Systems analysis of gut microbiome influence on metabolic disease in HIV and high-risk populations

Abigail J.S. Armstrong[1,2], Kevin Quinn[3], Sam X. Li[1], Jennifer M. Schneider[1], Nichole M. Nusbacher[1], Katrina A. Doenges[3], Suzanne Fiorillo[1], Tyson J. Marden[4], Janine Higgins[5], Nichole Reisdorph[3], Thomas B. Campbell[1], Brent E. Palmer[1#], Catherine A. Lozupone[1#]

Affiliations:

1. Department of Medicine, University of Colorado Denver, Anschutz Medical Campus, Aurora, CO, USA

2. Department of Immunology and Microbiology, University of Colorado Denver, Anschutz Medical Campus, Aurora, CO, USA

3. Skaggs School of Pharmacy and Pharmaceutical Sciences, University of Colorado Anschutz Medical Campus, 80045 CO, Aurora, USA

4. Colorado Clinical and Translational Sciences Institute; University of Colorado Anschutz Medical Campus, Aurora, Colorado, 80045. USA

5. Department of Pediatrics, Section of Endocrinology, University of Colorado Anschutz Medical Campus. Aurora, CO

Corresponding authors

Brent E. Palmer: 303-724-7203; brent.palmer@cuanschutz.edu

Catherine A. Lozupone: 303-724-7942; catherine.lozupone@cuanschutz.edu

Systems analysis of gut microbiome influence on metabolic disease in HIV and high-risk populations

Abstract

Poor metabolic health, characterized by insulin resistance and dyslipidemia, is higher in people living with HIV (PLWH) and has been linked with inflammation, anti-retroviral therapy (ART) drugs, and ART-associated lipodystrophy (LD). Metabolic disease is associated with gut microbiome composition outside the context of HIV but has not been deeply explored in HIV infection nor in high-risk men who have sex with men (HR-MSM), who have a highly altered gut microbiome composition. Furthermore, the contribution of increased bacterial translocation and associated systemic inflammation that has been described in HIV-positive and HR-MSM individuals has not been explored. We used a multi-omic approach to explore relationships between gut microbes, immune phenotypes, diet, and metabolic health across ART-treated PLWH with and without LD; untreated PLWH; and HR-MSM. For PLWH on ART, we further explored associations with the plasma metabolome. Sixty-nine measures of diet, gut microbes, inflammation, and demographics were associated with impaired metabolic health defined using fasting blood markers including lipids, glucose and hormones. We found microbiome-associated metabolites associated with metabolic disease including the microbially produced metabolites, dehydroalanine and bacteriohopane-32,33,34,35-tetrol. Our central result was that elevated plasma lipopolysaccharide binding protein (LBP) was the most important predictor of metabolic disease in PLWH and HR-MSM, with network analysis of predictors showing that LBP formed a hub joining correlated microbial and immune predictors of metabolic disease. Our results suggest the role of inflammatory processes linked with bacterial translocation (measured by LBP) and interaction with dietary components and the gut microbiome in metabolic disease among PLWH and HR-MSM.

48 **Importance Statement**

49 The role of the gut microbiome in the health of HIV infected individuals is of interest because
50 current therapies, while effective at controlling disease, still result in long term comorbidities.
51 Metabolic disease is prevalent in HIV-infected individuals even in well-controlled infection.
52 Metabolic disease has been linked with the gut microbiome in previous studies but little attention
53 has been given to HIV infected populations. Furthermore, integrated analyses that consider gut
54 microbiome composition together with data on diet, systemic immune activation, metabolites and
55 demographic data have been lacking. By conducting a systems level analysis of predictors of
56 metabolic disease in people living with HIV and men who are at high risk of acquiring HIV, we
57 found that increased LBP, an inflammatory marker indicative of compromised intestinal barrier
58 function, was associated with worse metabolic health. We also found this relationship to be
59 associated with dietary, microbial, and metabolic factors suggesting a systemic gut microbiome
60 influence on the presence of increased inflammatory markers which, in turn, influences the risk of
61 metabolic disease. This work lays the framework for mechanistic studies aimed at targeting the
62 microbiome and diet to prevent or treat metabolic endotoxemia in HIV-infected individuals.

63

64 **Keywords**

65 HIV; microbiome; men who have sex with men (MSM); metabolic disease

66

67 **Background**

68 The gut microbiome in people living with human immunodeficiency virus type 1(HIV) (PLWH) is
69 of interest as a potential contributor to infection, disease progression, and development of co-
70 morbidities. Poor metabolic health characterized by insulin resistance and dyslipidemia is frequent
71 in PLWH (1-3) and has been linked with chronic inflammation (4-7) and several anti-retroviral
72 therapy (ART) drugs (8). Metabolic disease is particularly prevalent in HIV-positive individuals
73 with lipodystrophy (LD), a disease linked with early ART drugs that is manifested by lipoatrophy

74 in the face, extremities, and buttocks with or without visceral fat accumulation. LD can have a
75 severe impact on the quality of life of PLWH and is associated with the development of diabetes
76 and cardiovascular disease (9).
77
78 Metabolic disease has been linked with gut microbiome structure and function outside the context
79 of HIV infection (10-14), but this relationship has not been explored deeply in PLWH. We and
80 others have found an altered gut microbiome composition in both PLWH (15-17) and men who
81 have sex with men at high-risk of contracting HIV (HR-MSM) (16, 18). Furthermore, we have
82 demonstrated that the altered microbiome in HIV (15) and HR-MSM (15, 19) are pro-inflammatory
83 both *in vitro* and/or in gnotobiotic mice (15, 19). This is of interest as peripheral inflammatory
84 signals have been implicated in both cardiovascular disease risk (7, 20) and insulin sensitivity (4,
85 5, 21-23) in PLWH. Increased peripheral immune activation in HIV-positive individuals is driven
86 in part by bacterial translocation (24, 25), as indicated by higher levels of the bacteria product
87 lipopolysaccharide (LPS) or LPS-binding protein (LBP) in blood. Increased blood LPS levels have
88 also been observed in MSM and linked with recent sexual behavior (26). An association between
89 LBP and metabolic disease in other diseases (e.g. hemodialysis patients) has been described
90 (27), however there are mixed data regarding a role in obesity associated metabolic disease (28-
91 30). Additionally, a recent study of metabolic syndrome in PLWH found greater immune
92 dysfunction and a more HIV-associated microbiome associated with risk of metabolic syndrome
93 (31).
94
95 We hypothesized that PLWH and HR-MSM with poor metabolic health would harbor a distinct gut
96 microbial signature that was in turn also associated with elevated peripheral immune activation.
97 We evaluated this relationship while considering other factors known to influence the microbiome,
98 immunity and metabolic health. This analysis included typical diet; HIV, ART, and LD status; and
99 other demographic characteristics such as age and body mass index (BMI). For HIV-positive

100 individuals on ART with and without LD, we further explored associations with the plasma
101 metabolome (Figure 1). Our results suggest a central role of inflammatory processes linked with
102 bacterial translocation as measured by LBP, and co-correlated intestinal microbes, dietary and
103 demographic attributes in metabolic disease risk.
104
105 **Results**
106 Study Population
107 This study examined a cohort of 113 men, including men who have sex with women (MSW; n=22,
108 19.5%) and men who have sex with men (MSM; n=91, 80.5%) (Table 1). Of the MSM, 32 were
109 HIV-negative (35.2%), 14 were HIV-positive and not on ART (15.4%), and 45 were HIV-positive
110 ART-treated (49.4%). The HIV-positive, treated group included those with lipodystrophy (LD;
111 n=25, 55.6%) and those without (n=20, 44.4%). The HIV-negative MSM participated in activities
112 that put them at high risk of contracting HIV as defined in a prior study of a candidate HIV vaccine:
113 1) a history of unprotected anal intercourse with one or more male or male-to-female transgender
114 partners; 2) anal intercourse with two or more male or male-to-female transgender partners; or 3)
115 being in a sexual relationship with a person who has been diagnosed with HIV (32). In order to
116 focus on HIV-associated metabolic disease, obese individuals (BMI >30) were excluded. There
117 was no significant difference in BMI between the cohorts (Kruskal-Wallis test, $p = 0.085$). HIV-
118 positive, treated cohorts were significantly older than HIV-negative MSM and HIV-positive,
119 untreated MSM (Kruskal-Wallis test, $p < 0.001$). Age matching across all cohorts was not feasible
120 in part because LD is associated with early-generation ART drugs and thus most common in older
121 HIV-positive individuals and HR-MSM behavior as well as new HIV infections are predominantly
122 in younger individuals. However, age is carefully considered in downstream analyses. All treated,
123 HIV-positive individuals were on successful ART with suppressed viral loads (Table 1).

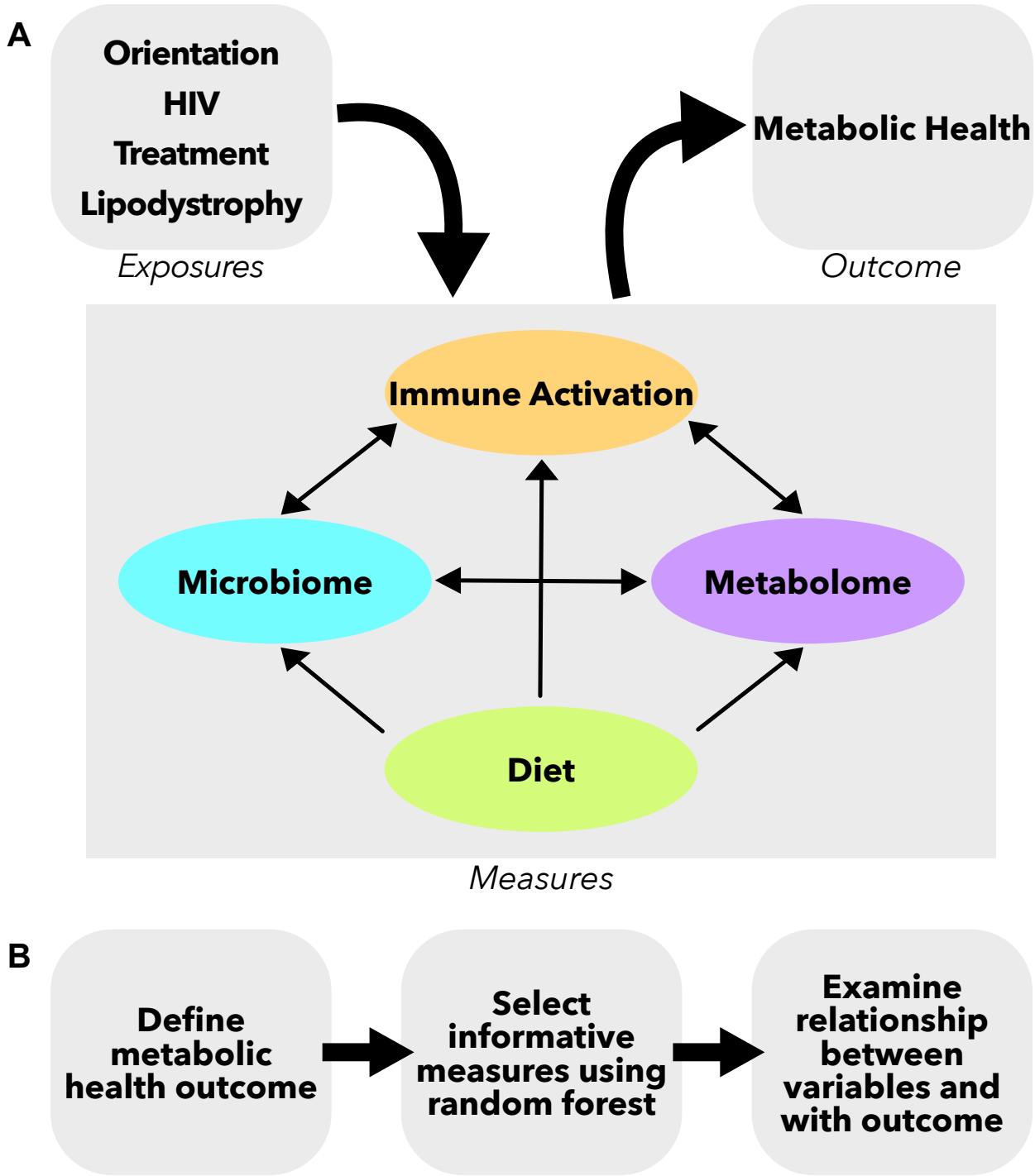


Figure 1. Study design schematic. Measures were collected from four compartments: gut microbiome, peripheral immune, diet questionnaire, and plasma metabolome. These separate compartments can all influence each other, forming complex systems that together influence metabolic health. Furthermore, the compartments can be influenced by other clinical and demographic characteristics such as HIV and treatment status. In this study we examine all of these measures together in order to investigate

