

1 **Amino acid variability, tradeoffs and optimality in human diet**

2
3 Ziwei Dai^{1,2} and Jason W. Locasale^{1*}

4 1 Department of Pharmacology and Cancer Biology, Duke University School of Medicine,
5 Durham, NC 27710, USA

6 2 Department of Biology, School of Life Science, Southern University of Science and
7 Technology, Shenzhen 518055, China

8
9 *Corresponding author: Jason W. Locasale. Email: dr.jason.locasale@gmail.com

10
11 **Abstract**

12 While the quality of fat (e.g. saturated/unsaturated) and carbohydrate (e.g. whole grain/simple
13 sugars) intake has been of great interest, less attention has been made to the type of protein and
14 resulting amino acid intake profiles in human diets. Studies at the molecular level however
15 demonstrate that dietary amino acid intake produces substantial effects on health and disease such
16 as cancer by modulating metabolism. How these effects may manifest in human food consumption
17 and dietary patterns is unknown. We developed a series of algorithms to map, characterize and
18 model the landscape of amino acid content in human food, dietary patterns, and individual
19 consumption including relations to health status, covering over 2,000 foods, ten dietary patterns,
20 and over 30,000 dietary records. We found that the type of amino acids contained in foods and
21 human consumption is highly dynamic with variability far exceeding that of fat and carbohydrate.
22 Some amino acids positively associate with diseases such as obesity while others contained in the
23 same food negatively link to disease. Using linear programming and machine learning, we show
24 that these health trade-offs among can be accounted to satisfy biochemical constraints in food and
25 human eating patterns to construct a Pareto front in dietary practice, a means of achieving
26 optimality in the face of tradeoffs that are commonly considered in economic and evolutionary
27 theories. Thus this study may enable the design of human protein quality intake guidelines based
28 on a quantitative framework.

33 **Introduction**

34 Diet is generally considered to be a major determinant of human health and disease¹⁻⁵. Numerous
35 dietary recommendations, such as the Dietary Guidelines for Americans⁶, have been developed.
36 These dietary recommendations often focus on two major goals: to increase the diversity and
37 nutrient density of the foods consumed, and to reduce the intake of certain components known to
38 increase risk of disease⁷⁻⁹. Such restrictions involve limiting the intake of certain types of
39 carbohydrate and fat such as added sugar, saturated fat and trans-fat, and has rationale based on
40 epidemiology, human¹⁰⁻¹² and model organism research^{13,14}. While it has been widely
41 acknowledged that the types of dietary carbohydrate and fat are important determinants of the
42 quality of a diet, protein the other macronutrient¹⁵, is often neglected. In most human nutritional
43 studies albeit with exceptions, protein is considered as a single variable and often held constant¹⁶.
44 Nevertheless, each amino acid has its specific metabolism¹⁷ and is important for numerous cellular
45 and physiological processes. A growing number of studies shows that variation in dietary intake
46 of amino acids such as serine, glycine, asparagine, histidine, and methionine mediates health and
47 disease including cancer through defined molecular mechanisms¹⁸⁻²⁸. Altogether there is a
48 rationale for investigating in a systematic manner amino acid intake in human diets and possible
49 consequences on health.

50 In this study, we investigated the variability of amino acids in human food and diets and find
51 variability commensurate with what is observed in fats and carbohydrates. Based on optimizing
52 associations with health status, we use these analyses to devise guidelines for dietary amino acids.
53 Finally, we implement machine learning algorithms to design personalized diets based on amino
54 acid intake that correspond to optimality in specified health statuses.

55

56 **Amino acid landscape of human food**

57 To characterize the variability of amino acid levels in human food, we first constructed a database
58 consisting of amino acid profiles in three levels of human dietary components: individual foods,
59 dietary patterns or representations of patterns of food consumption (e.g. Western, Mediterranean,

60 Japanese, Keto, etc), and dietary records containing daily reported food intake (Figure 1). The
61 abundance of 18 amino acids in 2,335 foods was collected based on nutritional profiles in the
62 United States of America Department of Agriculture National Nutrient Database for Standard
63 Reference Legacy Release (USDA SR) (Figure 1a, methods). 18 of the 20 amino acids were
64 considered because during quantitation, amino acids which largely exist in protein-bound forms,
65 require hydrolysis into free amino acids during which amino groups from glutamine and
66 asparagine are also hydrolyzed to make glutamic and aspartic acid. Thus, the abundance of
67 glutamic acid and aspartic acid from measurements of free amino acid levels reflects the total
68 abundance of glutamate and glutamine, and the total abundance of aspartate and asparagine,
69 respectively. The distributions of amino acid abundance over 2,335 foods show that each amino
70 acid has considerable variability across foods (Coefficient of variation > 0.2 for all amino acids,
71 Figure 1b), and amino acids most abundant in human food are glutamine/glutamate (median = 0.16
72 g/g total amino acids), asparagine/aspartate (median = 0.095 g/g total amino acids), leucine
73 (median = 0.082 g/g total amino acids), and lysine (median = 0.076 g/g total amino acids). On the
74 other hand, amino acids with the lowest abundance in human foods are cystine (median = 0.012
75 g/g total amino acids), tryptophan (median = 0.012 g/g total amino acids), methionine (median =
76 0.024 g/g total amino acids), and histidine (median = 0.028 g/g total amino acids). This ordering
77 largely resembles the abundance of amino acids in the proteomes which are conserved across living
78 organisms^{29,30}. Principal component analysis (PCA) shows that amino acid abundances can be
79 clustered by different categories of foods (Figure 1c, d, methods). Highly variable amino acids
80 include those whose dietary modulation has molecular links to cancer progression and health
81 outcomes, such as methionine (0.031 g/g total amino acids in eggs compared to 0.013 in legumes)
82 and serine (0.076 g/g total amino acids in eggs compared to 0.039 in lamb, veal, and game meat).
83 To quantify the variability of amino acid abundance across foods, we computed the F-statistic from
84 one-way analysis of variance (ANOVA), and compared the resulting F-statistic values with those
85 of carbohydrates (i.e. dietary fiber and sugar) and fats (i.e. saturated fat, monounsaturated fat, and
86 polyunsaturated fat). Notably, we found that the ANOVA F-statistics for amino acids were

87 comparable to or higher than those for carbohydrates and fats (Figure 1e, methods), especially for
88 the amino acids methionine, histidine, lysine, and proline (F-statistic = 816.2 for methionine, 566.1
89 for histidine, 504.3 for lysine, and 362.9 for proline compared to the range of 45.0 to 119.6 for
90 carbohydrates and the range of 125.2 to 746.3 for fats, Figure 1e, f), highlighting the variability of
91 amino acid abundance in foods which has been largely overlooked previously. Taken together,
92 these results suggest that differences in food intake due to the high variability in amino acid content
93 may lead to differences physiological and cellular effects on metabolism.

94

95 **Human dietary patterns are variable in amino acid content**

96 Dietary patterns can be grouped according to eating patterns that often have a cultural or societal
97 element. They can be characterized by a combination of certain types of foods consumed (e.g.
98 Mediterranean diet, which includes high amounts of plant-based foods, high to moderate amounts
99 of seafood, low consumption of red meat, and olive oil as the main source of added fat³¹), or a
100 specific intake profile of certain nutrients (e.g. ketogenic diet, which is defined by very high intake
101 of fat and very low intake of carbohydrate). Adherence to certain dietary patterns, such as the
102 Mediterranean diet or Japanese diet, has been associated with increased lifespan and lower risk of
103 disease³²⁻³⁴. Moreover, some emerging dietary patterns, such as the ketogenic diet and the Paleo
104 diet, have recently been shown in some settings to have benefits on metabolic health, neural
105 function, and longevity³⁵⁻³⁸. However, it is unclear whether these dietary patterns differ in their
106 amino acid content, and whether the variability in amino acid abundance across dietary patterns
107 contributes to the health outcomes associated with these diets.

108 To further understand the relationship between human dietary patterns and amino acid intake,
109 we next developed an algorithm to quantitatively evaluate amino acid abundance in ten
110 representative human dietary patterns (Figure 2a, S1, Supplementary Methods). Among these
111 dietary patterns, the Mediterranean diet and Japanese diet are two traditional diets believed to have
112 beneficial influences on health, while the Dietary Approaches to Stop Hypertension (DASH) diet
113 consists of consumption of a variety of low-fat and minimally processed foods, and the American

114 diet, which represents the dietary behaviors of a typical individual in western society is also
115 considered. We also include diets that restrict the consumption of certain foods (Paleo diet,
116 vegetarian diet, plant-based diet), diets limiting carbohydrate intake (ketogenic diet, Atkins diet),
117 and a USDA recommended diet defined based on the daily nutrient intake goals in the USDA 2015-
118 2020 dietary guidelines for Americans⁶. We first computed the range of amino acid intake (i.e.
119 grams of each amino acid consumed per day) for each dietary pattern using a linear programming
120 algorithm we developed (Figure 2b, Supplementary Methods) and found that, although none of
121 these dietary patterns includes any constraint on amino acid intake, they still differ greatly with
122 each other in the values of amino acid consumption. Moreover, each dietary pattern allowed for
123 substantial flexibility in the intake of all amino acids (maximal daily intake/minimal daily intake >
124 20 for all dietary patterns and amino acids, Figure 2b), revealing the possibility to modulate amino
125 acid intake under a certain dietary pattern.

126 To quantify the variability of amino acid composition that is independent of energy and protein
127 intake, we developed a sampling algorithm based on the accelerated convergence hit-and-run
128 method³⁹ to quantify the amino acid composition of each diet by sampling 50,000 instances of
129 each diet (Supplementary Methods). We first confirmed that the sample size of 50,000 was
130 sufficient to capture the distribution of amino acid abundance in a dietary pattern based on the
131 convergence of the sample mean and standard deviation values (Figure S2a). PCA of the sampled
132 diets (Figure 2c) and comparison of mean values (Figure 2d) showed that the ten dietary patterns
133 also have different signatures of amino acid composition. Notably, differences in amino acid
134 composition also exists between dietary patterns similar to each other such as the vegetarian diet
135 and plant-based diet. Indeed, we observed a 30% of difference in methionine abundance between
136 vegetarian diet and plant-based diet (0.019 g methionine/g total AAs in vegetarian diet compared
137 to 0.014 in plant-based diet), suggesting that small changes in the choice of foods result in
138 substantial differences in amino acid intake (Figure 2d, Figure S2b). We also estimated
139 compositions of carbohydrates and fats in these diets (Figure S2b), and quantified the variability
140 of amino acid composition across human diets using F-statistic values from one-way ANOVA, and

141 compared it with the variability of carbohydrates and fats across dietary patterns (Figure 2e).
142 Strikingly, we found that the variability of amino acid composition across diets was much higher
143 than that of carbohydrates and fats, with the amino acids lysine, methionine, proline and histidine
144 being the most highly variable across human dietary patterns (F-statistic > 50,000 compared to
145 less than 10,000 for carbohydrates and fats, Figure 2e-f, S2b-c). Among these amino acids, lysine,
146 histidine and methionine are significantly lower in instances of the plant-based diet, and proline is
147 significantly lower in Paleo diet (Figure 2f). The amino acid signatures of human dietary patterns
148 were further validated by measurements of fasting blood concentrations of the amino acids leucine,
149 isoleucine, and alanine in human subjects eating plant-based or ketogenic diet (Figure S2d)⁴⁰.
150 Taken together, these results reveal that the biggest difference in macronutrient composition across
151 human dietary patterns is in amino acid content, and not that of carbohydrates or fats. How the
152 diversity in dietary amino acids results in different health outcomes remains an open question,
153 which may begin to be answered with nutritional and health data in large populations of humans.

