ORIGINAL ARTICLE

Human metabolic correlates of body mass index

Steven C. Moore · Charles E. Matthews · Joshua N. Sampson ·
Rachael Z. Stolzenberg-Solomon · Wei Zheng · Qiuyin Cai ·
Yu Ting Tan · Wong-Ho Chow · Bu-Tian Ji · Da Ke Liu ·
Qian Xiao · Simina M. Boca · Michael F. Leitzmann · Gong Yang ·
Yong Bing Xiang · Rashmi Sinha · Xiao Ou Shu · Amanda J. Cross

Abstract A high body mass index (BMI) is a major risk factor for several chronic diseases, but the biology underlying these associations is not well-understood. Dyslipidemia, inflammation, and elevated levels of growth factors and sex steroid hormones explain some of the increased disease risk, but other metabolic factors not yet identified may also play a role. In order to discover novel metabolic biomarkers of BMI, we used non-targeted metabolomics to assay 317 metabolites in blood samples from 947 participants and examined the cross-sectional associations between metabolite levels and BMI. Participants were from three studies in the United States and China. Height, weight, and potential confounders were ascertained by questionnaire (US studies) or direct measurement (Chinese study). Metabolite levels were measured using liquid-phase chromatography and gas chromatography coupled with mass spectrometry. We evaluated study-specific associations using linear regression, adjusted for age, gender, and smoking, and we estimated combined associations using random effects meta-analysis. The meta-analysis revealed 37 metabolites significantly associated with BMI, including 19 lipids, 12 amino acids, and 6 others, at the Bonferroni significance threshold ($P < 0.00016$). Eighteen of these associations had not been previously reported, including histidine, an amino acid neurotransmitter, and butyrylcarnitine, a lipid marker of whole-body fatty acid oxidation. Heterogeneity by study was minimal (all $P_{heterogeneity} > 0.05$). In total, 110 metabolites were associated with BMI at the $P < 0.05$ level. These findings establish a baseline for the BMI metabolome, and suggest new targets for researchers attempting to clarify mechanistic links between BMI and disease risk.

S. C. Moore (✉) · C. E. Matthews · J. N. Sampson ·
R. Z. Stolzenberg-Solomon · B.-T. Ji · Q. Xiao ·
S. M. Boca · R. Sinha · A. J. Cross
Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, 9609 Medical Center Drive, Rockville, MD 20850, USA
e-mail: moorest@mail.nih.gov

W. Zheng · Q. Cai · G. Yang · X. O. Shu
Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Institute for Medicine and Public Health, Vanderbilt University School of Medicine, Nashville, TN, USA

Y. T. Tan · D. K. Liu · Y. B. Xiang
Shanghai Cancer Institute, Shanghai, China

W.-H. Chow
University of Texas MD Anderson Cancer Center, Houston, TX, USA

M. F. Leitzmann
Department of Epidemiology and Preventive Medicine, Regensburg University Medical Center, Franz-Josef-Strauss-Allee 11, 93053 Regensburg, Germany

1 Introduction

Substantial evidence indicates that a high body weight—as measured by body mass index (BMI)—is a major risk factor for several chronic diseases, including cardiovascular disease and cancer (Poirier et al. 2006; Chan et al. 1994; Renehan et al. 2008). The biological mechanisms linking a high BMI to these diseases are not well-understood,

however (Poirier et al. 2006; National Cancer Institute 2013). Elevated levels of cholesterol, triglycerides, inflammation, growth factors (e.g. insulin) and sex steroid hormones explain some of the increased disease risk related to a high BMI, but undiscovered metabolic factors may also play a major role (Poirier et al. 2006; Roberts et al. 2010; Chan et al. 1994; Renehan et al. 2008).

New, high-throughput metabolic profiling technologies, such as metabolomics, could facilitate discovery of BMI-related metabolic factors that are important for disease risk. Investigators have recently used metabolic profiling to identify amino acids associated with an elevated BMI (Newgard et al. 2009; Cheng et al. 2012; Gaudet et al. 2012) and levels of these amino acids were later found to predict future diabetes risk (Wang et al. 2011; Floegel et al. 2012). Overall, however, few studies have examined the metabolic correlates of BMI and they were small (fewer than 150 participants) (Newgard et al. 2009) or assessed fewer than 60 metabolites (Cheng et al. 2012; Gaudet et al. 2012). Many prior studies have, of course, used targeted assays to examine metabolite-BMI associations. But, because these studies examined one or only a few metabolic factors, they could not address whether their findings were simply due to confounding by other BMI associated metabolic factors.

A broader examination of the metabolic correlates of BMI—one that includes both a large sample size and hundreds of metabolites—may identify novel biomarkers related to BMI that could be examined as candidate disease markers in future etiologic studies. To this end, we measured 317 metabolites in blood samples from 947 participants from three study populations from the United States and China, and examined their relation to BMI.

