</i> = 608 (b), 222 (c , Normal), 185 (c , Overweight), 196 (c , Obese), 137 (d , Matched), 85
1209	(d, Mismatched), 139 (e, Matched), 57 (e, Mismatched) participants.
1210	



1212Figure 6. Plasma analyte correlation network in the metabolically obese group shifted toward a1213structure observed in metabolically healthier state following a healthy lifestyle intervention.

1214a Cross-omic interactions modified by metabolomics-inferred Body Mass Index (MetBMI) and days1215in the program. For each of the 608,856 pairwise relationships of plasma analytes (766 metabolites,

1216	274 proteomics, 64 clinical laboratory tests), the baseline relationship between analyte-analyte pair
1217	and MetBMI within the Arivale sub-cohort (Fig. 5a; 608 participants) was assessed using their
1218	interaction term in each generalized linear model (GLM; see Methods), while adjusting multiple
1219	testing with the Benjamini-Hochberg method. The 100 analyte-analyte pairs (82 metabolites, 33
1220	proteins, 16 clinical laboratory tests; Supplementary Data 7) that were significantly modified by the
1221	baseline MetBMI are presented. For each of these 100 pairs, the longitudinal relationship between
1222	analyte-analyte pair and days in the program within the metabolically obese group (i.e., the baseline
1223	obese MetBMI class; 182 participants) was further assessed using their interaction term in each
1224	generalized estimating equation (GEE; see Methods), while adjusting multiple testing with the
1225	Benjamini-Hochberg method. The 27 analyte-analyte pairs (21 metabolites, 3 proteins, 3 clinical
1226	laboratory tests) that were significantly modified by days in the program are highlighted by line width
1227	and label font size. b , c Representative examples of the analyte-analyte pair that was significantly
1228	modified by both baseline MetBMI (b) and days in the program (c) in a. The solid line in each panel is
1229	the ordinary least squares (OLS) linear regression line with 95% confidence interval (CI). $n = 530$ (b,
1230	left), 553 (b , center), 566 (b , right) participants; $n = 324$ (c , left), 339 (c , center), 347 (c , right)
1231	measurements from the 182 participants of the metabolically obese group. Of note, data points outside
1232	of plot range are trimmed in these presentations.
1233	