154

155 **Landscape of amino acid intake in human dietary records**

156 Next, we considered individual dietary amino acid intake records across a population of
157 individuals from diverse ethnic and cultural backgrounds. We reconstructed the dietary amino acid
158 intake profiles in more than 30,000 human subjects in the United States based on dietary records
159 in the National Health and Nutrition Examination Survey (NHANES) 2007-2014 datasets (Figure
160 3a). Since the NHANES datasets do not direct include dietary amino acid intake values, we
161 developed a set of computational tools for data imputation and mapping to reconstruct the amino
162 acid profiles for the dietary records based on two additional datasets, the USDA SR food nutritional
163 database and the Food and Nutrient Database for Dietary Studies (FNDDS) (Figure 3a). Data
164 imputation using random forest (RF) regression, which outperformed other methods in the
165 accuracy of imputation (Figure S3a, b), was applied to estimate the missing values of amino acid
166 levels in the USDA SR dataset. The imputed datasets were then used to construct amino acid
167 profiles for the FNDDS and NHANES records by mapping foods in the USDA dataset to foods in

168 the FNDDS dataset which were then used to compute nutrient intake values in the NHANES
169 dietary records (Figure 3a, Supplementary Methods). To assess the limitations of self-reported
170 dietary records in the NHANES data, we compared our computed nutrient intake values with
171 measurements of blood concentrations of related metabolites such as Vitamin D (Figure 3b). Next,
172 to validate the reconstructed amino acid intake levels, we first compared the total intake of amino
173 acids and intake of protein in each dietary record and confirmed that the reconstructed total amino
174 acid intake closely resembles the known total protein intake (Pearson correlation = 0.99, p-value
175 $< 10^{-323}$, Figure 3c), concentrations of amino acids in human blood (Spearman correlation = 0.52,
176 p-value = 0.03, Figure 3d), uptake fluxes of amino acids in human cell lines, which reflect demands
177 of amino acids in cultured human cells (Spearman correlation = 0.70, p-value = 0.01, Figure 3e),
178 and amino acid composition of several culture mediums (Spearman correlation > 0.5 and p-value
179 < 0.05 for 4 out of 7 culture media, Figure S3c). The high correlation between dietary amino acid
180 intake and physiological parameters related to amino acids suggests that our reconstructed amino
181 acid intake data may reflect some aspects of physiological metabolism, and suggest that the cellular
182 behaviors and tissue microenvironment in amino acid metabolism reflect to some extent dietary
183 intake of amino acids despite the many other factors that influence cellular metabolism.

184 We then evaluated the overall variability in the intake of each amino acid based on the ratio
185 of maximal to minimal intake values in the human dietary records (Figure 3f), and performed PCA
186 on the reconstructed dietary amino acid profiles to report the association between dietary amino
187 acid composition and demographic variables such as age, sex, and ethnicity (Figure 3g). We found
188 that among the population included in the NHANES 2007-2014 cohorts, daily intake of amino
189 acids typically varies by two to six fold (e.g. maximal intake/minimal intake = 4 for tryptophan,
190 2.5 for methionine, 6.2 for glycine, and so on). Dietary amino acid composition profiles showed
191 no difference between batches (Figure 3e), thus confirming that our reconstruction is not biased
192 by batch effect. Interestingly, dietary intake of amino acids was found to correlate with age, while
193 no dependency on other demographic variables such as sex and ethnicity was observed (Fig 3e,
194 Figure S4). These reconstructed dietary amino acid intake profiles allow us to examine the

195 quantitative relationship between dietary amino acids and human health.

196

197 **Dietary amino acid intake associations with human health**

198 We next attempted to link dietary amino acid intake and incidence of several human diseases based
199 on the reconstructed dietary amino acid intake profiles and clinical records available in the
200 NHANES database. We focused on chronic diseases that are a major concern to human health such
201 as cardiovascular disease, diabetes, and cancer. We retrieved the medical records of 18,196 adult
202 subjects in the NHANES 2007-2014 datasets and defined quantitative scores describing the
203 incidences of hypertension, obesity, cancer, and diabetes based on the examination, laboratory, and
204 questionnaire datasets (Figure 4a, Methods). We first computed partial Spearman's rank
205 correlation coefficients as a metric to evaluate the association between dietary amino acid
206 composition and the incidences of the four diseases while controlling for confounders including
207 demographic and lifestyle-related factors (Supplementary Figure 5). We identified many amino
208 acid intake-disease associations involving all four diseases considered (statistically significant
209 associations in 21 out of 72 amino acid-disease pairs, Figure 4b, methods), among which obesity
210 showed the strongest association with dietary amino acid composition (obesity incidence
211 positively correlated with the intake of threonine, histidine, alanine, glycine, lysine and methionine,
212 and negatively correlated with intake of tryptophan, phenylalanine, valine, serine, asparagine,
213 aspartate, glutamine, and glutamate, Figure 4b). These associations between dietary amino acid
214 intake and obesity were consistent with some observations in molecular studies, such as the anti-
215 obesity functions of dietary tryptophan and pro-obesity functions of methionine in mice^{41,42}. As a
216 control, we also correlated the incidence of the four diseases with dietary intake of different types
217 of carbohydrates and fats. Counterintuitively, we found much fewer statistically significant
218 associations between dietary intake of carbohydrate and fat (9 significant associations out of 40
219 disease-nutrient pairs, Figure 4c). These results together highlight the unexpectedly strong
220 association between that dietary intake of amino acids and human disease which exceeds the
221 association for dietary carbohydrates and fats. To further explore these questions, we performed a

222 comparison of the association between nutrients and human health using machine learning models
223 predicting health outcomes from different types of nutritional variables (Figure 4d). We
224 categorized nutritional variables included in the NHANES database into six groups, including
225 energy, macronutrients, macronutrient compositions (i.e. fractions of different types of
226 carbohydrate and fat in total carbohydrate and fat intake), vitamins, minerals, amino acid
227 compositions (i.e. intake of each amino acid with the unit g/g total AA), and other nutrients. For
228 each disease, nutritional variables in each group were used as covariates to build a logistic
229 regression model to predict the incidence of that disease. The area under receiver operating
230 characteristic curve (AUC) with 5-fold cross validation was used to assess the performance of each
231 group of nutritional variables in predicting disease incidence, which reflects strength of the
232 association between dietary intake of those nutrients and that disease. We found that dietary amino
233 acid composition was predictive of incidence of all diseases except for cancer ($AUC > 0.5$, 5-fold
234 cross validation), and achieved accuracy of prediction comparable to or higher than that of dietary
235 carbohydrate and fat intake for obesity and hypertension ($AUC = 0.55$ for amino acids compared
236 to 0.55 for macronutrient composition in predicting obesity, and $AUC = 0.53$ for amino acids
237 compared to 0.52 for macronutrient composition in predicting hypertension, Figure 4e). The reason
238 that dietary amino acid intake was unable to predict cancer outcome was probably for the reason
239 that different types of cancers were not distinguished in the analysis, the population included
240 remissions, and the frequency of cancer in the dataset is relatively low (1844 cases out of 18469
241 individuals). Nevertheless, the higher accuracy of amino acid intake in predicting obesity and
242 hypertension incidence in humans provides a rationale for optimization of dietary amino acid
243 intake.

244

245 **Guidelines for dietary amino acids and diet design**

246 Dietary recommendations, such as these in the USDA Dietary Guidelines for Americans, often
247 involve suggestions to consume a variety of minimally processed foods and recommended ranges
248 for intake of nutrients including macronutrients, vitamins, and minerals. Since dietary intake of

249 amino acids has been associated with health outcomes both in molecular studies and by our
250 analysis thus far, we sought to develop an Artificial Intelligence (AI) -based approach for
251 identification of dietary guidelines for amino acids and design of personalized human diets
252 optimizing their amino acid composition.

253 First, we developed an algorithm for identification of amino acid intake guidelines based on
254 the associations between dietary amino acid intake and human health (Figure 5a). We focused on
255 obesity since it had the highest incidence and was found to have the strongest association with
256 dietary amino acid intake among the four diseases considered in this study (Figure 4b). We
257 classified obesity-associated amino acids into three categories (Figure 5b), including amino acids
258 for which the intake positively associate with obesity incidence ('positive association'), negatively
259 associate with obesity incidence ('negative association'), or associate with obesity incidence with
260 a non-monotonic, U-shaped relationship ('U-shaped relationship'). The amino acids phenylalanine,
261 aspartate/asparagine, tryptophan, valine and glutamate/glutamine fell into the negative association
262 group. On the other hand, the amino acids glycine, alanine, methionine, lysine, histidine were
263 categorized into the positive association group. The association between intake of dietary amino
264 acids and obesity was not due to changes in calorie intake, since amino acids positively associated
265 with obesity were either negatively or positively correlated to calorie intake, and *vice versa* (Figure
266 S6a).

267 We also examined whether there exists a dietary pattern that can minimize the intake of the
268 amino acids positively associated with obesity while maximizing the intake of the amino acids
269 negatively associated with obesity. To our surprise, no dietary pattern was able to satisfy all of
270 these requirements. For instance, the Paleo diet has the highest levels of aspartate and asparagine,
271 which negatively associate with obesity. Nevertheless, the Paleo diet also has the highest intake of
272 alanine, methionine and lysine, which all positively associate with obesity incidence. These results
273 reveal the complexity in the relationship between dietary amino acid intake and obesity, indicating
274 trade-offs between the goals of maximizing or minimizing different groups of amino acids which
275 should be considered while designing dietary guidelines for amino acids.

276 We therefore sought to define dietary amino acid intake guidelines based on the association
277 between dietary amino acids and obesity (Figure 5c), that is, to minimize the total intake of amino
278 acids that positively associate with obesity (i.e. AAs-to-minimize, including glycine, alanine,
279 methionine, lysine, histidine), and to maximize the total intake of amino acids that negatively
280 associate with obesity (i.e. AAs-to-maximize, including tryptophan, phenylalanine, valine,
281 aspartate+asparagine, glutamate+glutamine). We first confirmed that both total AAs-to-minimize
282 and total AAs-to-maximize were significantly associated with obesity incidence (Chi-squared p-
283 value = 9.0×10^{-8} for total AAs-to-minimize and 4.9×10^{-10} for total AAs-to-maximize, Figure 5c).

284 We then further characterized the trade-off between the requirements of minimizing total AAs-
285 to-minimize and maximizing total AAs-to-maximize by constructing the Pareto surface based on
286 the two requirements (Figure 5d). The concept of Pareto optimality has been widely applied in
287 economics and engineering, and introduced to biology to characterize the trade-off between
288 multiple tasks of bacteria, cancer cells, and organisms⁴³⁻⁴⁶. For each dietary pattern, there exists a
289 Pareto surface consisting of diets that best balance the needs to minimize total AAs-to-minimize
290 and to maximize total AAs-to-maximize, meaning that for a diet within the Pareto surface, any
291 other diet following this dietary pattern would never have both higher total intake of AAs-to-
292 maximize and lower total intake of AAs-to-minimize at the same time. We hence developed an
293 algorithm to construct the Pareto surface for each of the ten dietary patterns considered in this
294 study (Figure 5d, S6b, Methods), and quantified the extent by which a specific diet satisfies the
295 two requirements of maximizing total AAs-to-maximize and minimizing total AAs-to-minimize
296 using the deviation from Pareto surface (Figure 5d). For each dietary pattern, we computed the
297 deviation of each NHANES dietary record from its Pareto surface, and found that the deviation
298 from the Pareto surface strongly correlates with obesity incidence (Chi-squared p-values $< 10^{-10}$
299 for all dietary patterns), implying that diets on the Pareto surface of each dietary pattern are
300 associated with lower risk of obesity. On average, an individual that eats a diet that is the top 20%
301 furthest away from the Pareto surface has a 34% higher chance of being obese compared to one
302 eating a diet among the 20% closest to the Pareto surface (Figure 5e).

303 These findings not only reveal novel relationship between dietary amino acid intake and health,
304 but also allow us to design diets that have amino acid profiles associated with lower risk of obesity
305 and satisfy personalized needs and requirements such as preferred dietary patterns according to the
306 constructed Pareto surface of the preferred dietary pattern. Hence, based on such strategy, we
307 developed an AI for designing diets including the Mediterranean, Paleo, and ketogenic diet (Figure
308 5f). Each diet contains a variety of foods from diverse sources and keeps the features of the
309 corresponding dietary pattern.

310

311 **Discussion**

312 This study develops data resources and computational techniques to begin to address two major
313 limitations in the nutritional sciences: 1) the lack of systematic collections of nutritional
314 information and 2) the lack of computational tools to probe the connections in food, dietary
315 patterns and practices, and health status. Consequentially, we made a number of findings about
316 the variability of amino acids across different types of human foods and dietary patterns and the
317 unexpected associations between dietary amino acid intake, food and dietary patterns, and health.
318 Unexpected links from amino acid intake to pathology such as obesity highlight non-intuitive diet-
319 disease associations and inherent tradeoffs in amino acid content in food.