2 Subjects and methods

2.1 Study participants and data collection

Our study included participants from three study populations: The Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial, the Navy Colon Adenoma Study (Navy), and the Shanghai Physical Activity Study (Shanghai). The PLCO study was a multicenter trial that recruited participants between 1993 and 2001 and randomly assigned them to cancer screening or no screening (Prorok et al. 2000). Eligible participants were 55–74 years of age, with no prior diagnosis of prostate, lung, colorectal, or ovarian cancer. Participants in our analysis had been specifically selected from PLCO for a nested case–control study of colorectal cancer. In this study, there were 254 cases and 255 matched controls who completed a self-administered demographics and health habits questionnaire

at baseline, reported their height and weight, and donated a baseline blood sample. Individuals were excluded if they had a rare cancer during follow-up, had self-reported Crohn's disease, ulcerative colitis, familial polyposis, Gardner's syndrome, or colorectal polyps as determined from the baseline questionnaire. All colorectal cancer cases in this nested case–control study were diagnosed at least 6 months after baseline, with the average diagnosis occurring 8 years after baseline. Controls were matched to cases on age, sex, year of randomization, and season of blood draw, and were alive and cancer free at the time that the matched case was diagnosed. Of these participants, height and weight data were available for 254 cases and 251 controls.

The Navy study was a case–control study of colorectal adenomas conducted at the National Naval Medical Center (Bethesda, MD) (Sinha et al. 1999). To be eligible, participants had to reside within 60 miles of the study center, be 18–74 years of age, and have no history of Crohn's disease, ulcerative colitis, or cancer (except non-melanoma skin cancer). Cases were patients diagnosed with colorectal adenoma by sigmoidoscopy (18 %) or colonoscopy (82 %) between April 1994 and September 1996. Controls were patients confirmed to be polyp-free during sigmoidoscopic screening and they were frequency matched to cases by age (±5 years) and gender. Study participants returned to the clinic to donate blood and, approximately 3 months later, completed a questionnaire inquiring about demographics, health habits, and height and weight during an in-person interview at a home visit. Participation rates were 84 % for cases and 74 % for controls. From the 244 adenoma cases diagnosed within this study, we identified 131 cases with no previous history of rectal bleeding or adenoma, complete questionnaire data, and serum samples available. Cases were matched to an equal number of controls on age, sex, and smoking status. Of these participants, 129 cases and 129 controls had reported their height and weight.

The Shanghai study includes men and women selected at random from the Shanghai Women's Health Study (Zheng et al. 2005) and the Shanghai Men's Health Study (Jurj et al. 2007) for multiple measurements of physical activity by accelerometer over a study year. These two parent studies are population based prospective cohort studies from 8 communities of Shanghai, China, with recruitment occurring between 1997 and 2006. To be eligible, women had to be 40–70 years of age and men had to be 40–75 years of age at baseline. Our analysis included 106 women and 78 men enrolled in the first wave of the Shanghai physical activity study who donated blood at the end of the study year and had valid accelerometry measurements.

All participants provided informed consent. Institutional review boards at the NCI and each participating screening center (PLCO) or institute (Navy—National Naval Medical

Center; Shanghai—Shanghai Cancer Institute and Vanderbilt University) approved the studies.

2.2 BMI assessment

In the PLCO and Navy studies, participants were asked to report their current height and weight by questionnaire at baseline; these data were then used to calculate BMIs. In the Shanghai study, height and weight were measured twice during an in person interview according to a standard protocol by trained interviewers who were retired medical professionals.

2.3 Blood samples and batching

Blood samples from PLCO were collected from non-fasted participants at the individual screening centers. A serum tube (red top with no additive) was allowed to clot at room temperature for one hour, centrifuged to obtain serum ($1,200 \times g$ for 15 min), aliquoted into cryovials, and frozen at -70 °C within two hours of blood draw. Once per month, serum samples were shipped from screening centers on dry ice to a central biorepository in Frederick, Maryland, and stored at -70 °C.

Blood samples from Navy were collected from participants who were not required to fast. Time since last meal data was collected concurrently with the blood and approximately 58 people (22 %) had fasted, as defined as six or more hours since consuming any food or drink other than water. A serum tube (tiger-top serum separator tube) was allowed to clot at room temperature for 30 min, sent on ice to Biotech Research (Rockville, MD), centrifuged to obtain serum, and frozen at -80 °C within two hours of blood draw.

Blood samples from Shanghai were collected from participants asked to fast. Time since last meal data was collected and 151 of 184 (84 %) had fasted prior to blood draw. An EDTA plasma vacutainer tube was sent from participants' homes on ice to the Shanghai Cancer Institute. The tube was centrifuged and frozen at -70 °C within two hours of blood draw. After a large number of samples had been collected, tubes were sent on dry ice to a central biorepository in Frederick, Maryland, and stored at -70 °C.

For each study, samples were grouped into batches of 30, corresponding to the number of samples run per day. For each sample, the batch and position within a batch were randomly assigned except that, for the PLCO and Navy studies, cases and their matched controls were placed next to each other in the same batch.