131 **Table 1. Description of full study cohort**

	HIV-negative MSW	HIV-negative MSM	HIV-positive MSM, untreated	HIV-positive MSM, treated	HIV-positive MSM, treated, with LD
n	22	32	14	20	25
Age (years)	33 (27.3-38.5)	34 (29.8-44.5)	34 (26.5-40.3)	46 (42.8-50.5)	60 (54-64)
BMI (kg/m2)	25.2 (23.0-27.0)	25.5 (20.2-28.0)	21.4 (20.2-25.6)	23.9 (22.6-26.2)	25.8 (23.0-28.0)
CD4 cell count	NA	NA	538 (408.5-731.8)	586 (419.5-878.0)	659 (550.0-908.0)
Viral load	NA	NA	101,400 (20,300-292,514)	20 (0-20)	0 (0-20)
Cholesterol drugs/statins n (%)	2 (9.1%)	3 (9.4%)	1 (7.1%)	4 (20%)	14 (56%)

132 Numbers reported are median (IQR)

133 Metabolic Disease Score as a Marker for Metabolic Health

134 We measured seven common clinical markers of metabolic health from fasting blood:
135 triglycerides, glucose, insulin, LDL, HDL, leptin, and adiponectin. Since these markers are often
136 correlated with each other, we used principal component analysis (PCA) to define a single
137 continuous measure of overall metabolic health of study participants, as has been done previously
138 (Figure 2A) (33, 34). The first principal component (PC1) explained 28.8% of the variability within
139 the clinical marker data and separated individuals by multiple correlated markers of metabolic
140 disease. Specifically, individuals with higher values of PC1 generally had high triglycerides and
141 low HDL, indicating dyslipidemia, and higher levels of fasting blood glucose and insulin, indicating
142 insulin resistance (Figure 2A, Supplemental Figure 1). We thus decided to use values of PC1 as
143 an outcome. Values of PC1 were shifted to a minimum of one and log transformed to define the
144 metabolic disease score, which ranged from 0 as healthy and 2.5 as impaired. In order to
145 determine how this score related to metabolic health, we performed regressions between the
146 metabolic disease score and individual measures to define a metabolic score threshold that
147 corresponded with clinically defined cutoffs for normal levels (Supplemental Figure 1). For
148 example, triglycerides positively correlated with metabolic disease score and almost all individuals
149 with a score above 1.45 had triglyceride levels in the unhealthy range of greater than 200 mg/dL.
150 Similar patterns and cutoffs were true for HDL, LDL, and glucose (Supplemental Figure 1). The
151 intersect of the regression with these cutoffs were all averaged to a single number of 1.4.
152 Individuals below the cutoff were categorized as metabolically normal and those above were
153 categorized as metabolically impaired.

154

155 When comparing the metabolic disease score across cohorts, we found that ART-treated, HIV-
156 positive individuals with LD trended higher in both the average metabolic disease score and the
157 proportion of individuals with scores in the metabolically impaired group but intergroup

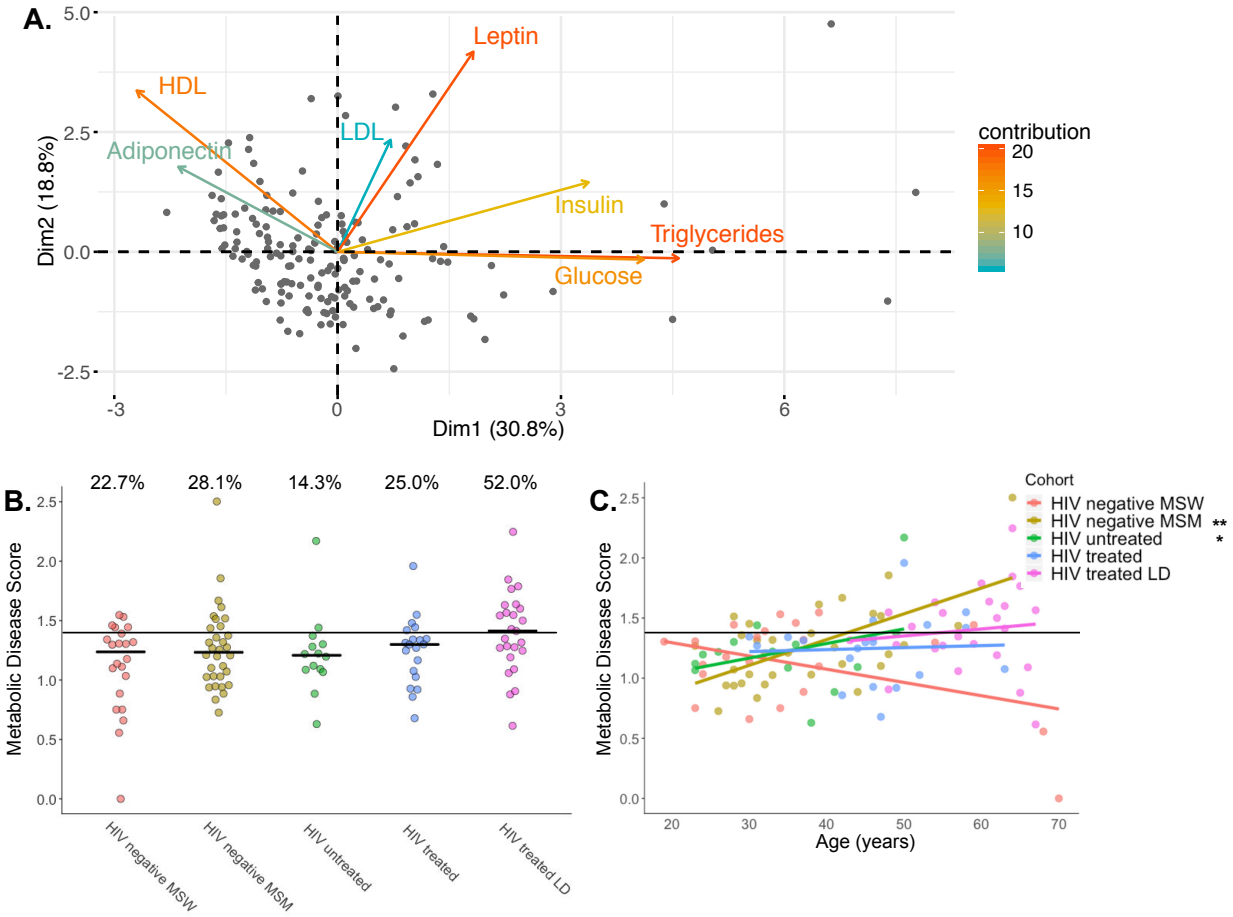


Figure 2. Metabolic disease score for marker of metabolic health. A. PCA of metabolic measures in fasting blood of 164 men and women: 113 participants described in this paper along with 51 individuals recruited at the same time and under the same exclusion criteria as study participants. Metabolic disease score is calculated as the PC1 coordinates shifted to a minimum of one and log transformed. **B.** Metabolic disease scores broken up by cohort. The percentages noted above groups are the percent of individuals with a score above our metabolic impairment cutoff (Supplemental Figure 1). There is no significant difference between the proportions in each group (Fischer's exact test, $p = 0.11$) or between mean ranks in each group (Kruskal-Wallis test $p = 0.13$). **C.** Relationships between metabolic disease score and age stratified by cohort. Statistical significance of slopes are indicated and were calculated with the linear model: score ~ age + cohort + age*cohort. P-value annotations: ** < 0.01; * < 0.05.

170 significance was lost after multiple test corrections (Figure 2B). Furthermore, because our HIV-
171 positive, treated cohorts were significantly older than our HIV-negative MSM and HIV-positive,
172 untreated MSM, we used a linear model to explore differences in the metabolic disease score
173 across cohorts while accounting for age (Figure 2C). This score was positively associated with
174 age only in HIV-negative MSM and HIV-positive, untreated MSM (Figure 2C; linear model; $p <$
175 0.001 and $p = 0.036$ respectively) and only HIV-negative MSM had significantly higher
176 metabolic disease score compared to HIV-negative MSW when accounting for age (linear
177 model; $p < 0.001$).
178
179 Selection of Features that Predict the Metabolic Disease Score and Interactions Between
180 Selected Features
181 To explore the complex relationships of the gut microbiome, peripheral immune activation, and
182 diet to the metabolic disease score and to each other, we first selected features that were
183 important predictors of the metabolic score using the tool VSURF (Variable Selection Using
184 Random Forest) tool (35). The VSURF implementation of random forest is optimized for feature
185 selection, returning all features that are highly predictive of the response variable, even when a
186 smaller subset of highly predictive variables with redundant features removed could be just as
187 accurate for prediction (35). We input the following features into the VSURF tool: 1) 130 microbial
188 features (99% identity Operational Taxonomic Units (OTUs) with highly co-correlated OTUs
189 binned into modules as described in the methods (detailed in Supplemental Table 1) and filtered
190 to OTUs only in >20% of samples). 2) 21 immune features that were measured in plasma using
191 multiplex ELISA (detailed in Supplemental Table 2). These immune measures were selected
192 based on a literature search for those previously shown to be altered in HIV infection and/or
193 metabolic disease. 3) 21 clinical/demographic features that were collected in questionnaires or
194 measured in study participants such as age, BMI, HIV infection and treatment status, typical
195 gastrointestinal symptoms including constipation, diarrhea and bloating, and sexual behavior

196 (detailed in Supplemental Table 3). 4) 29 dietary features that were collected using a food
197 frequency questionnaire of typical dietary intake over the prior year as detailed in the methods.
198
199 From the initial 201 measures, VSURF identified 69 important variables (four clinical data
200 measures, six diet measures, 14 immune measures, 45 microbes) and a subset of ten highly
201 predictive variables (Supplemental Table 4). These 69 features were sufficient to accurately
202 predict metabolic disease score using traditional random forest (linear model: $r^2 = 31.05\%$, $p <$
203 0.001). Additionally, permutation testing revealed that VSURF performed better at selecting
204 explanatory variables than a null model where the outcome was randomly permuted (permutation
205 test; $p = 0.049$, Supplemental Figure 2). We found that 21 of the 69 selected variables were
206 positively or negatively correlated with the metabolic disease score, indicating either increased or
207 decreased risk respectively (Spearman rank correlation, FDR $p < 0.1$, Supplemental Table 4).
208 Since random forest can detect non-linear relationships and/or features that are only important
209 when also considering another feature, it is not surprising that all features were not correlated
210 linearly with the metabolic disease score.
211
212 All VSURF selected clinical measures were positively correlated with metabolic disease score
213 and included age, BMI, lipodystrophy, and bloating (Supplemental Table 4). None of the six
214 selected diet measures correlated with metabolic disease score (Supplemental Table 4). VSURF
215 selected several inflammatory immune measures that were positively correlated with metabolic
216 disease score: LBP, intercellular adhesion molecule 1 (ICAM-1), interleukin (IL) 16, IL-12, and
217 granulocyte-macrophage colony-stimulating factor (GM-CSF) (Supplemental Table 4). The most
218 important feature as determined by random forest importance score was LBP.
219
220 Diet, the microbiome, and immune phenotypes can all influence each other (Figure 1). They can
221 also relate to the measured clinical/demographic factors that we had identified as predictors of