320 While we were able to use the tools we devised to study and make discoveries about the
321 landscape of amino acid intake, these capabilities are generalizable to any systematic analysis of
322 human food and diet. For instance, it is still unclear how dietary patterns and human dietary
323 records differ with each other in micronutrients such as vitamins, minerals, dietary fiber, added
324 sugars, and how personalized diets can be designed to cover more nutritional goals. Application of
325 the algorithms we developed in this study may help address these questions.

326 This study has some limitations. First, the association between dietary amino acids and human
327 diseases is observational and does not directly imply causality. Nevertheless, some amino acid-
328 disease associations identified by our analysis have been observed in experimental studies. For
329 instance, tryptophan, which was found to be negatively associated with obesity in our study, was

330 shown in mice to reduce appetite and weight gain through the production of serotonin in brain⁴¹.
331 On the other hand, dietary restriction of methionine in mice and human has been shown to improve
332 metabolic health and increase fat oxidation, which may contribute to the anti-obesity effects of
333 dietary methionine restriction^{42,47,48}. Further studies, such as randomized controlled trials that
334 directly compare the health outcomes of diets differing with each other in amino acids, are
335 necessary but also limited to the cohort in consideration and the pre-determined end points.

336 We also note that the datasets used in this study are not completely free of bias. The majority
337 of records in the databases of foods and human dietary records are western, while foods frequently
338 consumed in other geographical regions and by other cultural groups, such as Asians and Africans,
339 are largely underrepresented. Therefore, application of our findings to non-western populations
340 may be limited. Nevertheless, we are optimistic that this limitation could be addressed by
341 extending the coverage of the existing nutritional and epidemiological datasets to non-western
342 populations^{49,50}.

343

344 **Methods**

345 **Computer algorithms and their implementation**

346 Details about the computer algorithms used in this study, including these for reconstruction of
347 amino acid landscape in human foods, dietary patterns, and dietary records, are explained in the
348 Supplementary Methods. The algorithms for imputation and reconstruction of amino acid profiles
349 in the NHANES database, including imputation of missing data, and mapping of foods in the
350 USDA SR, FNDDS, and NHANES databases, were implemented in R. All other algorithms used
351 in this study were implemented in MATLAB. The database for amino acid abundance in human
352 foods, dietary patterns and dietary records was implemented in both Microsoft Access database
353 file and Microsoft Excel files. All database files are freely available for download at the GitHub
354 repository: https://github.com/ziweidai/AA_human_diet/tree/main/6-Database.

355 **Statistical analysis**

356 Principal component analysis was performed using the MATLAB built-in function ‘pca()’. One-

357 way ANOVA was performed using the MATLAB built-in function ‘anova1()’. Logistic regression
358 models were constructed, trained, and evaluated using the MATLAB built-in functions ‘glmfit()’,
359 ‘glmval()’, and ‘perfcurve()’. Chi-squared test was performed using the MATLAB built-in
360 function ‘crosstab()’. Relationships with p-value < 0.05 were considered significant. Partial
361 Spearman’s rank correlation coefficients were computed using the MATLAB built-in function
362 ‘partialcorr()’ with p-values adjusted using the Benjamini-Hochberg procedure. Associations with
363 adjusted p-value < 0.05 were considered significant. Average amino acid abundances in food
364 categories or dietary patterns were computed using the mean values of amino acid abundances
365 across all foods in that food category or instances in that dietary pattern.

366 **Data and code availability**

367 All code, scripts, and datasets used or generated in this study are available at the GitHub page of
368 Ziwei Dai: https://github.com/ziweidai/AA_human_diet.

369 **Acknowledgments**

370 The authors thank all members of the Locasale Lab, especially Dr. Zhengtao Xiao, Shiyu Liu, and
371 Yudong Sun, for helpful discussions. JWL thanks the Marc Lustgarten Foundation, the National
372 Institutes of Health (R01CA193256 to JWL), and the American Cancer Society 129832-RSG-16-
373 214-01-TBE for their generous support. Support for computational resources from the Duke
374 Compute Cluster and Data Commons Storage is gratefully acknowledged.

375 **Author contributions**

376 Z.D. and J.W.L. designed the study, wrote and edited the paper. Z.D. developed the algorithms and
377 analyzed the data.

378 **Competing interests**

379 J.W.L. advises Restoration Foodworks, Nanocare Technologies and Raphael Pharmaceuticals. Z.D.
380 declares no competing interests.

381

382 **References**

383 1 Bose, S., Allen, A. E. & Locasale, J. W. The Molecular Link from Diet to Cancer Cell
384 Metabolism. *Mol Cell* **80**, 554, doi:10.1016/j.molcel.2020.10.006 (2020).

385 2 Steck, S. E. & Murphy, E. A. Dietary patterns and cancer risk. *Nat Rev Cancer* **20**, 125-
386 138, doi:10.1038/s41568-019-0227-4 (2020).

387 3 Kanarek, N., Petrova, B. & Sabatini, D. M. Dietary modifications for enhanced cancer
388 therapy. *Nature* **579**, 507-517, doi:10.1038/s41586-020-2124-0 (2020).

389 4 Lien, E. C. & Vander Heiden, M. G. A framework for examining how diet impacts tumour
390 metabolism. *Nat Rev Cancer* **19**, 651-661, doi:10.1038/s41568-019-0198-5 (2019).

391 5 Afshin, A. *et al.* Health effects of dietary risks in 195 countries, 1990-2017: a systematic
392 analysis for the Global Burden of Disease Study 2017. *Lancet* **393**, 1958-1972,
393 doi:10.1016/S0140-6736(19)30041-8 (2019).

394 6 United States. Department of Health and Human Services., United States. Department of
395 Agriculture. & United States. Dietary Guidelines Advisory Committee. *Dietary guidelines
396 for Americans, 2015-2020*. Eighth edition. edn, (U.S. Department of Health and Human
397 Services and U.S. Department of Agriculture, 2015).

398 7 Katz, D. L. & Meller, S. Can We Say What Diet Is Best for Health? *Annu Rev Publ Health*
399 **35**, 83-103, doi:10.1146/annurev-publhealth-032013-182351 (2014).

400 8 Willett, W. C. & Stampfer, M. J. Current evidence on healthy eating. *Annu Rev Public
401 Health* **34**, 77-95, doi:10.1146/annurev-publhealth-031811-124646 (2013).

402 9 Willett, W. *et al.* Food in the Anthropocene: the EAT-Lancet Commission on healthy diets
403 from sustainable food systems. *Lancet* **393**, 447-492, doi:10.1016/S0140-6736(18)31788-
404 4 (2019).

405 10 Seidelmann, S. B. *et al.* Dietary carbohydrate intake and mortality: a prospective cohort
406 study and meta-analysis. *Lancet Public Health* **3**, e419-e428, doi:10.1016/S2468-
407 2667(18)30135-X (2018).

408 11 Jakobsen, M. U. *et al.* Major types of dietary fat and risk of coronary heart disease: a pooled
409 analysis of 11 cohort studies. *Am J Clin Nutr* **89**, 1425-1432, doi:10.3945/ajcn.2008.27124
410 (2009).

411 12 Yang, Q. *et al.* Added sugar intake and cardiovascular diseases mortality among US adults.
412 *JAMA Intern Med* **174**, 516-524, doi:10.1001/jamainternmed.2013.13563 (2014).

413 13 Goncalves, M. D. *et al.* High-fructose corn syrup enhances intestinal tumor growth in mice.
414 *Science* **363**, 1345-1349, doi:10.1126/science.aat8515 (2019).

415 14 Hopkins, B. D. *et al.* Suppression of insulin feedback enhances the efficacy of PI3K
416 inhibitors. *Nature* **560**, 499-503, doi:10.1038/s41586-018-0343-4 (2018).

417 15 Fulgoni, V. L., 3rd. Current protein intake in America: analysis of the National Health and
418 Nutrition Examination Survey, 2003-2004. *Am J Clin Nutr* **87**, 1554S-1557S,
419 doi:10.1093/ajcn/87.5.1554S (2008).

420 16 Newgard, C. B. *et al.* A branched-chain amino acid-related metabolic signature that
421 differentiates obese and lean humans and contributes to insulin resistance. *Cell Metab* **9**,
422 311-326, doi:10.1016/j.cmet.2009.02.002 (2009).

423 17 Nelson, D. L., Cox, M. M. & Lehninger, A. L. *Lehninger principles of biochemistry*.
424 Seventh edition. edn, (W.H. Freeman and Company ;
425 Macmillan Higher Education, 2017).

426 18 Grandison, R. C., Piper, M. D. & Partridge, L. Amino-acid imbalance explains extension
427 of lifespan by dietary restriction in *Drosophila*. *Nature* **462**, 1061-1064,
428 doi:10.1038/nature08619 (2009).

429 19 Piper, M. D. W. *et al.* Matching Dietary Amino Acid Balance to the In Silico-Translated
430 Exome Optimizes Growth and Reproduction without Cost to Lifespan. *Cell Metab* **25**,
431 1206, doi:10.1016/j.cmet.2017.04.020 (2017).

432 20 Barcena, C. *et al.* Methionine Restriction Extends Lifespan in Progeroid Mice and Alters
433 Lipid and Bile Acid Metabolism. *Cell Rep* **24**, 2392-2403,
434 doi:10.1016/j.celrep.2018.07.089 (2018).

435 21 Maddocks, O. D. K. *et al.* Modulating the therapeutic response of tumours to dietary serine
436 and glycine starvation. *Nature* **544**, 372-376, doi:10.1038/nature22056 (2017).

437 22 Gao, X. *et al.* Dietary methionine influences therapy in mouse cancer models and alters
438 human metabolism. *Nature* **572**, 397-401, doi:10.1038/s41586-019-1437-3 (2019).

439 23 Kanarek, N. *et al.* Histidine catabolism is a major determinant of methotrexate sensitivity.
440 *Nature* **559**, 632-636, doi:10.1038/s41586-018-0316-7 (2018).

441 24 Knott, S. R. V. *et al.* Asparagine bioavailability governs metastasis in a model of breast
442 cancer. *Nature* **554**, 378-381, doi:10.1038/nature25465 (2018).

443 25 Tajan, M. *et al.* Serine synthesis pathway inhibition cooperates with dietary serine and
444 glycine limitation for cancer therapy. *Nat Commun* **12**, 366, doi:10.1038/s41467-020-
445 20223-y (2021).

446 26 Muthusamy, T. *et al.* Serine restriction alters sphingolipid diversity to constrain tumour
447 growth. *Nature* **586**, 790-795, doi:10.1038/s41586-020-2609-x (2020).

448 27 Sanderson, S. M., Gao, X., Dai, Z. & Locasale, J. W. Methionine metabolism in health and
449 cancer: a nexus of diet and precision medicine. *Nat Rev Cancer* **19**, 625-637,
450 doi:10.1038/s41568-019-0187-8 (2019).

451 28 Mendez-Lucas, A. *et al.* Identifying strategies to target the metabolic flexibility of tumours.
452 *Nat Metab* **2**, 335-350, doi:10.1038/s42255-020-0195-8 (2020).

453 29 Krick, T. *et al.* Amino Acid metabolism conflicts with protein diversity. *Mol Biol Evol* **31**,
454 2905-2912, doi:10.1093/molbev/msu228 (2014).

455 30 Zhang, H. *et al.* Biosynthetic energy cost for amino acids decreases in cancer evolution.
456 *Nat Commun* **9**, 4124, doi:10.1038/s41467-018-06461-1 (2018).

457 31 Bach-Faig, A. *et al.* Mediterranean diet pyramid today. Science and cultural updates. *Public*
458 *Health Nutr* **14**, 2274-2284, doi:10.1017/S1368980011002515 (2011).

459 32 Estruch, R. *et al.* Primary prevention of cardiovascular disease with a Mediterranean diet.
460 *N Engl J Med* **368**, 1279-1290, doi:10.1056/NEJMoa1200303 (2013).

461 33 Asano, M. *et al.* Effects of Japanese diet in combination with exercise on visceral fat
462 accumulation. *Nutrition* **57**, 173-182, doi:10.1016/j.nut.2018.05.023 (2019).

463 34 Tsugane, S. & Sawada, N. The JPHC Study: Design and Some Findings on the Typical
464 Japanese Diet. *Jpn J Clin Oncol* **44**, 777-782, doi:10.1093/jjco/hyu096 (2014).