222 the metabolic disease score such as BMI and age. For this reason, we also investigated the
223 relationship between the 69 important factors using pairwise Spearman rank correlation and
224 network visualization (Figure 3, Supplemental Figure 3, Supplemental Table 5). This approach
225 revealed many within data type associations such as positive correlations within many selected
226 dietary, microbiome, and immune features. It also identified correlations between data types such
227 as a negative relationship between LBP and several Gram-negative bacteria or a positive
228 correlation between age and immune measures such as LBP or IL-6 (Figure 3A, Supplemental
229 Table 5).
230
231 The selected important microbes included many that positively or negatively correlated with the
232 metabolic disease score (Supplemental Table 4) and that were highly co-correlated with each
233 other and with dietary, clinical/demographic, and inflammatory phenotypes (Supplemental Figure
234 3A, Supplemental Table 5). For example, a module of bacteria identified within the *Prevotella*
235 genus and the *Paraprevotellaceae* family, negatively associated with metabolic disease score
236 and positively associated with dietary fiber (Supplemental Figure 3B). Because bacterial
237 translocation is known to occur at increased levels in both HIV-positive individuals (36) and HR-
238 MSM (26), we were specifically interested in looking at the selected microbes and other features
239 that correlated with LBP, a marker of bacterial translocation. The network of LBP neighbors is
240 shown in Figure 3B. All of the microbes correlating with LBP were classified to the order
241 *Clostridiales*. More specifically, LBP is negatively correlated with several butyrate or putative
242 butyrate producing bacteria/bacterial modules such as OTUs in the genera *Coprococcus* (37, 38).
243 LBP was also positively correlated with *Dorea* species (Supplemental Table 5).
244
245 In addition to correlations, we evaluated interactions between variables using the tool iRF
246 (iterative Random Forest)(39). These interactions represent variables that are in adjacent nodes

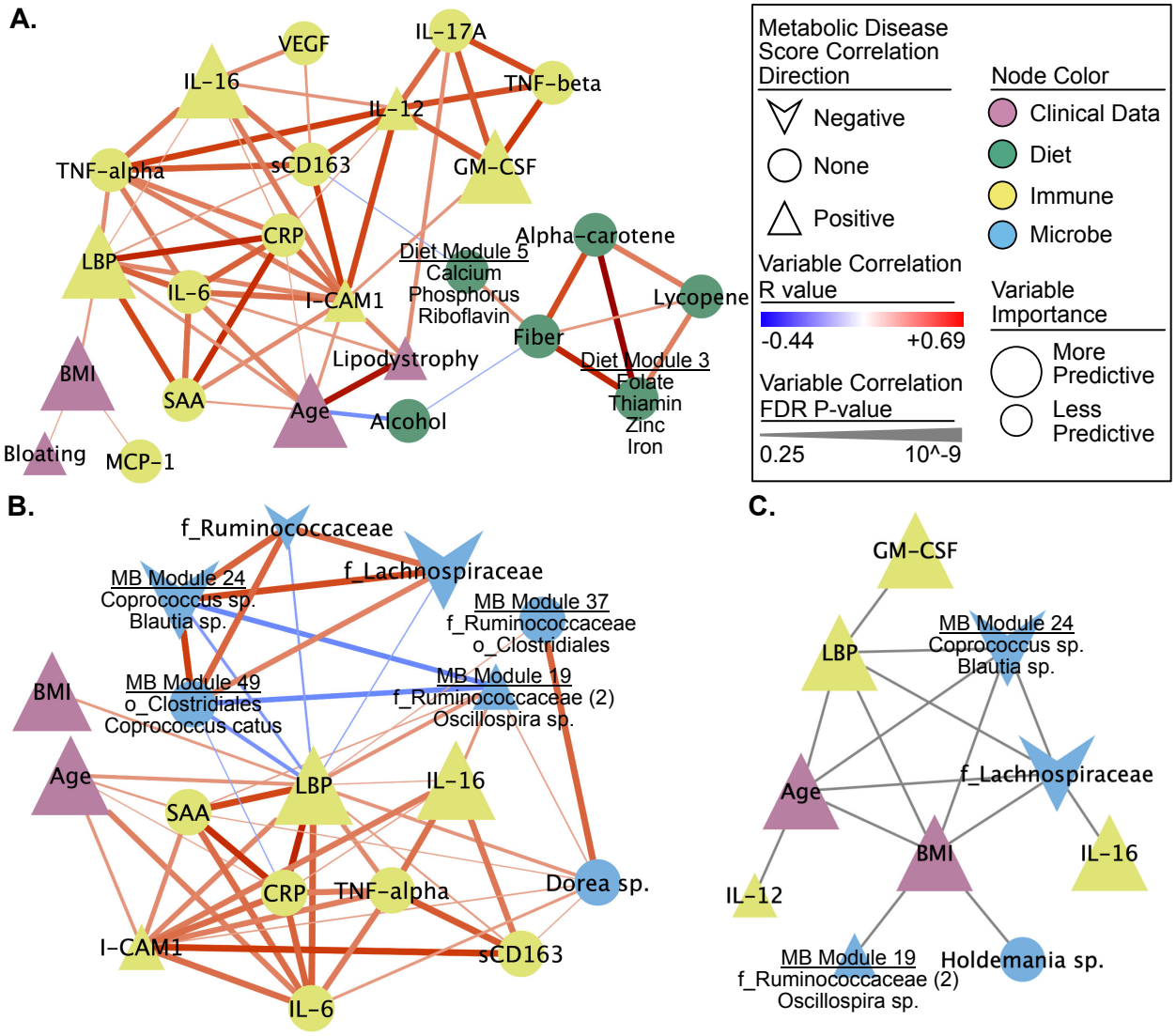


Figure 3. Networks of selected measures reveal several strong associations with metabolic disease score and between measures. Correlation sub-networks of **A.** all the non-microbe selected measures. **B.** the nearest neighbors of LBP. All Spearman rank correlations with an FDR $p < 0.25$ are shown. Subnetworks were pulled from a larger network of all VSURF selected measures (Supplemental Figure 3, Supplemental Table 2). **C.** Network of interactions between measures calculated using iRF. All edges represent an interaction (i.e. proximity in a decision tree) that occurred in 30% or more of the decision trees.

255 in a random forest tree in which the value of one influences the predictability of the other.
256 Interactions between variables in more than 30% of the trees were kept for further analysis (Figure
257 3C). This analysis identified a group of 5 interactive features: LBP, age, BMI, an OTU in the
258 *Lachnospiraceae* family, and microbiome module 24 (*Coprococcus* sp. and *Blautia* sp.). These
259 features were also significantly correlated with LBP and suggest a subset of features that when
260 taken together may be predictive of metabolic disease risk.
261
262 <u>The Plasma Metabolome as a Potential Mechanism of Microbial Influence on Metabolic Disease</u>
263 <u>Score in ART-treated HIV-positive Individuals with and without LD</u>
264 To pursue a further mechanistic understanding of how the gut microbiome may influence the
265 metabolic disease score in PLWH, we performed untargeted metabolomics (LC/MS) on plasma
266 from our cohort of ART-treated, HIV-positive individuals with and without LD (n=44). Metabolite
267 identities were then validated using untargeted MS/MS. We used two approaches to determine
268 which plasma metabolites were either directly produced or indirectly influenced by the presence
269 of a microbiome. First, metabolites found in the human plasma were run through the
270 computational tool AMON (40), which uses the KEGG database (41) and inferred metagenomes
271 (calculated using PICRUSt2 (42)) to determine which of the measured metabolites could have
272 been produced by the microbiome. Second, plasma from both germ free (GF) and humanized
273 mice was analyzed using metabolomics to determine metabolites that had significantly altered
274 levels in mice with human microbiomes compared to GF mice, i.e. microbiome influenced
275 metabolites. For this purpose, GF mice were gavaged using fecal samples from eight men from
276 the study cohort (humanized mice) while two mice were gavaged using PBS as control
277 (Supplemental Table 6). Plasma was collected before and after gavage. All mice were fed a high-
278 fat western diet.
279

280 We found that 820 metabolites were different in abundance between GF and humanized mice
281 after multiple test corrections (Student's t-test, FDR $p < 0.05$), 493 of which were also present in
282 the human plasma samples. However, only 376 of these 493 metabolites could be annotated (see
283 Methods and Supplemental Table 7) while the remaining 148 were only assigned a mass. From
284 the full set of 5,332 metabolites identified in the human plasma, 416 were able to be annotated
285 with KEGG IDs. These were further analyzed using AMON. 146 microbiome-associated
286 metabolites were identified that are putatively produced by the gut microbiome; however, many
287 of these could also be produced by the host. Twenty-six of the 134 microbiome-associated
288 metabolites identified by AMON were also identified in the gnotobiotic mouse analysis
289 (Supplemental Table 7).
290
291 Of the 5,332 total measured metabolites in the human samples, 150 correlated with metabolic
292 score (Spearman rank correlation, FDR $p < 0.05$; Supplemental Table 8). Of these 68 could not
293 be annotated. The annotated correlated compounds were enriched in a number of different
294 metabolic pathways with both the Phospholipid and the Glycerolipid pathways of the Small
295 Molecule Pathway Database (SMDPH)(43) being highly enriched (Supplemental Table 9).
296 Consistent with the metabolic score being defined in part by dyslipidemia, 17 of the significant
297 compounds were annotated as triglycerides. Of the 150 correlated metabolites, seven were
298 associated with the microbiome either because they were predicted microbial products of the gut
299 microbiome (as determined with AMON (40)), because they were significantly different while
300 comparing the metabolome of germ-free mice to that of mice colonized with the feces of study
301 participants, or by literature search. We confirmed the identity of 5 of the 7 of these with MS/MS
302 (Supplemental Table 8). Of these seven microbiome-associated metabolites, two could
303 exclusively be explained by direct production by the microbiome. Specifically, dehydroalanine was
304 identified as a microbial product with AMON and negatively correlated with metabolic disease
305 score (Supplemental Table 8). Bacteriohopane-32,33,34,35-tetrol is a bacterial metabolite that

306 positively correlated with the metabolic disease score. Two additional microbiome-associated
307 compounds were triglycerides (TG(54:6) and TG (16:0/18:2.20:4)) that were positively correlated
308 with metabolic disease score and elevated in humanized mice compared to GF mice. Another of
309 these metabolites, 1-Linleoyl-2-oleoyl-rac-glycerol is a 1,2-diglyceride in the triglyceride
310 biosynthesis pathway. Finally, phosphatidylcholine (PC(17:0/18:2)) and
311 phosphatidylethanolamine (PE(20:3/18:0)) compounds were identified as microbiome associated
312 and positively and negatively correlated with the metabolic disease score respectively (44).
313
314 **Discussion**
315 In this study, we identified several bacterial, diet, and immune measures that predicted higher
316 metabolic disease score in a cohort of MSM with and without HIV, ART, and LD. Notably, we
317 identified a strong relationship between circulating LBP and higher metabolic disease score which
318 in turn correlated with other markers of systemic inflammation, a loss of beneficial microbes such
319 as Gram-positive, butyrate-producing bacteria, and higher BMI, indicating that diverse modifiable
320 factors may influence LPS/inflammation driven metabolic disease in this population.
321
322 There was a positive association between metabolic disease score and age, as has been reported
323 previously for non-HIV populations (45), but linear modeling suggested that this relationship was
324 driven by an association in HIV-negative MSM and HIV-positive untreated MSM in our study,
325 revealing a possibly larger effect size than in our other cohorts. Also, when controlling for age,
326 HIV-negative and positive untreated MSM had the highest metabolic disease score, even
327 compared to HIV positive individuals on ART with LD, a population that has previously been
328 reported to have higher incidence (46). This result is intriguing given our results supporting a role
329 for LBP driven inflammation in metabolic disease and prior research linking increased levels of
330 LPS in blood with high-risk behavior in MSM (26). Larger cohorts and more detailed behavior

331 information are required, however, to make any definitive claims on impaired metabolic health in
332 ageing in HR-MSM.
333
334 Consistent with prior studies that have associated high BMI with dyslipidemia, insulin resistance,
335 and/or metabolic syndrome (9, 47, 48), BMI was a positive predictor of metabolic disease score
336 in our cohort even though our study excluded obese individuals, but did include overweight. This
337 suggests the importance of weight management even among overweight, non-obese individuals
338 as a strategy for reducing metabolic health impairment in this population.
339
340 We did not find a positive association between ART treatment status and metabolic disease score,
341 but this may be because individuals in our study were on a wide variety of drug combinations with
342 the potential to have varied/contrasting effects. For instance, both integrase stand transfer
343 inhibitors (ISTI) (49) and regimens including the nucleoside reverse transcriptase inhibitor (NRTI)
344 tenofovir have been shown to increase risk of weight gain (50). Conversely, the CCR5 antagonist,
345 maraviroc, may confer a benefit to cardiovascular function and body weight maintenance and
346 evidence in mice suggests that this may be linked to differences in gut microbiome composition
347 with treatment (21, 51). Thus, future studies more targeted to particular ART regimes will be
348 required to look at factors important in particular drug contexts.
349
350 Several of the dietary components identified as important predictors of metabolic disease score
351 in our cohort have been previously associated with metabolic health, including dietary carotenoid,
352 lycopene, and fiber (52-56). Fiber's benefit in glucose response has been linked with the activity
353 of *Prevotella copri*. Individuals who had improved glucose response upon 3 days of a high-fiber
354 diet consumption were characterized by a higher increase in *P. copri* (55) and beneficial effects
355 of *P. copri* were confirmed in mice fed a high-fiber diet (55). Interactions between *Prevotella*,
356 dietary fiber, and metabolic health were of particular interest in this cohort of HIV positive and

357 negative MSM since these individuals have much higher Prevotella, including *P. copri*, than in
358 non-MSM (16, 18). However, other published studies suggested that high Prevotella might predict
359 increased risk of metabolic disease. One group observed that *P. copri* in mice fed a western diet
360 low in fiber could promote poor glucose response through the production of branched chain amino
361 acids (BCAAs) (12). Additionally, our prior study using *in vitro* stimulations of human immune cells
362 with fecal bacteria of HIV positive and negative MSM indicated that the Prevotella-rich
363 microbiomes of MSM could drive systemic inflammation (15). Interestingly, in our data, a module
364 of three OTUs, two of which are identified as within the genus *Prevotella* and the other within the
365 family *Paraprevotellaceae,* negatively associated with metabolic disease score and positively
366 associated with dietary fiber (Supplemental Figure 3B), supporting a relationship between
367 particular *Prevotella* strains and dietary fiber towards improved metabolic health, and not
368 supporting deleterious effects. Further work will be needed to decompose the complex
369 relationship between dietary fiber, particular Prevotella strains, and metabolic health in HIV
370 positive and negative MSM with unique Prevotella-dominated communities.
371
372 LBP was the most important feature in the random forest analysis and also a highly interactive
373 measure in the iRF analysis. LBP binds to both microbial LPS and lipoteichoic acid (LTA) in the
374 blood (57) and the presence of elevated LBP is indicative of increased intestinal barrier
375 permeability (58). LBP levels were correlated with other inflammatory markers that have been
376 linked with worse metabolic health in HIV-negative populations suggesting a role as a central
377 mediator of metabolic-disease associated immune phenotypes. These included 1) I-CAM 1,
378 whose expression in adipose tissue has been associated with diet-induced obesity in mice (59)
379 and metabolic syndrome in humans (60) 2) IL-6, a pro-inflammatory cytokine that has been shown
380 to play a direct role in insulin resistance (61), and 3) SAA, which is regulated in part by IL-6 and
381 plays a role in cholesterol metabolism (60); SSA3 specifically has been shown to be produced in
382 response to gut bacteria in obesity mice (62). We observed a positive association between

383 metabolic disease score and frequency of abdominal bloating, further supporting a role of
384 intestinal dysfunction in this population. Taken together these associations suggest that
385 inflammation originating from an impaired intestinal barrier is promoting worse metabolic health.
386
387 Although prior studies have connected LBP-associated inflammation with worse metabolic health
388 (27, 28, 63); the strength of this relationship is disease specific with less clear results in obesity-
389 associated metabolic disease (29, 30). An importance of bacterial translocation in HIV-associated
390 metabolic syndrome was demonstrated in a recent study of metabolic comorbidities in HIV-
391 positive individuals which found that lower CD4 nadir and/or AIDS events, HIV-associated
392 microbiota, and low alpha diversity was correlated with increases in sCD14 and LBP and increase
393 risk of metabolic syndrome (31). Additionally, in our study LBP was correlated with age and BMI,
394 a relationship that was previously observed in a cohort of HIV-negative men of African ancestry
395 with this trio being further associated with adiposity and pre-diabetes (64). Lastly, the negative
396 association of LBP with putative butyrate producing bacteria suggests that a lack of microbes that
397 promote intestinal barrier integrity contributes to increased intestinal permeability and thus
398 microbial components in circulating blood.
399
400 In our metabolomic analysis, we identified 150 metabolites in blood that correlated with metabolic
401 disease score. In order to identify compounds whose prevalence may be related to the gut
402 microbiome we used two complimentary approaches. First, we predicted which of these
403 compounds could have been produced by the microbiome using information in KEGG and the
404 bioinformatics tool AMON (40). Second, we measured which compounds changed in relative
405 abundance in germ-free versus mice colonized with feces from our study cohort. The AMON
406 analysis allows us to specifically evaluate which compounds could have been directly produced
407 by the gut microbiome but is limited by a lack of KEGG annotations for many compounds (40).
408 The gnotobiotic mouse experiments can identify microbial influence in unannotated compounds

409 but cannot differentiate between direct production/consumption by microbes versus indirect
410 influence. The results will also be influenced by physiological differences between mice and
411 humans and the incomplete colonization of human microbes in humanized mice. Although these
412 weaknesses may have led us to underestimate which of the 150 metabolic disease associated
413 compounds may have been related to the microbiome, it still identified compounds that supported
414 a mechanistic link between gut microbes, metabolites, and metabolic disease in HIV-infected
415 individuals on ART.
416
417 Firstly, we found a negative correlation between the microbially-produced non-canonical amino
418 acid, dehydroalanine, and metabolic disease score. Dehydroalanine is a component of lantibiotics
419 that are active against Gram-positive bacteria. We observed Gram-positive *Dorea* to positively
420 correlate with LBP, suggesting a role of lantibiotics in regulating our proposed LBP-centered
421 metabolic disease in this population.
422
423 Secondly, we observed that bacteriohopane-32,33,34,35-tetrol positively correlated with the
424 metabolic disease score. This compound has been found to be a lipoxygenase inhibitor that
425 prevents the formation of hydroxyicosatetraenoic acid and various leukotrienes from arachidonic
426 acid (65), which have been linked with the development of cardiovascular disease and metabolic
427 syndrome (66). This association of a potentially protective metabolite increased in metabolic
428 impairment seems counterintuitive; however, it may be indicative of larger systemic changes in
429 arachidonic acid metabolism and is worthy of further exploration.
430
431 Thirdly, we identified a PC and a PE associated with both the microbiome and metabolic disease
432 score. Changes in PCs and/or PEs have been previously implicated in atherosclerosis, insulin
433 resistance and obesity (44). AMON analysis indicated that both PCs and PEs can be synthesized
434 by intestinal bacteria; however, these compounds can also be synthesized in the host and may

435 be found in the diet. In our analysis, PE(18:1/20:1) levels were higher in colonized compared to
436 germ-free mice indicating that intestinal bacteria do influence overall levels despite diverse
437 potential sources.
438
439 Lastly, we observed increased levels of several plasma triglycerides in the humanized compared
440 to germ-free mice, including two plasma triglycerides that were significantly associated with
441 metabolic disease score. This confirms the influence of the gut microbiome on host plasma
442 triglycerides (67-69). However, we did not find any strong associations between these
443 triglycerides and specific microbes within our dataset, indicating a potential need for studies
444 conducted in larger cohorts or with shotgun metagenomics to look for functional correlates.
445
446 In conclusion, we observed a relationship between diet, gut microbiome, plasma metabolome,
447 and peripheral immune markers of inflammation and metabolic disease in MSM. These data pull
448 together several previously described relationships between pairs of these compartments
449 observed in other populations. However, in this work we demonstrate both a novel collection of
450 measures (microbiome, peripheral immune signaling, peripheral metabolites, demographic and
451 diet information) and a novel approach for integrating several host compartments in order to
452 examine a more complex system and applied it to the little studied population of HIV negative and
453 positive MSM with and without LD. Our results suggest that an overall gut environment driven by
454 low fiber, key vitamins, and microbes that promote intestinal barrier integrity and high in potentially
455 pathogenic bacteria may work in conjunction to increase levels of LBP and other inflammatory
456 cytokines to drive poor metabolic health. These results illuminate potential microbiome-targeted
457 therapies and personalized diet recommendation given an interacting set of gut microbes and
458 other host factors. Understanding these relationships further may provide novel treatments to
459 improve the metabolic disease and inflammatory outcomes of MSM living with HIV.
460

461 **Methods**

462 Subject Recruitment

463 Participants were residents of the Denver, Colorado metropolitan area and the study was
464 conducted at the Clinical Translational Research Center of the University of Colorado Hospital.
465 The study was reviewed and approved by the Colorado Multiple Institutional Review Board and
466 informed consent was obtained from all participants. For detailed criteria on recruitment of our
467 five cohorts (HIV negative MSW; HIV negative MSM; HIV positive, ART naïve MSM; HIV positive
468 ART-treated MSM with LD; and HIV positive ART-treated MSM without LD) see supplemental
469 methods.

470

471 Feces, a fasting blood sample, and clinical surveys were collected from participants in order to
472 obtain analytes for the study design outlined in Figure 3 (Supplemental Table 2). To evaluate
473 metabolic health, we measured seven common clinical markers from fasting blood: triglycerides,
474 glucose, insulin, LDL, HDL, leptin, and adiponectin. Additional information about relevant clinical
475 measures such as probiotic use were also collected via a questionnaire and study participants
476 also filled out information on typical frequency of high-risk sexual practices and on typical levels
477 of gastrointestinal issues such as bloating, constipation, nausea and diarrhea.

478

479 Diet Data FFQ Collection

480 Typical dietary consumption over the prior year was collected using Diet History Questionnaire II
481 (70). Diet composition was processed using the Diet*Calc software and the
482 dhq2.database.092914 database (71). All reported values are based on USDA nutrition
483 guidelines. Reported dietary levels were normalized per 1000 kcal. To reduce the number of
484 comparisons within the diet survey data, we binned highly co-correlating groups of measures
485 within the data types into modules (Supplemental Table 3). These modules were defined using
486 the tool, SCNIC (72).

487

488 Immune Data Collection

489 Whole blood was collected in sodium heparin vacutainers and centrifuged at 1700rpm for 10
490 minutes for plasma collection. Plasma was aliquoted into 1mL microcentrifuge tubes and stored
491 at -80. For ELISA preparation, plasma was thawed, kept cold, and centrifuged at 2000xg for 20
492 minutes before ELISA plating. Markers for sCD14, sCD163, and FABP-2 were measured from
493 plasma using standard ELISA kits from R&D Systems (DC140, DC1630 &DFBP20). Positive
494 testing controls for each ELISA kit were also included (R&D Systems QC20, QC61, & QC213).
495 LBP was measured by standard ELISA using Hycult Biotech kit HK315-02. Markers for IL-6, IL-
496 10, TNF-α, MCP-1, and IL-22 were measured using Meso Scale Discovery's U-PLEX Biomarker
497 Group 1 multiplex kit K15067L-1. Markers for SAA, VCAM-1, ICAM-1, and CRP were measured
498 using Meso Scale Discovery's V-Plex Plus Vascular Injury Panel 2 multiplex kit K15198G-1.
499 Vascular Injury Control Pack 1 C4198-1 was utilized as a positive control for this assay. Markers
500 for GM-CSF, IL-7, IL-12/23p40, IL-15, IL-16, IL-17A, TNF-β, and VEGF were measured using
501 Meso Scale Discovery's V-Plex Plus Cytokine Panel 1 multiplex kit K151A0H-1. Cytokine Panel
502 1 Control Pack C4050-1 was utilized as a positive control for this assay. Plasma samples were
503 diluted per manufacturer's recommendation for all assays. Standard ELISA kit plates were
504 measured using a Vmax® Kinetic Microplate Reader with Softmax® Pro Software from Molecular
505 Devices LLC. Multiplex ELISA kits from Meso Scale Discovery were measured using the
506 QuickPlex SQ 120 with Discovery Workbench 4.0 software.