465 35 Roberts, M. N. *et al.* A Ketogenic Diet Extends Longevity and Healthspan in Adult Mice.
466 *Cell Metab* **26**, 539-546 e535, doi:10.1016/j.cmet.2017.08.005 (2017).

467 36 Newman, J. C. *et al.* Ketogenic Diet Reduces Midlife Mortality and Improves Memory in
468 Aging Mice. *Cell Metab* **26**, 547-557 e548, doi:10.1016/j.cmet.2017.08.004 (2017).

469 37 Olson, C. A. *et al.* The Gut Microbiota Mediates the Anti-Seizure Effects of the Ketogenic
470 Diet. *Cell* **174**, 497, doi:10.1016/j.cell.2018.06.051 (2018).

471 38 Masharani, U. *et al.* Metabolic and physiologic effects from consuming a hunter-gatherer
472 (Paleolithic)-type diet in type 2 diabetes. *Eur J Clin Nutr* **69**, 944-948,
473 doi:10.1038/ejcn.2015.39 (2015).

474 39 Kaufman, D. E. & Smith, R. L. Direction Choice for Accelerated Convergence in Hit-and-
475 Run Sampling. *Oper Res* **46**, 84-95 (1998).

476 40 Hall, K. D. *et al.* Effect of a plant-based, low-fat diet versus an animal-based, ketogenic
477 diet on ad libitum energy intake. *Nat Med* **27**, 344-353, doi:10.1038/s41591-020-01209-1
478 (2021).

479 41 Solon-Biet, S. M. *et al.* Branched chain amino acids impact health and lifespan indirectly
480 via amino acid balance and appetite control. *Nat Metab* **1**, 532-545, doi:10.1038/s42255-
481 019-0059-2 (2019).

482 42 Ables, G. P., Perrone, C. E., Orentreich, D. & Orentreich, N. Methionine-restricted
483 C57BL/6J mice are resistant to diet-induced obesity and insulin resistance but have low
484 bone density. *PLoS One* **7**, e51357, doi:10.1371/journal.pone.0051357 (2012).

485 43 Schuetz, R., Zamboni, N., Zampieri, M., Heinemann, M. & Sauer, U. Multidimensional
486 optimality of microbial metabolism. *Science* **336**, 601-604, doi:10.1126/science.1216882
487 (2012).

488 44 Shoval, O. *et al.* Evolutionary trade-offs, Pareto optimality, and the geometry of phenotype
489 space. *Science* **336**, 1157-1160, doi:10.1126/science.1217405 (2012).

490 45 Dai, Z. W. *et al.* Identification of Cancer-associated metabolic vulnerabilities by modeling
491 multi-objective optimality in metabolism. *Cell Commun Signal* **17**, doi:ARTN 124
492 10.1186/s12964-019-0439-y (2019).

493 46 Hausser, J. & Alon, U. Tumour heterogeneity and the evolutionary trade-offs of cancer. *Nat*
494 *Rev Cancer* **20**, 247-257, doi:10.1038/s41568-020-0241-6 (2020).

495 47 Plaisance, E. P. *et al.* Dietary methionine restriction increases fat oxidation in obese adults
496 with metabolic syndrome. *J Clin Endocrinol Metab* **96**, E836-840, doi:10.1210/jc.2010-
497 2493 (2011).

498 48 Wang, L. *et al.* Methionine restriction alleviates high-fat diet-induced obesity: Involvement
499 of diurnal metabolism of lipids and bile acids. *Biochim Biophys Acta Mol Basis Dis* **1866**,
500 165908, doi:10.1016/j.bbadis.2020.165908 (2020).

501 49 Hankin, J. H. *et al.* Singapore Chinese Health Study: development, validation, and
502 calibration of the quantitative food frequency questionnaire. *Nutr Cancer* **39**, 187-195,
503 doi:10.1207/S15327914nc392_5 (2001).

504 50 Nishioka, S., Wakabayashi, H. & Momosaki, R. Nutritional Status Changes and Activities
505 of Daily Living after Hip Fracture in Convalescent Rehabilitation Units: A Retrospective
506 Observational Cohort Study from the Japan Rehabilitation Nutrition Database. *J Acad Nutr*
507 *Diet* **118**, 1270-1276, doi:10.1016/j.jand.2018.02.012 (2018).

508

509 **Figure captions**

510 **Figure 1. Amino acid landscape of human foods**

- 511 a. Workflow for construction of the database for amino acid abundances in human foods.
- 512 b. Ranges of amino acid abundance in human foods. The horizontal lines indicate median
- 513 values.
- 514 c. Principal Components Analysis (PCA) of amino acid profiles in human foods. Each dot
- 515 represents a food. Colors of the dots indicate different categories of the foods.
- 516 d. Average amino acid abundance in different categories of human foods.
- 517 e. F-statistic values from one-way ANOVA comparing abundance of amino acids, different
- 518 types of carbohydrate, and different types of fat across human foods.
- 519 f. Violin plots showing the distributions of abundance of amino acids, carbohydrates, and fats
- 520 that are the most variable across human foods. The circles indicate median values. Green
- 521 dots indicate individual values.

522

523 **Figure 2. Amino acid landscape of human diets**

- 524 a. Workflow for the computational modeling of amino acid abundance in human dietary
- 525 patterns.
- 526 b. Absolute levels of amino acids in human dietary patterns quantified by the minimal and
- 527 maximal daily intake values of amino acids in each dietary pattern.
- 528 c. PCA of relative amino acid compositions of human diets sampled for all ten dietary patterns.
- 529 Each dot represents for a diet. Colors of the dots indicate different dietary patterns.
- 530 d. Average amino acid composition of the ten human dietary patterns.
- 531 e. F-statistic values from one-way ANOVA comparing the composition of amino acids,
- 532 carbohydrates, and fats across human dietary patterns.
- 533 f. Violin plots showing the distributions of amino acids that are the most variable across
- 534 human dietary patterns. The circles indicate median values.

535

536 **Figure 3. Landscape of human dietary amino acid intake**

- 537 a. Workflow for reconstruction of the database consisting of amino acid intake profiles in
538 human dietary records.
- 539 b. Comparison between nutrient intake values in the self-reported dietary records and
540 laboratory measurements of nutrient-related metabolites in blood.
- 541 c. Comparison between total dietary amino acid intake in the reconstructed amino acid intake
542 database and dietary protein intake in the original dietary records.
- 543 d. Comparison of the reconstructed human dietary amino acid intake values to blood
544 concentrations of amino acids. The dots represent for mean values and error bars for
545 standard deviations.
- 546 e. Comparison of the reconstructed human dietary amino acid intake values to uptake fluxes
547 of amino acids. The dots represent for mean values and error bars for standard deviations.
- 548 f. Distributions of amino acid intake in human dietary records. The circles indicate median
549 values.
- 550 g. PCA analysis of amino acid intake values in human dietary records showing their
551 association with age, sex, ethnicity, and batch of the data.

552

553 **Figure 4. Amino acid intake is predictive of human health**

- 554 a. Workflow for the analysis of association between dietary amino acid intake and human
555 health.
- 556 b. Partial Spearman correlation between incidences of human diseases and dietary intake of
557 amino acids.
- 558 c. Partial Spearman correlation between incidences of human diseases and dietary intake of
559 different types of carbohydrate and fat.
- 560 d. Framework of the machine learning model predicting incidence of human diseases from
561 different groups of dietary variables.

562 e. AUC values for predicting incidence of human diseases from different groups of dietary
563 variables. Error bars indicate standard deviations.

564

565 **Figure 5. AI for dietary amino acid guidelines and personalized diet design**

566 a. Workflow for AI-assisted identification of dietary amino acid guidelines and design of
567 personalized diets.

568 b. Three types of association between dietary amino acid intake and obesity in humans.

569 c. Identification and confirmation of amino acid intake guidelines based on the association
570 between dietary amino acids and obesity.

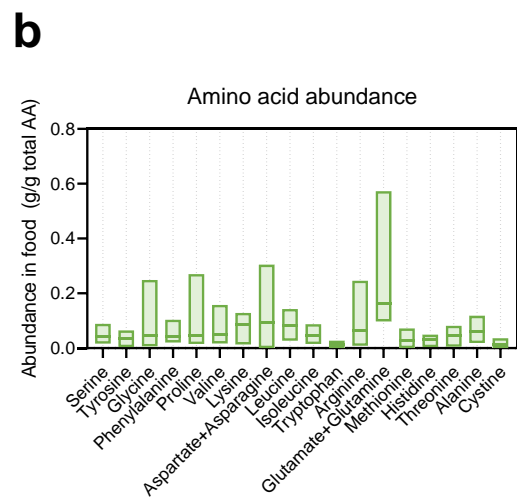
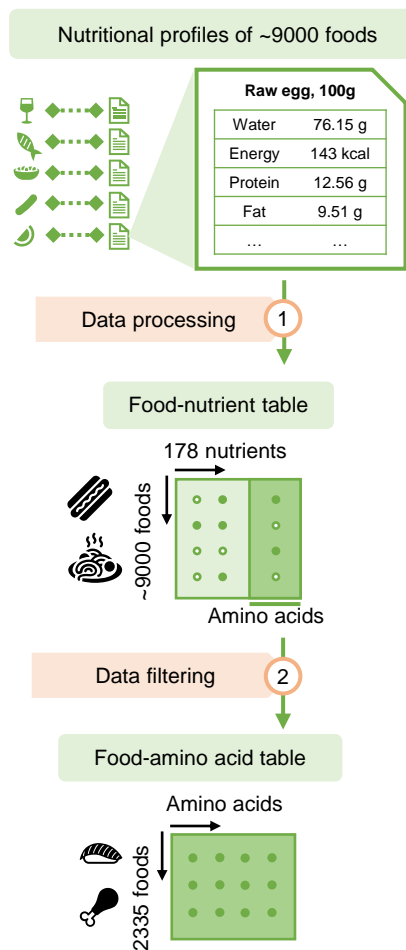
571 d. Ranges of intake of total amino-acids-to-maximize and amino-acids-to-minimize in the
572 dietary pattern of USDA-recommended diet (grey shaded region) and the Pareto surface
573 (orange bold curve) corresponding to the two guidelines, i.e. maximizing total amino-
574 acids-to-maximize, and minimizing total amino-acids-to-minimize.

575 e. Associations between the obesity incidence and deviation of dietary records from the
576 Pareto surface. Chi-squared p-values were computed to assess the significance levels of the
577 associations.

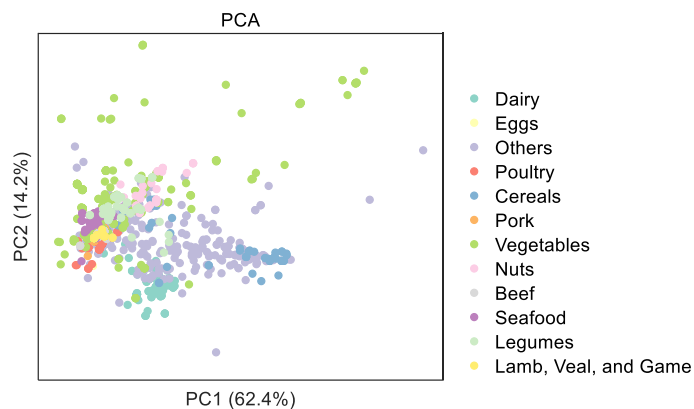
578 f. Examples of diets designed according to the amino acid intake guidelines and personalized
579 preferences of dietary patterns.

Figure 1. Amino acid landscape of human foods

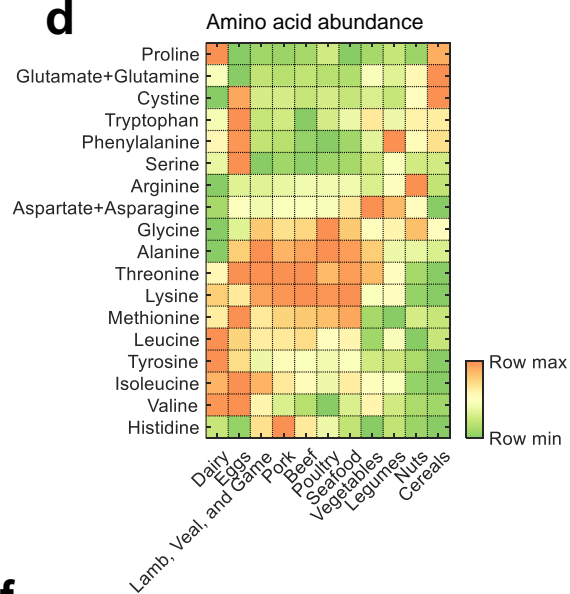
a Construction of database for amino acid abundance in foods



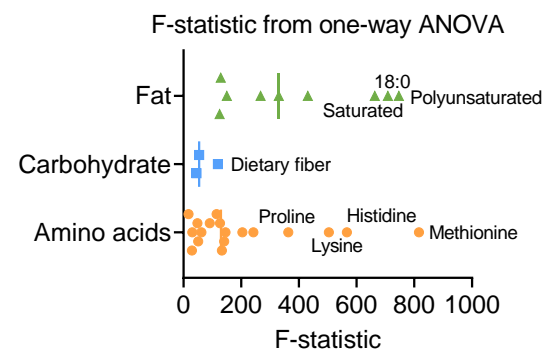
c



d



e



f

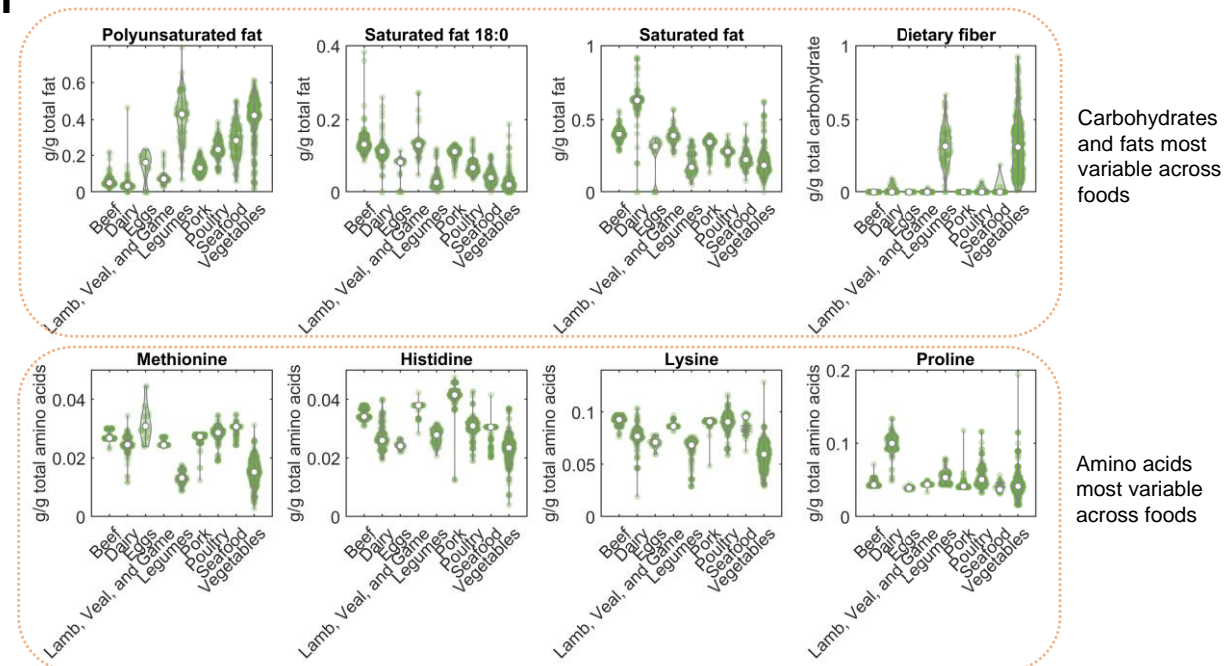
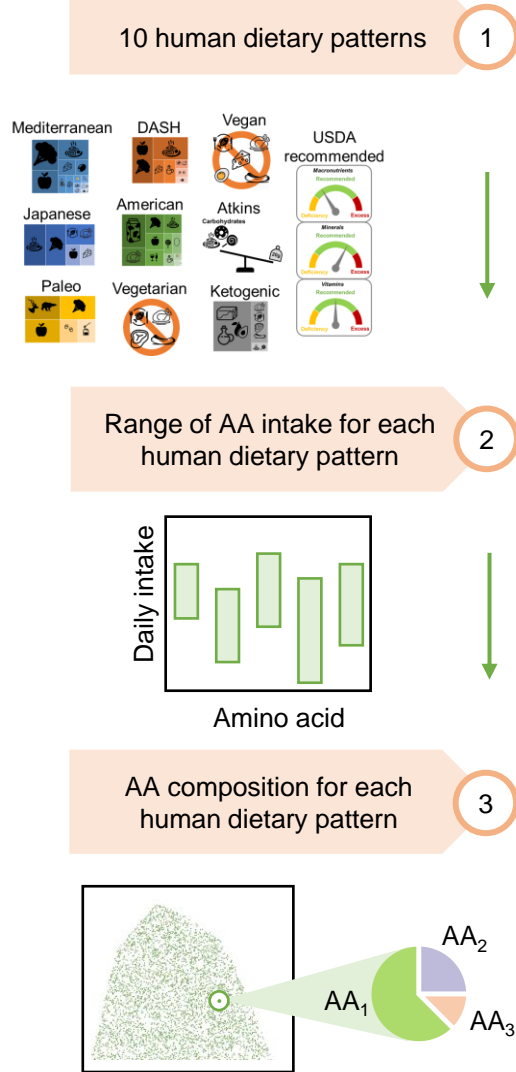