507

508 Gnotobiotic Mouse Protocols

509 Germ-free C57/BL6 mice were purchased from Taconic and bred and maintained in flexible film
510 isolator bubbles, fed with standard mouse chow. Three days before they were gavaged, male
511 mice between 5-7 weeks of age were switched over to a western high-fat diet and were fed this
512 diet for the remainder of the experiment. Diets were all obtained from Envigo (Indiana): Standard

513 chow - Teklad global soy protein-free extruded (item 2920X -
514 https://www.envigo.com/resources/data-sheets/2020x-datasheet-0915.pdf), Western Diet – New
515 Total Western Diet (item TD.110919). See Supplemental Table 10 for detailed diet composition.
516 Mice were gavaged with 200 µL of fecal solutions prepared from 1.5 g of donor feces mixed in 3
517 mL of anaerobic PBS (19). Mice were housed individually following gavage for three weeks in a
518 Tecniplast iso-positive caging system, with each cage having HEPA filters and positive
519 pressurization for bioexclusion. Feces were collected from mice at day 21 for 16S rRNA gene
520 sequencing. Mice were euthanized at 21 days post gavage using isoflurane overdose and all
521 efforts were made to minimize suffering. Blood from euthanized animals was collected using
522 cardiac puncture and cells were pelleted in K2-EDTA tubes; plasma was then aliquoted and
523 stored at -80° C.
524
525 Metabolomics Methods
526 *Plasma Sample Preparation*
527 A modified liquid-liquid extraction protocol was used to extract hydrophobic and hydrophilic
528 compounds from the plasma samples (73). Briefly, 50 µL of plasma spiked with internal standards
529 underwent a protein crash with 250 µL ice cold methanol. 750 µL methyl tert-butyl ether (MTBE)
530 and 650 µL 25% methanol in water were added to extract the hydrophobic and hydrophilic
531 compounds, respectively. 500 µL of the upper hydrophobic layer and 400 µL of the lower
532 hydrophilic layer were transferred to separate autosampler vials and dried under nitrogen. The
533 hydrophobic layer was reconstituted with 100 µL of methanol and the hydrophilic layer was
534 reconstituted with 50 µL 5% acetonitrile in water. Both fractions were stored at -80 °C until LC/MS
535 analysis.
536
537 *Liquid Chromatography Mass Spectrometry*

538 The hydrophobic fractions were analyzed using reverse phase chromatography on an Agilent
539 Technologies (Santa Clara, CA) 1290 ultra-high precision liquid chromatography (UHPLC)
540 system on an Agilent Zorbax Rapid Resolution HD SB-C18, 1.8um (2.1 x 100mm) analytical
541 column as previously described (73, 74). The hydrophilic fractions were analyzed using
542 hydrophilic interaction liquid chromatography (HILIC) on a 1290 UHPLC system using an Agilent
543 InfinityLab Poroshell 120 HILIC-Z (2.1 x 100mm) analytical column with gradient conditions as
544 previously described (75) with mass spectrometry modifications as follows: nebulizer pressure:
545 35psi, gas flow: 12L/min, sheath gas temperature: 275C, sheath gas flow: 12L/min, nozzle
546 voltage: 250V, Fragmentor: 100V. The hydrophobic and hydrophilic fractions were run on Agilent
547 Technologies (Santa Clara, CA) 6545 Quadrupole Time of Flight (QTOF) mass spectrometer.
548 Both fractions were run in positive electrospray ionization (ESI) mode
549
550 *Mass Spectrometry Data Processing*
551 Compound data was extracted using Agilent Technologies (Santa Clara, CA) MassHunter
552 Profinder Version 10 software in combination with Agilent Technologies Mass Profiler
553 Professional Version 14.9 (MPP) as described previously (40). Briefly, Batch Molecular Feature
554 Extraction (BMFE) was used in Profinder to extract compound data from all samples and sample
555 preparation blanks. The following BMFE parameters were used to group individual molecular
556 features into compounds: charge state 1-2, with +H, +Na, +NH4 and/or +K charge carriers. To
557 reduce the presence of missing values, a theoretical mass and retention time database was
558 generated for compounds present in samples only from a compound exchange format (.cef) file.
559 This .cef file was then used to re-mine the raw sample data in Profinder using Batch Targeted
560 Feature Extraction.
561
562 An in-house database containing KEGG, METLIN, Lipid Maps, and HMDB spectral data was used
563 to putatively annotate metabolites based on accurate mass (≤ 10 ppm), isotope ratios and isotopic

564 distribution. This corresponds to a Metabolomics Standards Initiative metabolite identification
565 level three (76). To improve compound identification, statistically significant compounds
566 underwent tandem MS using 10, 20, and 40V. Fragmentation patterns of identified compounds
567 were matched to either NIST14 and NIST17 MSMS libraries, or to the *in silico* libraries, MetFrag
568 (77) and Lipid Annotator 1.0 (Agilent) (78).
569
570 *Microbiome-associated metabolites*
571 Microbiome-associated metabolites were defined using metabolites identified as significantly
572 different in abundance between germ-free compared to humanized gnotobiotic mice and/or
573 metabolites identified as microbially produced by the tool AMON (40).
574
575 For the gnotobiotic mouse analysis aqueous and lipid metabolites were analyzed separately (see
576 mouse protocol above for details on experimental set-up). Metabolites that were present in <20%
577 of samples were filtered out before analysis. Significant difference was determined using a
578 Student's t-test with FDR p-value correction. FDR-corrected p values < 0.05 were deemed
579 significant. Significant metabolites also present in the human samples were retained for further
580 analysis.
581
582 For the AMON-identified metabolites, the tool used an inferred metagenome, which was
583 calculated using the PICRUSt2 QIIME2 plugin (42) and default parameters; a list of all identified
584 KEGG IDs from the metabolite data (see metabolome methods); and KEGG flat files (downloaded
585 2019/06/10). AMON determined metabolites observed that could be produced by the given
586 genome. These metabolites were kept for analysis in addition to the gnotobiotic mouse identified
587 metabolites. Those without any putative classification were removed from analysis.
588
589 <u>Microbiome Methods</u>

590 *Sample Collection, Extraction, and Sequencing*

591 Stool samples were collected by the patient within 24 hours prior to their clinic visit on sterile
592 swabs dipped into a full fecal sample deposited into a commode specimen collector. Samples
593 were kept cold or frozen at -20°C during transport prior to being stored at -80°C. DNA was
594 extracted using the standard DNeasy PowerSoil Kit protocol (Qiagen). Extracted DNA was PCR
595 amplified with barcoded primers targeting the V4 region of 16S rRNA gene according to the Earth
596 Microbiome Project 16S Illumina Amplicon protocol with the 515F:806R primer constructs (79).
597 Control sterile swab samples that had undergone the same DNA extraction and PCR amplification
598 procedures were also processed. Each PCR product was quantified using PicoGreen (Invitrogen),
599 and equal amounts (ng) of DNA from each sample were pooled and cleaned using the UltraClean
600 PCR Clean-Up Kit (MoBio). Sequences were generated on six runs on a MiSeq sequencing
601 platform (Illumina, San Diego, CA).

602

603 *Microbiome Sequence Processing and Analysis*

604 Microbiome processing was performed using QIIME2 version 2018.8.0 (80). Data was sequenced
605 across five sequencing runs. Each run was demultiplex and denoised separately using the
606 DADA2 q2 plugin (81). Individual runs were then merged together and 99% *de novo* OTUs were
607 defined using vSEARCH (82). Features were classified using the skLearn classifier in QIIME2
608 with a classifier that was pre-trained on GreenGenes13_8 (83). The phylogenetic tree was
609 building using the SEPP plugin (84). Features that did not classify at the phylum level or were
610 classified as mitochondria or chloroplast were filtered from the analysis. Samples were rarefied
611 at 19,986 reads. To reduce the number of comparisons within the microbiome, we binned highly
612 co-correlating groups of measures within the data types into modules (Supplemental Table 1).
613 These modules were defined using the tool, SCNIC (72). For statistical analysis features present
614 in <20% of samples were filtered out.

615

616 Bioinformatics

617 *Module definition*

618 Modules were called on microbiome and diet data. Modules were defined using the tool SCNIC
619 (72). The q2-SCNIC plugin was used with default parameters for the microbiome data and
620 standalone SCNIC was used for the diet data (https://github.com/shafferm/SCNIC). Specifically,
621 for each data type SCNIC was used to first identify pairwise correlations between all features.
622 Pearson correlation was used for diet and SparCC (85), which takes into account
623 compositionality, was used for microbiome data to. Modules were then selected with a shared
624 minimum distance (SMD) algorithm. The SMD method defines modules by first applying complete
625 linkage hierarchical clustering to correlation coefficients to make a tree of features. Modules are
626 defined as subtrees where all pairwise correlations between all pairs of tips have an R value
627 greater than defined minimum. The diet modules were defined using a Pearson r^2 cutoff of 0.75.
628 The microbiome modules were defined using a SparCC minimum r cutoff of 0.35. To summarize
629 modules SCNIC uses a simple summation of count data from all features in a module. Application
630 of SCNIC reduced the number of evaluated features from 6,913 to 6,818 for microbiome and 59
631 to 29 for diet data.

632

633 *Statistical Analysis*

634 All statistics were performed in R. For non-parametric tests Spearman rank correlation and
635 Kruskal-Wallis test were used. For parametric tests linear models and Student's t-test were used.

636

637 *Data analysis tools*

638 Metabolic disease score was calculated using PCA in R with prcomp. Data was scaled using
639 default method within the prcomp library. All random forest analysis tools were used in R.
640 Standard random forest was performed using randomForest. Variable selection was performed
641 in R using the tool VSURF (35). Interaction analysis was performed in R using the tool iRF (39).

642

643 Data Availability

644 All data will be publicly available upon publishing. Microbiome data in QIITA
645 (https://qiita.ucsd.edu) Study ID 13338 and available upon request and will be publicly available
646 in EBI/ENA (https://www.ebi.ac.uk/ena) upon publishing. Immune and diet data are available
647 along with the microbiome data as associated metadata. Metabolomics data will be available on
648 Metabolomic Workbench (https://www.metabolomicsworkbench.org) upon publishing. Until
649 publicly available it is available upon request.

650

651 **Statements**

652 Acknowledgements

653 We would like to thank our study participants for contributing their samples and time to this study.
654 We would also like to thank Christine Griesmer for her contributions to subject recruitment and
655 Brandi Wagner and Maggie Stanislawski for insights into statistical analysis methods.

656

657 Funding sources

658 This work was funded by R01 DK104047 and R01 DK108366 with additional support from
659 NIH/NCATS Colorado CTSA Grant Number UL1 TR002535. High performance computing was
660 supported by a cluster at the University of Colorado Boulder funded by National Institutes of
661 Health 1S10OD012300. AJS Armstrong was supported by T32 AI052066. Contents are the
662 authors' sole responsibility and do not necessarily represent official NIH views.

663

664 Declarations of interests

665 The authors declare that they have no conflicts of interest.

666

667 Author contributions

AJSA analyzed and interpreted all data. AJSA and CAL wrote the manuscript. NR guided generation and interpretation of metabolomics data. KQ and KAD prepared, ran, and processed metabolomics. KQ ran metabolic pathway analysis. SXL prepared and conducted mouse experiments. JMS ran immunological assays. NMN prepared and ran sequencing and coordinated fecal sample and metadata collection from study subjects. SF recruited subjects, collected samples, and maintained regulatory compliance. TJM and JH collected and aided in interpretation and processing of diet data. CAL, BEP, and TC conceptualized and led the study.TC guided all clinical data collection and subject recruitment and provided clinical insight into study populations. BEP guided generation and interpretation of immune data. CAL guided microbiome data generation and multi'omic data analysis. All authors read and approved the final manuscript.