Figure 2. Amino acid landscape of human dietary patterns

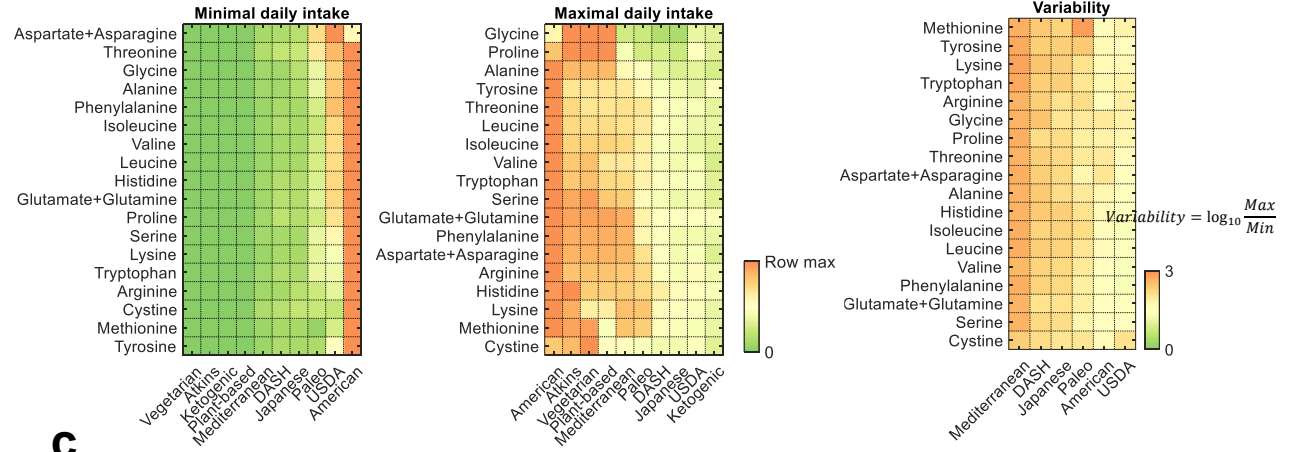
a

Amino acid landscape analysis



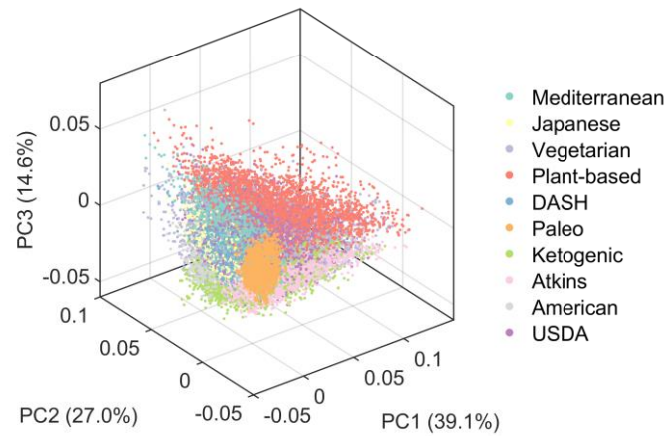
b

Absolute amino acid levels in human dietary patterns



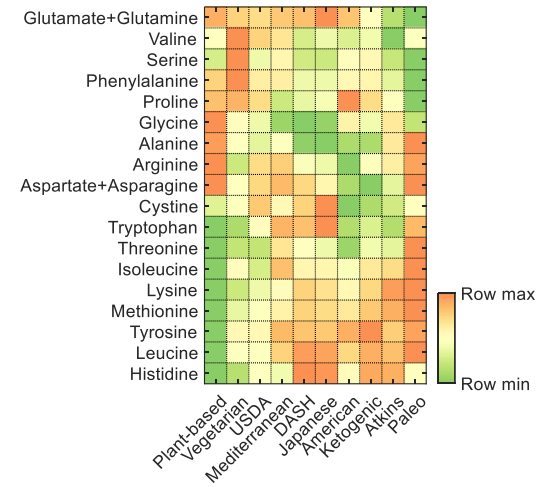
c

PCA of relative amino acid compositions of dietary patterns



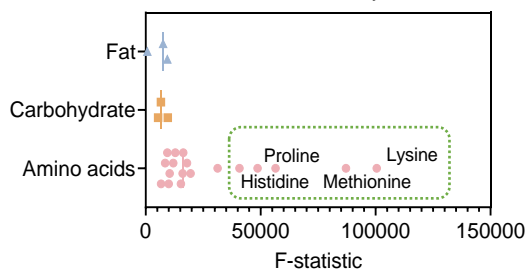
d

AA composition of human dietary patterns



e

F-stastic from one-way ANOVA



f

Top variable amino acids

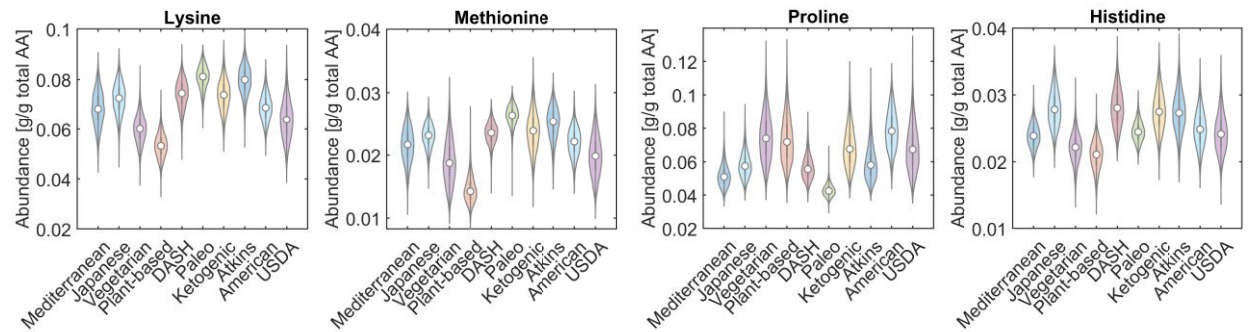


Figure 3. Landscape of human dietary amino acid intake

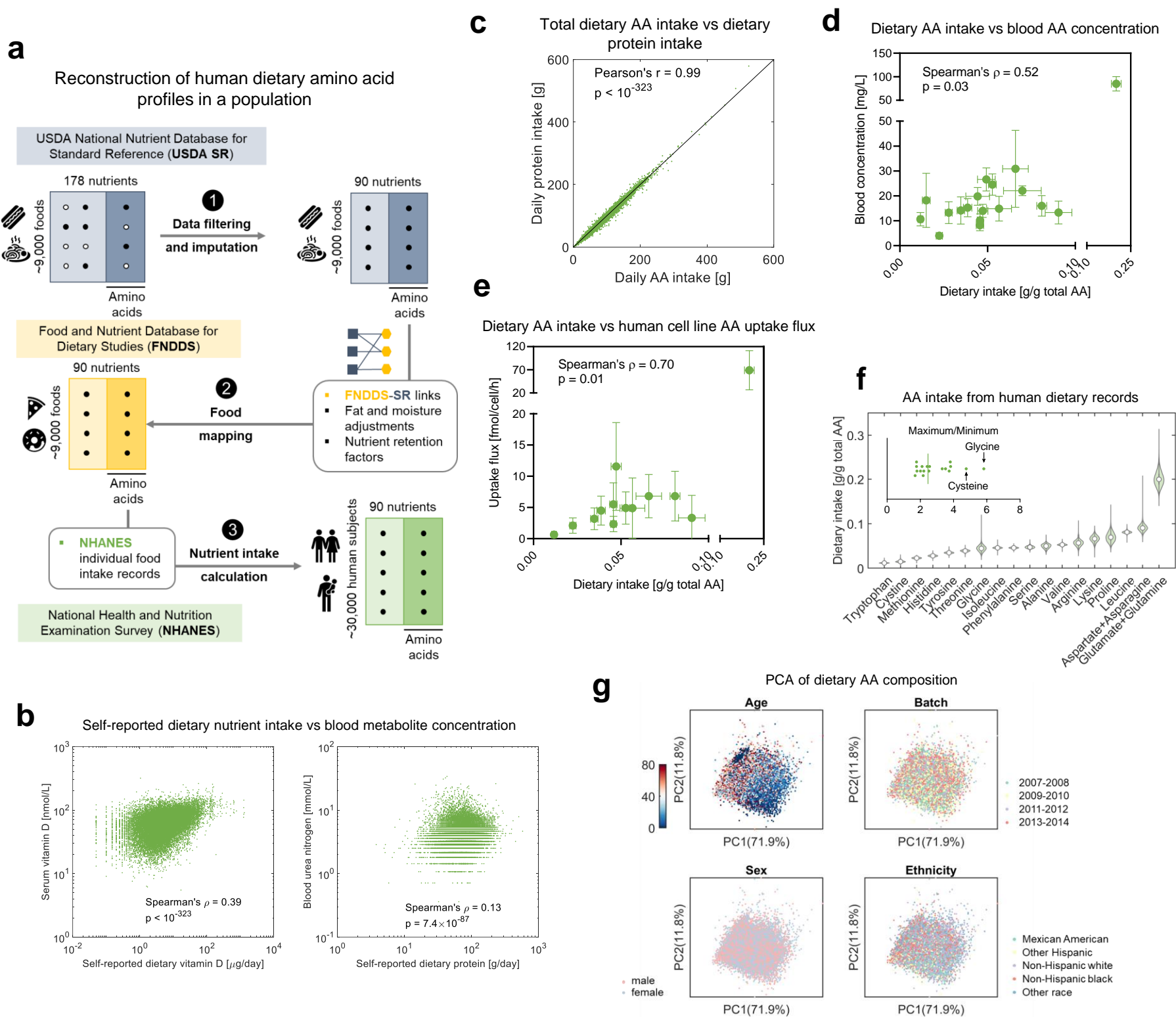


Figure 5. AI for dietary amino acid guidelines and personalized diet design

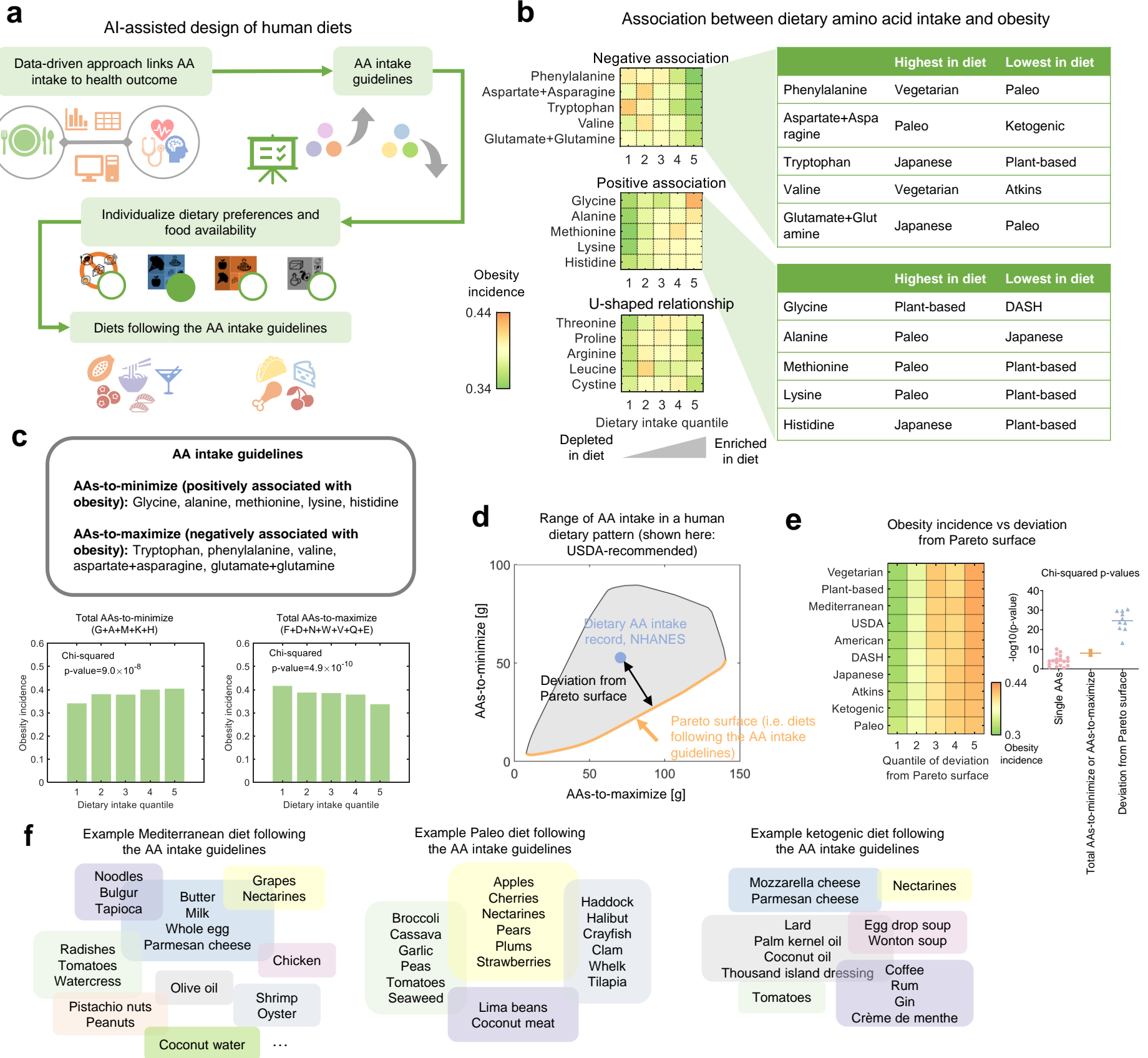


Figure S1 (Related to Figure 2)

Mediterranean diet

Food	Servings
Cereals	3~6/day
Fish	>=2/week
Fruits	3~6/day
Legumes	>=2/week
White meat	2/week
Processed meat	<=1/week
Sweets	<=2/week
Vegetables	>=6/day
Potatoes	<=3/week
Red meat	<=2/week
Olives/nuts/seeds	1~2/day
Olive oil	>=1/day
Eggs	2~4/week
Dairy	2/day
Others	0/day

Paleo diet

Food	Servings
Game meat/seafood	unlimited
Nuts/seeds	unlimited
Fruits/vegetables	unlimited
Honey	unlimited
Others	0/day

Vegetarian diet

Food	Servings
Meat	0/day
Others	unlimited

Japanese diet

Food	Servings
Cereals	5~7/day
Fruits	2/day
Vegetables	5~6/day
Dairy	2/day
Proteins	3~5/day
Others	0/day

American diet

Food	Consumption (+/-20%)
Meat	381g/day
Sweets	166g/day
Fats/oils	114g/day
Grains	291g/day
Starchy roots	164g/day
Vegetables	310g/day
Fruits	266g/day
Eggs	38g/day
Milk	704g/day
Legumes	9g/day
Alcoholic beverages	258g/day
Others	28g/day
Nutrient	Intake (+/-20%)
Calories	3641 kcal/day

Ketogenic diet

Nutrient	% kcal
Fat	>=70%
Carbohydrate	<=5%

USDA-recommended diet

Nutrient	Intake
Calories	1800~2200kcal/day
Protein	>=46g/day, 10~35%kcal
Carbohydrate	>=130g/day, 45~65%kcal
Dietary fiber	>=28g/day
Sugar	<=10%kcal
Fat	20~35%kcal
Saturated fat	<=10%kcal
Linoleic acid	>=12g/day
Linolenic acid	>=1.1g/day
Calcium	>=1g/day
Iron	>=18mg/day
Magnesium	>=310mg/day
Phosphorus	>=700mg/day
Potassium	>=4.7g/day
Sodium	<=2.3g/day
Zinc	>=8mg/day
Copper	>=0.9mg/day
Manganese	>=1.8mg/day
Selenium	>=55µg/day
Vitamin A	>=700mg/day
Vitamin E	>=15mg/day
Vitamin D	>=600IU/day
Vitamin C	>=75mg/day

Atkins diet

Nutrient	Intake
Carbohydrate	<=20g/day

USDA-recommended diet (continued)

Nutrient	Intake
Thiamin	>=1.1mg/day
Riboflavin	>=1.1mg/day
Niacin	>=14mg/day
Vitamin B6	>=1.3mg/day
Vitamin B12	>=2.4µg/day
Choline	>=425mg/day
Vitamin K	>=90µg/day
Folate	>=400µg/day

DASH diet

Food	Servings
Grains	6~8/day
Vegetables	4~5/day
Fruits	4~5/day
Low-fat dairy	2~3/day
Lean meat	<=6/day
Nuts/seeds/legumes	4~5/week
Fats/oils	2~3/day
Sweets	<=5/week
Others	0/day
Nutrient	Intake
Sodium	<=2.3g/day

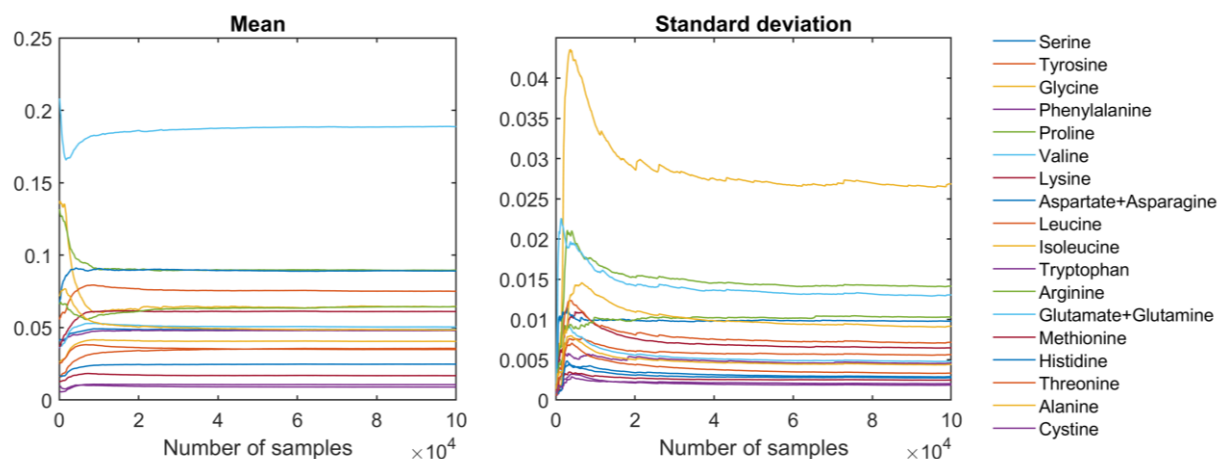
Plant-based diet

Food	Servings
Meat	0/day
Dairy	0/day
Eggs	0/day
Others	unlimited

Figure S2 (Related to Figure 2)