678 **Supplemental Tables**

679 **Supplemental Table 1.** Microbiome modules calculated by SCNIC

680

681 **Supplemental Table 2.** Study measures by datatype

682

683 **Supplemental Table 3.** Diet survey data modules calculated by SCNIC

684

685 **Supplemental Table 4.** VSURF-selected features and correlation with metabolic disease score

686

687 **Supplemental Table 5.** Edge table for VSURF-selected inter-variable correlations

688

689 **Supplemental Table 6.** Gnotobiotic mouse experiment set-up

690

691 **Supplemental Table 7.** Microbiome-associated metabolites list and source of identification

692

693 **Supplemental Table 8.** Metabolites correlating with metabolic score

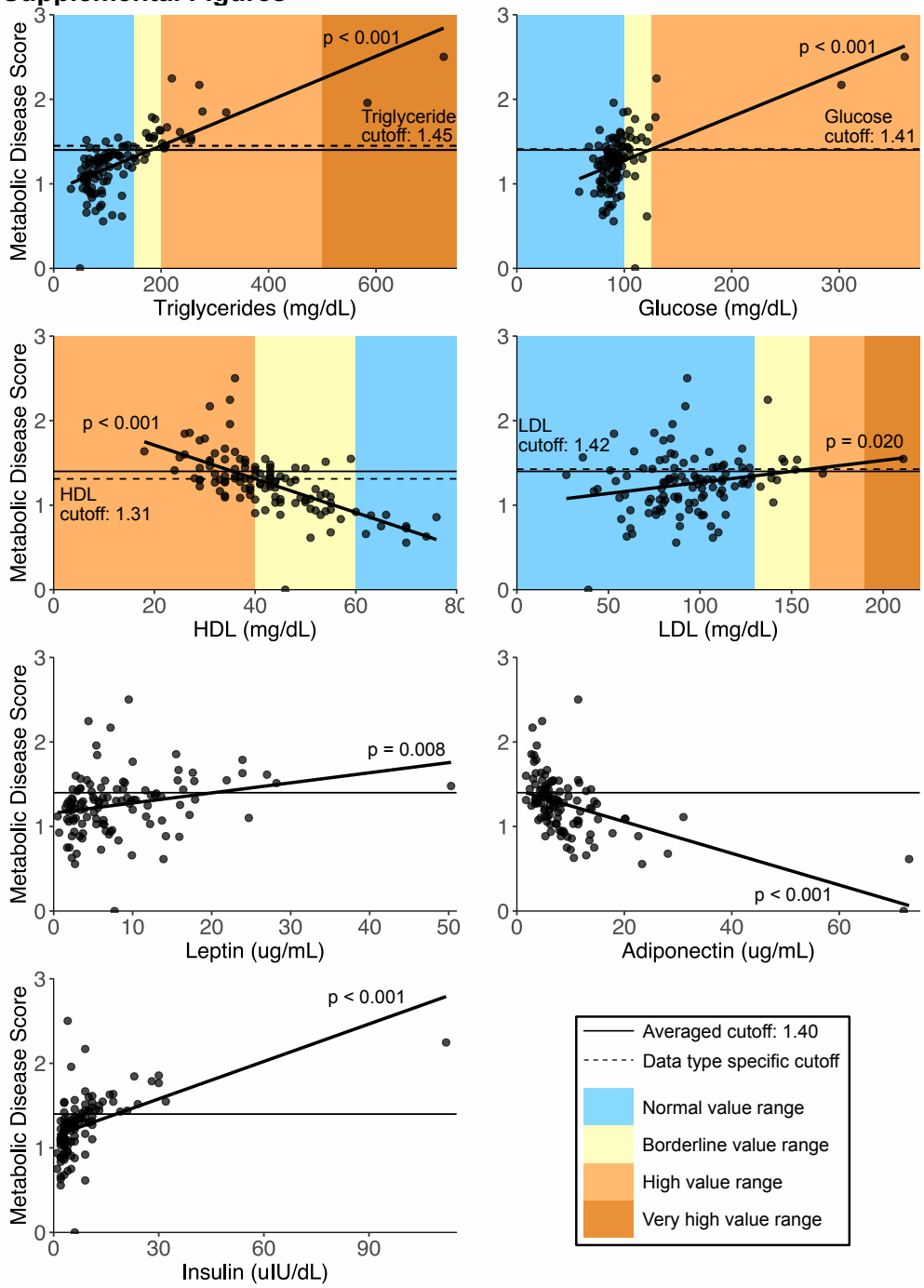
694

695 **Supplemental Table 9.** mBrole pathway analysis results

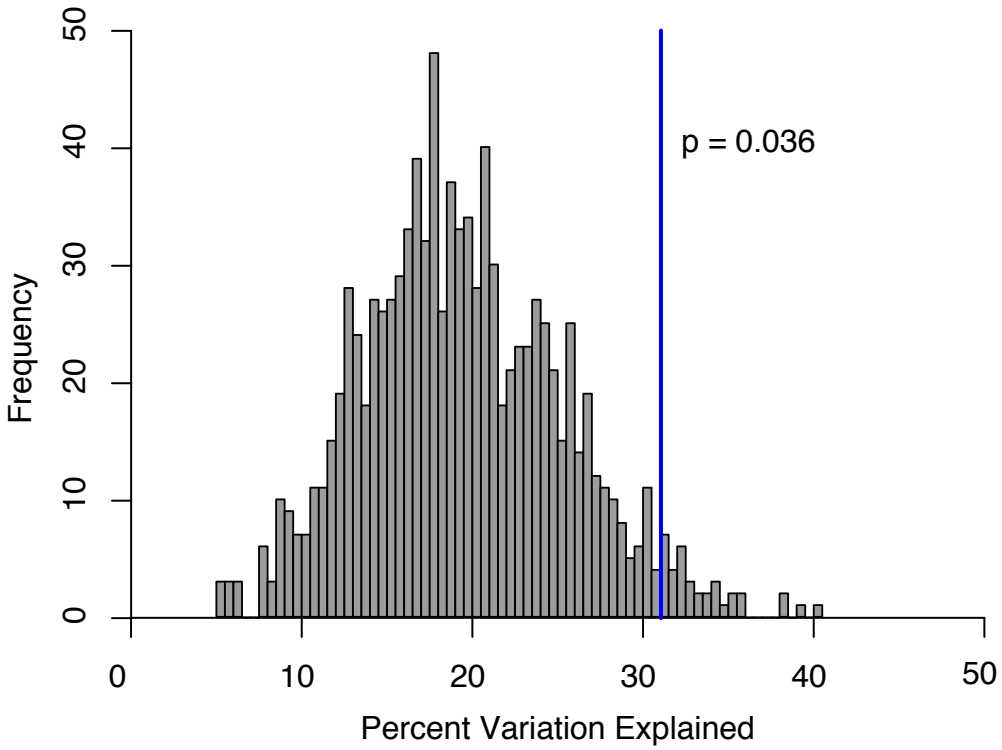
696

697 **Supplemental Table 10.** Western diet for gnotobiotic mice

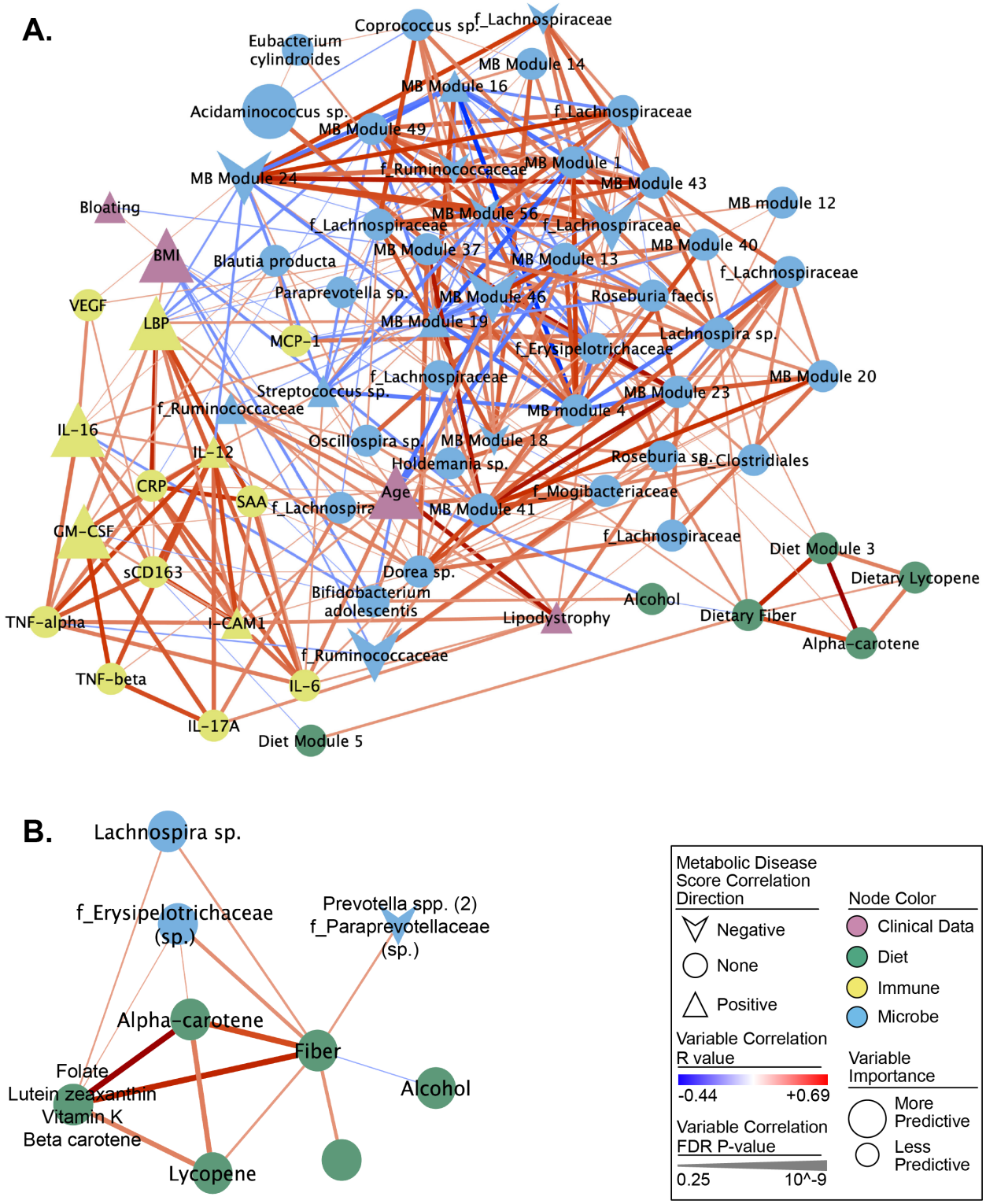
Supplemental Figures



Supplemental Figure 1. Metabolic Score Cutoff Calculations using regressions between metabolic disease score and the metrics used in the PCA analysis. Triglycerides, fasting glucose, HDL and LDL have well-define clinical cut-offs for high values and were used to calculate the healthy-unhealth cutoff. Linear model was calculated modeling metabolic disease score by each marker. The high value intercept of the regression line is marked with a dotted line and value annotated on the plot. The solid line is the defined healthy-unhealthy cutoff calculated as the mean of the four cutoff values. P values are from the linear model.



Supplemental Figure 2. Histogram of percent variation explained in permuted VSURF. Metabolic disease score was permuted 1,000 times and passed through VSURF. The resulting variables were run through a standard random forest and the percent variation explained was calculated. The blue line represents the percent variation explained for the true VSURF. P value was calculated using a one tailed test.



Supplemental Figure 3. Correlation network of VSURF-selected variables. Correlation network of **A)** all VSURF-selected variables and **B)** neighboring nodes of dietary fiber. All Spearman rank correlations with an FDR $p < 0.25$ are shown. See Supplemental Table 5 for the edge table and Supplemental Table 4 for the node table.

References

1. Non LR, Escota GV, Powderly WG. 2016. HIV and its relationship to insulin resistance and lipid abnormalities. Translational Research 183:41-56.
2. Nix LM, Tien PC. 2014. Metabolic syndrome, diabetes, and cardiovascular risk in HIV. Curr HIV/AIDS Rep 11:271-8.
3. Lake JE, Li X, Palella FJ, Jr., Erlandson KM, Wiley D, Kingsley L, Jacobson LP, Brown TT. 2018. Metabolic health across the BMI spectrum in HIV-infected and HIV-uninfected men. AIDS 32:49-57.
4. Willig AL, Overton ET. 2016. Metabolic Complications and Glucose Metabolism in HIV Infection: A Review of the Evidence. Curr HIV/AIDS Rep 13:289-96.
5. Monczor AN, Li X, Palella FJ, Jr., Erlandson KM, Wiley D, Kingsley LA, Post WS, Jacobson LP, Brown TT, Lake JE. 2018. Systemic Inflammation Characterizes Lack of Metabolic Health in Nonobese HIV-Infected Men. Mediators Inflamm 2018:5327361.
6. Funderburg NT, Mehta NN. 2016. Lipid Abnormalities and Inflammation in HIV Inflection. Curr HIV/AIDS Rep 13:218-25.
7. Beltran LM, Rubio-Navarro A, Amaro-Villalobos JM, Egido J, Garcia-Puig J, Moreno JA. 2015. Influence of immune activation and inflammatory response on cardiovascular risk associated with the human immunodeficiency virus. Vasc Health Risk Manag 11:35-48.
8. Tsiodras S, Mantzoros C, Hammer S, Samore M. 2000. Effects of Protease Inhibitors on Hyperglycemia, Hyperlipidemia, and Lipodystrophy A 5-Year Cohort Study. Arch Intern Med 160:2050-2056.
9. Galescu O, Bhangoo A, Ten S. 2013. Insulin resistance, lipodystrophy and cardiometabolic syndrome in HIV/AIDS. Reviews in Endocrine and Metabolic Disorders 14:133-140.
10. Tilg H, Kaser A. 2011. Gut microbiome, obesity, and metabolic dysfunction. The Journal of Clinical Investigation 121:2126-2132.
11. Sonnenburg JL, Bäckhed F. 2016. Diet–microbiota interactions as moderators of human metabolism. Nature 535:56-64.
12. Pedersen HK, Gudmundsdottir V, Nielsen HB, Hyotylainen T, Nielsen T, Jensen BA, Forslund K, Hildebrand F, Prifti E, Falony G, Le Chatelier E, Levenez F, Dore J, Mattila I, Plichta DR, Poho P, Hellgren LI, Arumugam M, Sunagawa S, Vieira-Silva S, Jorgensen T, Holm JB, Trost K, Meta HITC, Kristiansen K, Brix S, Raes J, Wang J, Hansen T, Bork P, Brunak S, Oresic M, Ehrlich SD, Pedersen O. 2016. Human gut microbes impact host serum metabolome and insulin sensitivity. Nature 535:376-81.
13. Matey-Hernandez ML, Williams FMK, Potter T, Valdes AM, Spector TD, Menni C. 2018. Genetic and microbiome influence on lipid metabolism and dyslipidemia. Physiol Genomics 50:117-126.
14. Karlsson F, Tremaroli V, Nielsen J, Bäckhed F. 2013. Assessing the human gut microbiota in metabolic diseases. Diabetes 62:3341-9.
15. Neff CP, Krueger O, Xiong K, Arif S, Nusbacher N, Schneider JM, Cunningham AW, Armstrong A, Li S, McCarter MD, Campbell TB, Lozupone CA, Palmer BE. 2018. Fecal Microbiota Composition Drives Immune Activation in HIV-infected Individuals. EBioMedicine 30:192-202.
16. Armstrong AJS, Shaffer M, Nusbacher NM, Griesmer C, Fiorillo S, Schneider JM, Preston Neff C, Li SX, Fontenot AP, Campbell T, Palmer BE, Lozupone CA. 2018. An exploration of Prevotella-rich microbiomes in HIV and men who have sex with men. Microbiome 6:198.
17. Vujkovic-Cvijin I, Somsouk M. 2019. HIV and the Gut Microbiota: Composition, Consequences, and Avenues for Amelioration. Curr HIV/AIDS Rep doi:10.1007/s11904-019-00441-w.

18. Noguera-Julian M, Rocafort M, Guillén Y, Rivera J, Casadellà M, Nowak P, Hildebrand F, Zeller G, Parera M, Bellido R, Rodríguez C, Carrillo J, Mothe B, Coll J, Bravo I, Estany C, Herrero C, Saz J, Sirera G, Torrela A, Navarro J, Crespo M, Brander C, Negredo E, Blanco J, Guarner F, Calle ML, Bork P, Sönnerborg A, Clotet B, Paredes R. 2016. Gut Microbiota Linked to Sexual Preference and HIV Infection. EBioMedicine 5:135-46.
19. Li SX, Sen S, Schneider JM, Xiong KN, Nusbacher NM, Moreno-Huizar N, Shaffer M, Armstrong AJS, Severs E, Kuhn K, Neff CP, McCarter M, Campbell T, Lozupone CA, Palmer BE. 2019. Gut microbiota from high-risk men who have sex with men drive immune activation in gnotobiotic mice and in vitro HIV infection. PLoS Pathog 15:e1007611.
20. Cui HL, Ditiatkovski M, Kesani R, Bobryshev YV, Liu Y, Geyer M, Mukhamedova N, Bukrinsky M, Sviridov D. 2014. HIV protein Nef causes dyslipidemia and formation of foam cells in mouse models of atherosclerosis. FASEB Journal 28:2828-2839.
21. Auclair M, Guenantin AC, Fellahi S, Garcia M, Capeau J. 2020. HIV antiretroviral drugs, dolutegravir, maraviroc and ritonavir-boosted atazanavir use different pathways to affect inflammation, senescence and insulin sensitivity in human coronary endothelial cells. PLoS One 15:e0226924.
22. Ahmed D, Roy D, Cassol E. 2018. Examining Relationships between Metabolism and Persistent Inflammation in HIV Patients on Antiretroviral Therapy. Mediators Inflamm 2018:6238978.
23. Reid M, Ma Y, Scherzer R, Price JC, French AL, Plankey MW, Grunfeld C, Tien PC. 2017. Higher CD163 levels are associated with insulin resistance in hepatitis C virus-infected and HIV-infected adults. AIDS 31:385-393.
24. Somsouk M, Estes JD, Deleage C, Dunham RM, Albright R, Inadomi JM, Martin JN, Deeks SG, McCune JM, Hunt PW. 2015. Gut epithelial barrier and systemic inflammation during chronic HIV infection. AIDS (London, England) 29:43-51.
25. Klase Z, Ortiz A, Deleage C, Mudd JC, Quiñones M, Schwartzman E, Klatt NR, Canary L, Estes JD, Brenchley JM. 2015. Dysbiotic bacteria translocate in progressive SIV infection. Mucosal Immunology 8:1009-1020.
26. Palmer CD, Tomassilli J, Sirignano M, Romero-Tejeda M, Arnold KB, Che D, Lauffenburger DA, Jost S, Allen T, Mayer KH, Altfeld M. 2014. Enhanced immune activation linked to endotoxemia in HIV-1 seronegative MSM. AIDS 28:2162-6.
27. Lim PS, Chang YK, Wu TK. 2019. Serum Lipopolysaccharide-Binding Protein is Associated with Chronic Inflammation and Metabolic Syndrome in Hemodialysis Patients. Blood Purif 47:28-36.
28. Hersoug LG, Moller P, Loft S. 2018. Role of microbiota-derived lipopolysaccharide in adipose tissue inflammation, adipocyte size and pyroptosis during obesity. Nutr Res Rev 31:153-163.
29. Awoyemi A, Troseid M, Arnesen H, Solheim S, Seljeflot I. 2018. Markers of metabolic endotoxemia as related to metabolic syndrome in an elderly male population at high cardiovascular risk: a cross-sectional study. Diabetol Metab Syndr 10:59.
30. Liu X, Lu L, Yao P, Ma Y, Wang F, Jin Q, Ye X, Li H, Hu FB, Sun L, Lin X. 2014. Lipopolysaccharide binding protein, obesity status and incidence of metabolic syndrome: a prospective study among middle-aged and older Chinese. Diabetologia 57:1834-41.
31. Gelpi M, Vestad B, Hansen SH, Holm K, Drivsholm N, Goetz A, Kirkby NS, Lindegaard B, Lebech AM, Hoel H, Michelsen AE, Ueland T, Gerstoft J, Lundgren J, Hov JR, Nielsen SD, Troseid M. 2020. Impact of Hiv-Related Gut Microbiota Alterations on Metabolic Comorbidities. Clin Infect Dis doi:10.1093/cid/ciz1235.
32. Hammer SM, Sobieszczyk ME, Janes H, Karuna ST, Mulligan MJ, Grove D, Koblin BA, Buchbinder SP, Keefer MC, Tomaras GD, Frahm N, Hural J, Anude C, Graham BS,

820 Enama ME, Adams E, DeJesus E, Novak RM, Frank I, Bentley C, Ramirez S, Fu R,
821 Koup RA, Mascola JR, Nabel GJ, Montefiori DC, Kublin J, McElrath MJ, Corey L, Gilbert
822 PB, Team HS. 2013. Efficacy trial of a DNA/rAd5 HIV-1 preventive vaccine. N Engl J
823 Med 369:2083-92.
824 33. Wiley JF, Carrington MJ. 2016. A metabolic syndrome severity score: A tool to quantify
825 cardio-metabolic risk factors. Prev Med 88:189-95.
826 34. Hillier TA, Rousseau A, Lange C, Lepinay P, Cailleau M, Novak M, Calliez E,
827 Ducimetiere P, Balkau B, Cohort DESIR. 2006. Practical way to assess metabolic
828 syndrome using a continuous score obtained from principal components analysis.
829 Diabetologia 49:1528-35.
830 35. Genuer R, Poggi J-M, Tuleau-Malot C. 2015. VSURF: An R Package for Variable
831 Selection Using Random Forests. The R Journal, R Foundation for Statistical Computing
832 7:12-33.
833 36. Marchetti G, Tincati C, Silvestri G. 2013. Microbial translocation in the pathogenesis of
834 HIV infection and AIDS. Clinical microbiology reviews 26:2-18.
835 37. Duncan SH, Holtrop G, Lobley GE, Calder AG, Stewart CS, Flint HJ. 2004. Contribution
836 of acetate to butyrate formation by human faecal bacteria. Br J Nutr 91:915-23.
837 38. Duncan SH, Barcenilla A, Stewart CS, Pryde SE, Flint HJ. 2002. Acetate utilization and
838 butyryl coenzyme A (CoA):acetate-CoA transferase in butyrate-producing bacteria from
839 the human large intestine. Appl Environ Microbiol 68:5186-90.
840 39. Basu S, Kumbier K, Brown JB, Yu B. 2018. Iterative random forests to discover
841 predictive and stable high-order interactions. Proc Natl Acad Sci U S A 115:1943-1948.
842 40. Shaffer M, Thurimella K, Quinn K, Doenges K, Zhang X, Bokatzian S, Reisdorph N,
843 Lozupone CA. 2019. AMON: annotation of metabolite origins via networks to integrate
844 microbiome and metabolome data. BMC Bioinformatics 20:614.
845 41. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. 2017. KEGG: new
846 perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Research
847 45:D353-D361.
848 42. Douglas GM, Maffei VJ, Zaneveld J, Yurgel SN, Brown JR, Taylor CM, Huttenhower C,
849 Langille MGI. 2020. PICRUSt2: An improved and extensible approach for metagenome
850 inference. doi:10.1101/672295.
851 43. Frolkis A, Knox C, Lim E, Jewison T, Law V, Hau DD, Liu P, Gautam B, Ly S, Guo AC,
852 Xia J, Liang Y, Shrivastava S, Wishart DS. 2010. SMPDB: The Small Molecule Pathway
853 Database. Nucleic Acids Res 38:D480-7.
854 44. van der Veen JN, Kennelly JP, Wan S, Vance JE, Vance DE, Jacobs RL. 2017. The
855 critical role of phosphatidylcholine and phosphatidylethanolamine metabolism in health
856 and disease. Biochim Biophys Acta Biomembr 1859:1558-1572.
857 45. Verhaegen P, Borchelt M, Smith J. 2003. Relation between cardiovascular and
858 metabolic disease and cognition in very old age: cross-sectional and longitudinal findings
859 from the berlin aging study. Health Psychol 22:559-69.
860 46. Montessori V, Press N, Harris M, Akagi L, Montaner JSG. 2004. Adverse effects of
861 antiretroviral therapy for HIV infection. CMAJ : Canadian Medical Association journal =
862 journal de l'Association medicale canadienne 170:229-38.
863 47. Alexander CM, Landsman PB, Teutsch SM, Haffner SM, Third National H, Nutrition
864 Examination S, National Cholesterol Education P. 2003. NCEP-defined metabolic
865 syndrome, diabetes, and prevalence of coronary heart disease among NHANES III
866 participants age 50 years and older. Diabetes 52:1210-4.
867 48. Meigs JB. 2003. Epidemiology of the insulin resistance syndrome. Curr Diab Rep 3:73-9.
868 49. Kerchberger AM, Sheth AN, Angert CD, Mehta CC, Summers NA, Ofotokun I, Gustafson
869 D, Weiser SD, Sharma A, Adimora AA, French AL, Augenbraun M, Cocohoba J,
870 Kassaye S, Bolivar H, Govindarajulu U, Konkle-Parker D, Golub ET, Lahiri CD. 2019.