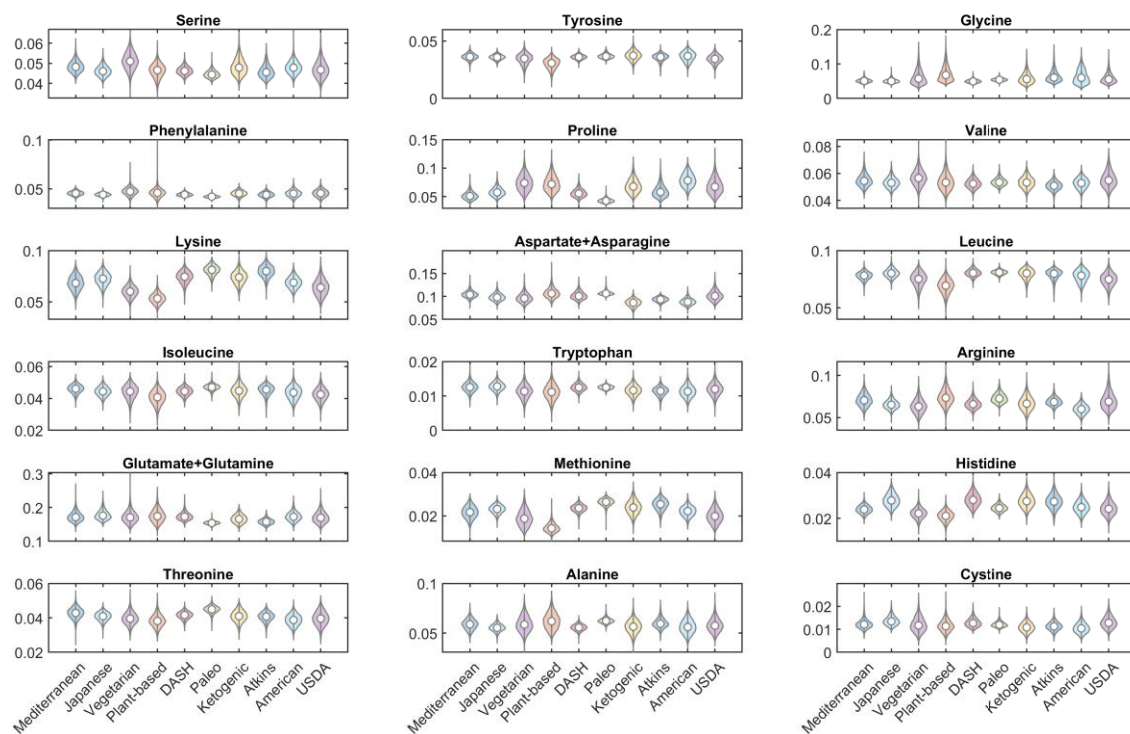
a

Convergence of mean and standard deviation values during the sampling process for USDA-recommended diet



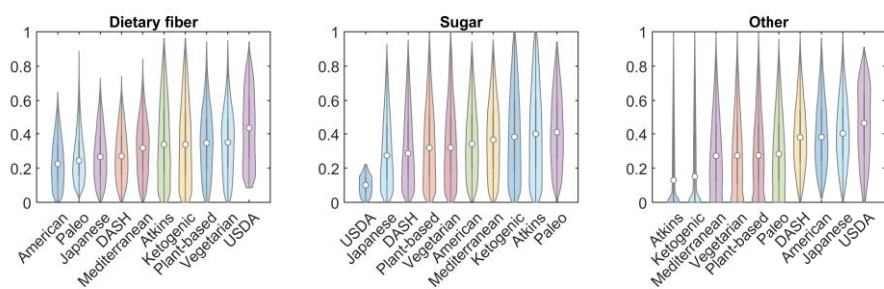
b

Distributions of relative AA composition in human dietary patterns

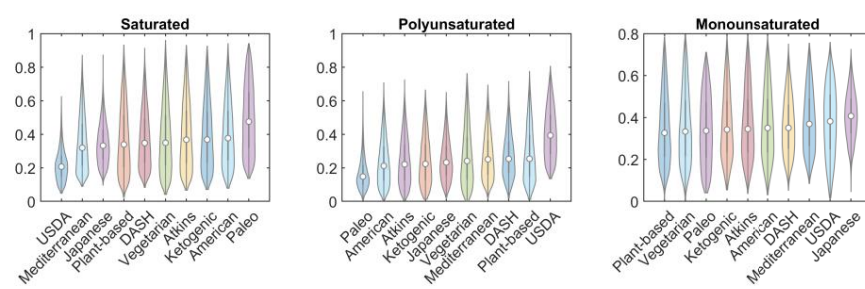


c

Distributions of carbohydrate composition in human dietary patterns

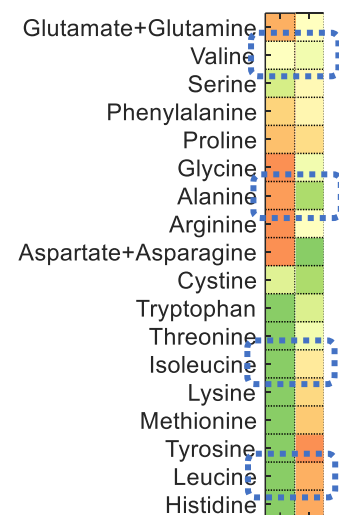


Distributions of fat composition in human dietary patterns



d

AA levels in plant-based diet and ketogenic diet



Plant-based diet
Ketogenic diet

Fasting blood concentrations of amino acids in individuals eating ketogenic or plant-based diets

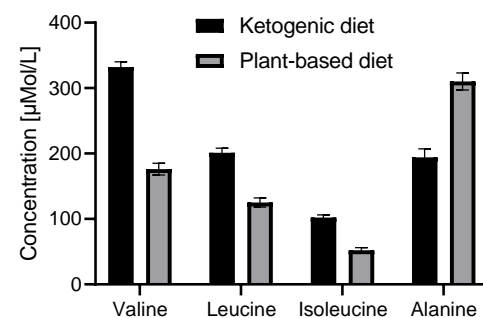
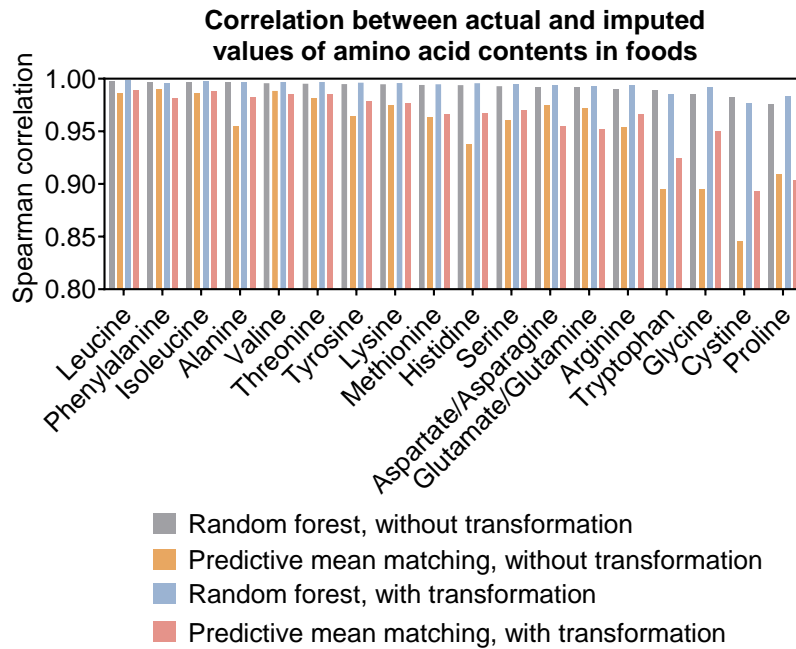


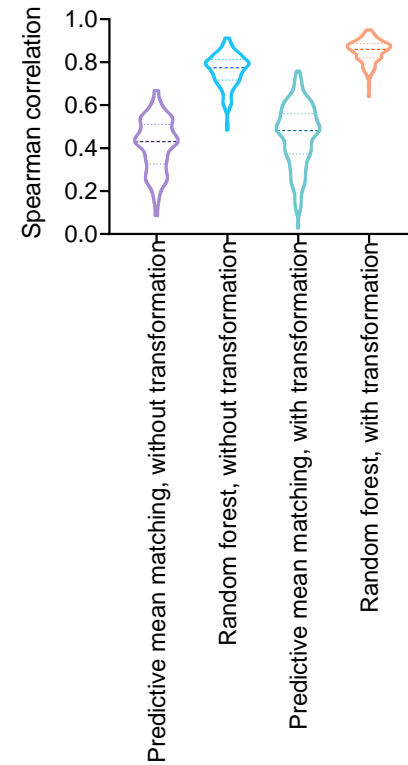
Figure S3 (Related to Figure 3)

a



b

Distributions of Spearman's rank correlation coefficient between actual and imputed pairwise ratios of AA contents in foods



c

Correlation between AA composition of human diets and culture medium

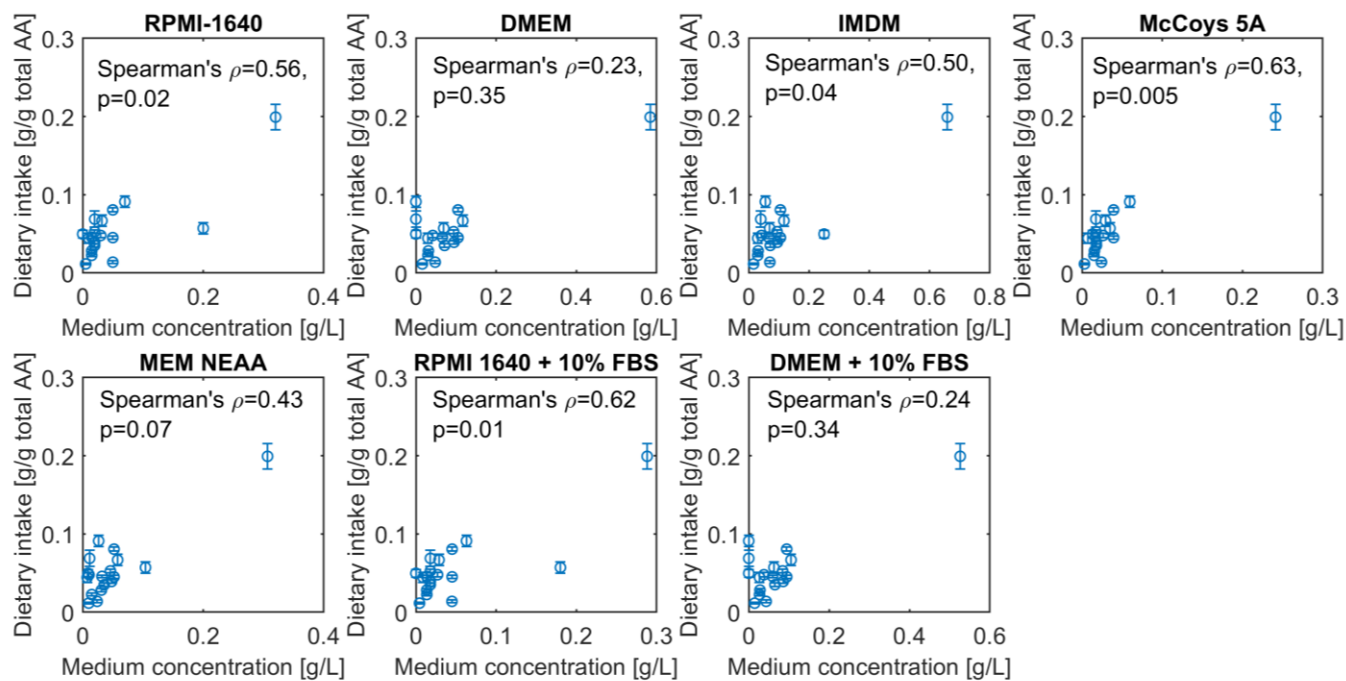


Figure S4 (Related to Figure 3)

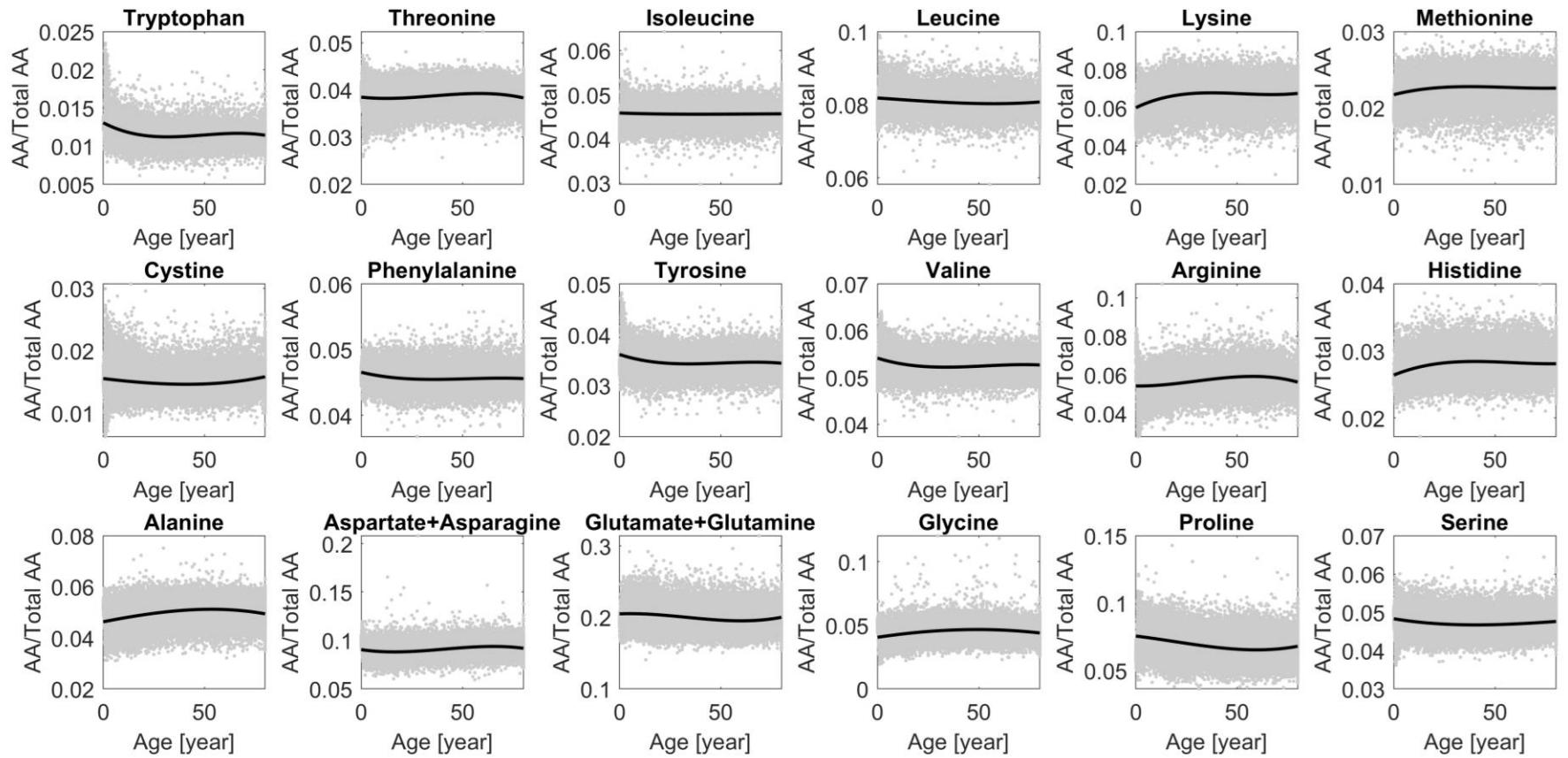


Figure S5 (Related to Figure 4)

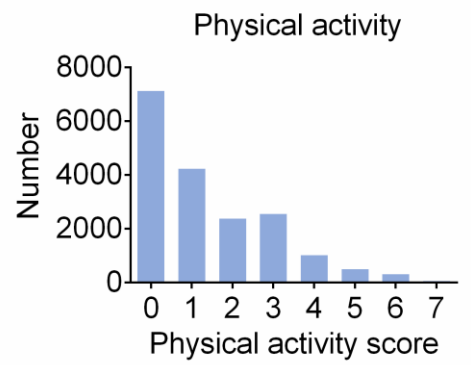
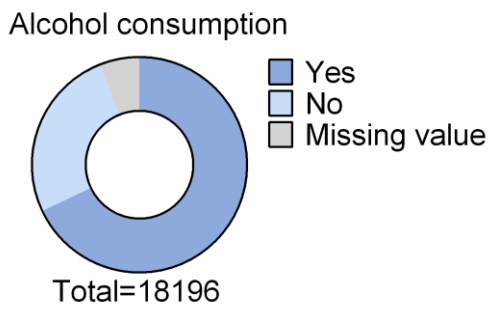
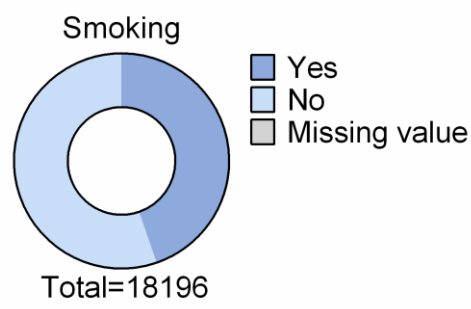
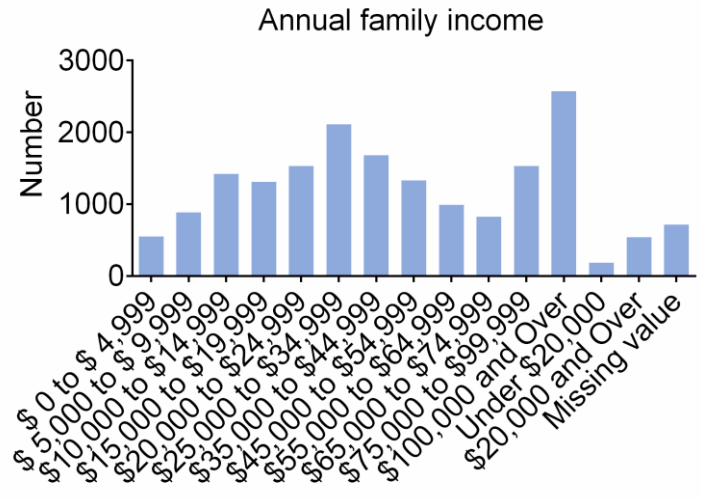
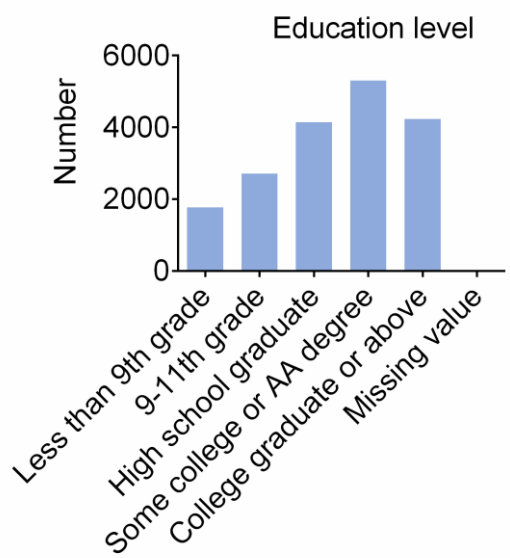
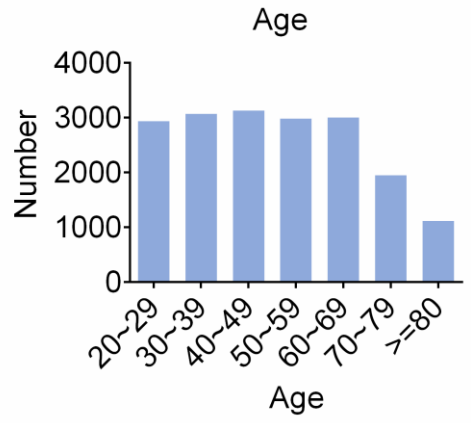
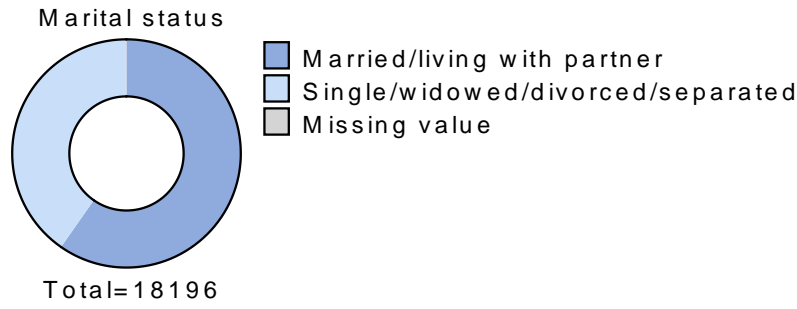
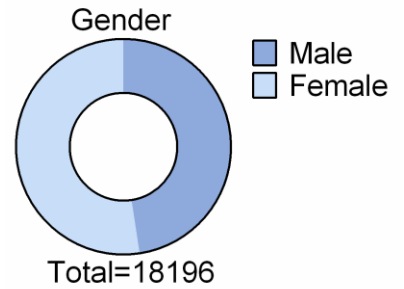
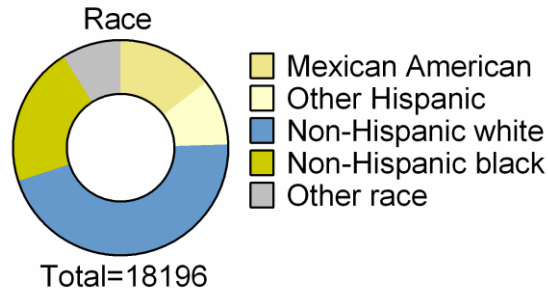
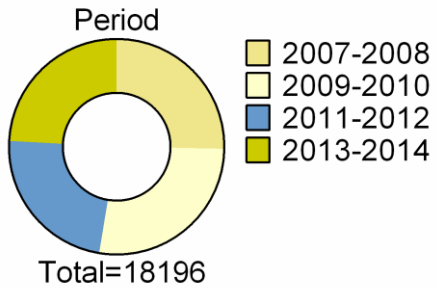
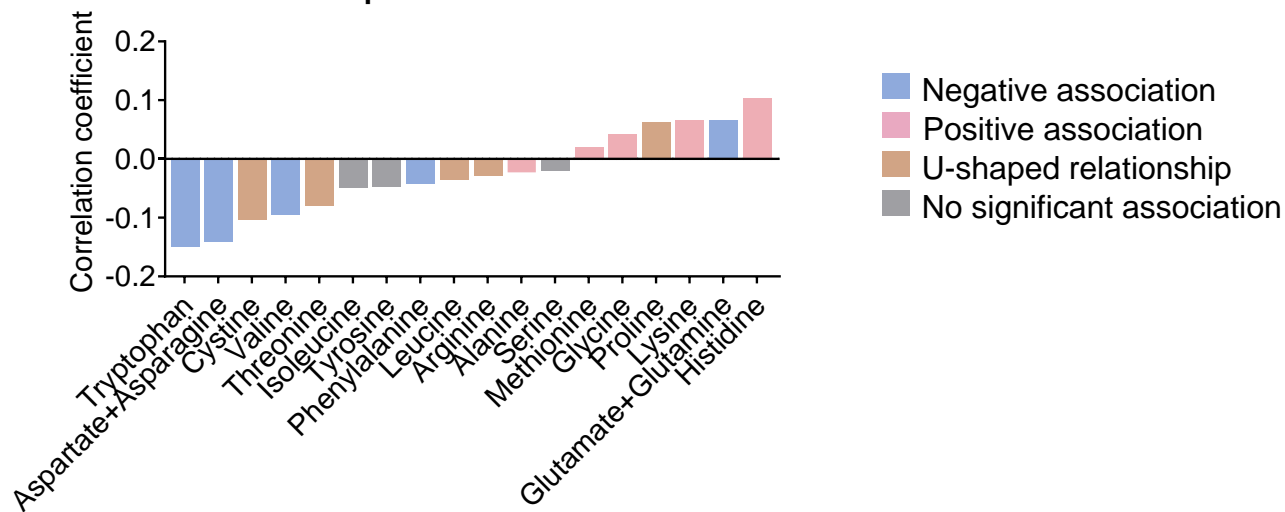


Figure S6 (Related to Figure 5)

a Spearman's rank correlation between dietary amino acid composition and total calorie intake



b