871 Weight Gain Associated with Integrase Stand Transfer Inhibitor Use in Women. Clin
872 Infect Dis doi:10.1093/cid/ciz853.
873 50. Venter WDF, Moorhouse M, Sokhela S, Fairlie L, Mashabane N, Masenya M, Serenata
874 C, Akpomiemie G, Qavi A, Chandiwana N, Norris S, Chersich M, Clayden P, Abrams E,
875 Arulappan N, Vos A, McCann K, Simmons B, Hill A. 2019. Dolutegravir plus Two
876 Different Prodrugs of Tenofovir to Treat HIV. N Engl J Med 381:803-815.
877 51. Pérez-Matute P, Pérez-Martínez L, Aguilera-Lizarraga J, Blanco JR, Oteo JA. 2015.
878 Maraviroc modifies gut microbiota composition in a mouse model of obesity: a plausible
879 therapeutic option to prevent metabolic disorders in HIV-infected patients. Revista
880 española de quimioterapia : publicación oficial de la Sociedad Española de
881 Quimioterapia 28:200-6.
882 52. Han GM, Meza JL, Soliman GA, Islam KM, Watanabe-Galloway S. 2016. Higher levels
883 of serum lycopene are associated with reduced mortality in individuals with metabolic
884 syndrome. Nutr Res 36:402-7.
885 53. Zeng YC, Peng LS, Zou L, Huang SF, Xie Y, Mu GP, Zeng XH, Zhou XL, Zeng YC.
886 2017. Protective effect and mechanism of lycopene on endothelial progenitor cells
887 (EPCs) from type 2 diabetes mellitus rats. Biomed Pharmacother 92:86-94.
888 54. Sluijs I, Beulens JW, Grobbee DE, van der Schouw YT. 2009. Dietary carotenoid intake
889 is associated with lower prevalence of metabolic syndrome in middle-aged and elderly
890 men. J Nutr 139:987-92.
891 55. Kovatcheva-Datchary P, Nilsson A, Akrami R, Lee Ying S, De Vadder F, Arora T, Hallen
892 A, Martens E, Björck I, Bäckhed F. 2015. Dietary Fiber-Induced Improvement in Glucose
893 Metabolism Is Associated with Increased Abundance of Prevotella. Cell Metabolism
894 22:971-982.
895 56. O'Keefe SJD, Li JV, Lahti L, Ou J, Carbonero F, Mohammed K, Posma JM, Kinross J,
896 Wahl E, Ruder E, Vipperla K, Naidoo V, Mtshali L, Tims S, Puylaert PGB, DeLany J,
897 Krasinskas A, Benefiel AC, Kaseb HO, Newton K, Nicholson JK, de Vos WM, Gaskins
898 HR, Zoetendal EG. 2015. Fat, fibre and cancer risk in African Americans and rural
899 Africans. Nature communications 6:6342.
900 57. Schröder NWJ, Schumann RR. 2016. Non-LPS targets and actions of LPS binding
901 protein (LBP). Journal of Endotoxin Research 11:237-242.
902 58. Boulangé CL, Neves AL, Chilloux J, Nicholson JK, Dumas M-E. 2016. Impact of the gut
903 microbiota on inflammation, obesity, and metabolic disease. Genome Medicine 8:42.
904 59. Brake DK, Smith EO, Mersmann H, Smith CW, Robker RL. 2006. ICAM-1 expression in
905 adipose tissue: effects of diet-induced obesity in mice. Am J Physiol Cell Physiol
906 291:C1232-9.
907 60. Hsu LA, Chang CJ, Wu S, Teng MS, Chou HH, Chang HH, Chang PY, Ko YL. 2010.
908 Association between functional variants of the ICAM1 and CRP genes and metabolic
909 syndrome in Taiwanese subjects. Metabolism 59:1710-6.
910 61. Senn JJ, Klover PJ, Nowak IA, Mooney RA. 2002. Interleukin-6 induces cellular insulin
911 resistance in hepatocytes. Diabetes 51:3391-9.
912 62. Reigstad CS, Lunden GO, Felin J, Backhed F. 2009. Regulation of serum amyloid A3
913 (SAA3) in mouse colonic epithelium and adipose tissue by the intestinal microbiota.
914 PLoS One 4:e5842.
915 63. Cani PD, Bibiloni R, Knauf C, Waget A, Neyrinck AM, Delzenne NM, Burcelin R. 2008.
916 Changes in gut microbiota control metabolic endotoxemia-induced inflammation in high-
917 fat diet-induced obesity and diabetes in mice. Diabetes 57:1470-81.
918 64. Tilves CM, Zmuda JM, Kuipers AL, Nestlerode CS, Evans RW, Bunker CH, Patrick AL,
919 Miljkovic I. 2016. Association of Lipopolysaccharide-Binding Protein With Aging-Related
920 Adiposity Change and Prediabetes Among African Ancestry Men. Diabetes Care 39:385-
921 91.

65. Moreau RA, Agnew J, Hicks KB, Powell MJ. 1997. Modulation of lipoxygenase activity by bacterial hopanoids. J Nat Prod 60:397-8.
66. Tsai IJ, Croft KD, Mori TA, Falck JR, Beilin LJ, Puddey IB, Barden AE. 2009. 20-HETE and F2-isoprostanes in the metabolic syndrome: the effect of weight reduction. Free Radic Biol Med 46:263-70.
67. Raza GS, Putaala H, Hibberd AA, Alhoniemi E, Tiihonen K, Makela KA, Herzig KH. 2017. Polydextrose changes the gut microbiome and attenuates fasting triglyceride and cholesterol levels in Western diet fed mice. Sci Rep 7:5294.
68. Rabot S, Membrez M, Bruneau A, Gérard P, Harach T, Moser M, Raymond F, Mansourian R, Chou CJ. 2010. Germ-free C57BL/6J mice are resistant to high-fat-diet-induced insulin resistance and have altered cholesterol metabolism. FASEB journal : official publication of the Federation of American Societies for Experimental Biology 24:4948-59.
69. Velagapudi VR, Hezaveh R, Reigstad CS, Gopalacharyulu P, Yetukuri L, Islam S, Felin J, Perkins R, Boren J, Oresic M, Backhed F. 2010. The gut microbiota modulates host energy and lipid metabolism in mice. J Lipid Res 51:1101-12.
70. National Institues of Health EaGRP, Nactional Cancer Institute. 2010. Diet History Questionnaire, Verson 2.0.
71. National Cancer Institute EaGRP. 2012. Diet*Calc Analysis Program, Version 1.5.0. .
72. Shaffer M. SCNIC: Sparse Cooccurrence Network Investigation for Compositional data. https://github.com/shafferm/SCNIC. Accessed
73. Yang Y, Cruickshank C, Armstrong M, Mahaffey S, Reisdorph R, Reisdorph N. 2013. New sample preparation approach for mass spectrometry-based profiling of plasma results in improved coverage of metabolome. J Chromatogr A 1300:217-26.
74. Halper-Stromberg E, Gillenwater L, Cruickshank-Quinn C, O'Neal WK, Reisdorph N, Petrache I, Zhuang Y, Labaki WW, Curtis JL, Wells J, Rennard S, Pratte KA, Woodruff P, Stringer KA, Kechris K, Bowler RP. 2019. Bronchoalveolar Lavage Fluid from COPD Patients Reveals More Compounds Associated with Disease than Matched Plasma. Metabolites 9.
75. Kennedy A, Bivens A. 2017. Methods for the Analysis of Underivatized Amino Acids by LC/MS: For Food, Life Science, and Metabolomics Applications, Agilent Application Note.
76. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TW, Fiehn O, Goodacre R, Griffin JL, Hankemeier T, Hardy N, Harnly J, Higashi R, Kopka J, Lane AN, Lindon JC, Marriott P, Nicholls AW, Reily MD, Thaden JJ, Viant MR. 2007. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics 3:211-221.
77. Ruttkies C, Schymanski EL, Wolf S, Hollender J, Neumann S. 2016. MetFrag relaunched: incorporating strategies beyond in silico fragmentation. J Cheminform 8:3.
78. Agilent. 2020. Lipidomics Analysis with Lipid Annotator and Mass Profiler Professional, vol 5994-1111EN., Technical Overview.
79. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, Prill RJ, Tripathi A, Gibbons SM, Ackermann G, Navas-Molina JA, Janssen S, Kopylova E, Vazquez-Baeza Y, Gonzalez A, Morton JT, Mirarab S, Zech Xu Z, Jiang L, Haroon MF, Kanbar J, Zhu Q, Jin Song S, Kosciolek T, Bokulich NA, Lefler J, Brislawn CJ, Humphrey G, Owens SM, Hampton-Marcell J, Berg-Lyons D, McKenzie V, Fierer N, Fuhrman JA, Clauset A, Stevens RL, Shade A, Pollard KS, Goodwin KD, Jansson JK, Gilbert JA, Knight R, Earth Microbiome Project C. 2017. A communal catalogue reveals Earth's multiscale microbial diversity. Nature 551:457-463.
80. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ,

973 Brown CT, Callahan BJ, Caraballo-Rodriguez AM, Chase J, Cope EK, Da Silva R,
974 Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M,
975 Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo
976 J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch
977 AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, et al. 2019.
978 Reproducible, interactive, scalable and extensible microbiome data science using QIIME
979 2. Nat Biotechnol 37:852-857.
980 81. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. 2016.
981 DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods
982 13:581-3.
983 82. Rognes T, Flouri T, Nichols B, Quince C, Mahe F. 2016. VSEARCH: a versatile open
984 source tool for metagenomics. PeerJ 4:e2584.
985 83. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL,
986 Knight R, Hugenholtz P. 2012. An improved Greengenes taxonomy with explicit ranks
987 for ecological and evolutionary analyses of bacteria and archaea. The ISME journal
988 6:610-8.
989 84. Janssen S, McDonald D, Gonzalez A, Navas-Molina JA, Jiang L, Xu ZZ, Winker K, Kado
990 DM, Orwoll E, Manary M, Mirarab S, Knight R. 2018. Phylogenetic Placement of Exact
991 Amplicon Sequences Improves Associations with Clinical Information. mSystems 3.
992 85. Friedman J, Alm EJ. 2012. Inferring correlation networks from genomic survey data.
993 PLoS Comput Biol 8:e1002687.

994