2.4 Metabolite assessment

Metabolites were measured by Metabolon Inc. whose platform and process have been described previously (Evans et al. 2009). They used a single non-targeted extraction with methanol followed by protein precipitation to recover a diverse set of metabolites. Samples were then analyzed using ultra high performance liquid-phase chromatography and gas chromatography coupled with tandem mass spectrometry and mass spectrometry. The three studies were run in sequence on the same equipment. The mass spectra peaks were compared to a chemical reference library generated from 2,500 standards to identify individual metabolites. There were 317 metabolites of known identity measured at detectable levels in all three studies and these constitute the focus of our current investigation. The metabolites were grouped into 8 chemical classes (amino acids, carbohydrates, cofactors and vitamins, energy metabolites, lipids, nucleotide metabolites, peptides, and xenobiotics) based upon the classifications of the Kyoto Encyclopedia of Genes and Genomes (Ogata et al. 1999). Amino acids here include standard amino acids (the 20 amino acids that are protein precursors and are directly encoded by the universal genetic code) as well as non-standard amino acids (amino acids that are not incorporated into proteins or encoded by the universal genetic code).

We have previously reported on the level of reliability for the metabolomics platform used in this analysis (Sampson et al. 2013). The reliability was high, with technical intraclass correlation coefficients greater than 0.8 for more than 50 % of the metabolites. We also calculated intra-assay coefficients of variation, averaged over all individuals, for each of the 317 metabolites in PLCO, Shanghai, and Navy. Each batch of 30 samples included blinded quality control samples—2 samples per batch for the PLCO and Shanghai studies, and 3 samples per batch for the Navy study. The median coefficients of variation over the 317 metabolites were 0.10 (IQR: 0.04, 0.21) for PLCO, 0.21 (IQR: 0.14, 0.33) for Navy, and 0.14 (IQR: 0.09, 0.21) for Shanghai.

2.5 Statistical analysis

To account for variability by run day, we standardized the non-missing (i.e. detectable) values as a proportion of the median value observed that day. Thus, the median value for each metabolite for each run day was set to 1, and metabolite values twice that of the median for that day were set to 2, and so on. Within each study, we then transformed metabolite values to their natural logarithm to normalize distributions and centered them (mean = 0, standard deviation = 1). For each metabolite, we assumed that any missing values were values below the limit of detection and imputed the minimum of the non-missing values. Over the 317 metabolites, the median level of "missingness" before imputation was 1 %.

We examined study-specific associations between metabolite levels and BMI as a continuous variable using linear regression, adjusted for age, gender, and smoking status (never, former, current). PLCO models were additionally adjusted for study center and case status and Navy models were also adjusted for case status. We then used DerSimonian and Laird random effects models to conduct a combined meta-analysis of the three studies (DerSimonian and Laird 1986). Effect sizes for each study and in the meta-analysis indicate the increase in BMI units per one standard deviation increase in the metabolite level (on the log scale). Pearson correlations, adjusted for the same covariates, were also estimated in each study. For meta-analysis of study-specific Pearson correlations, we used Fisher's r to Z transformation and estimated the standard error as the square root of the sampling variance (Field 2001). Meta-analysis results were then back-transformed for presentation. The threshold for statistical significance was Bonferroni adjusted for 317 statistical tests, i.e. alpha = 0.05/317 = 0.000158.

We assessed consistency of the statistically significant associations across studies using the Q statistic (Takkouche et al. 1999) and by examining consistency of study-specific effect estimates, particularly whether they were all within the bounds of the meta-analysis 95 % confidence interval. To assess replication of findings, we examined whether P-values were consistently low—less than 0.05—across all 3 studies and whether effects were in the same direction.

To assess whether associations were independent, we evaluated the pairwise correlations between all metabolites significantly associated with BMI. Metabolites with pairwise correlations greater than 0.5 were considered to be highly correlated and to have possible redundancy. In supplementary analyses, we also used forward stepwise regression to identify metabolites independently associated with BMI. Metabolites were entered or removed from the model based on the meta-analysis P-value, with the threshold for entry and/or exit into the model set at $P = 0.05$. This model helps to identify a parsimonious set of BMI-related metabolites that could be carried forward to targeted analyses.

In a sensitivity analysis, we examined whether there was heterogeneity in metabolite-BMI associations according to fasting status in the Navy and Shanghai studies. We also examined correlations between fasting status and BMI to determine whether there is potential confounding by fasting status. In another sensitivity analysis, we excluded participants who developed colorectal cancer during follow-up (PLCO) and prevalent adenoma cases (Navy), reran models, and compared results to the main analysis.

All analyses were performed with SAS software version 9.1.3 (SAS Institute, Cary, NC) and the R statistical language version 2.14.0.

3 Results

The participant characteristics of our three study populations are shown in Table 1. Overall, participants from PLCO were older than those of the other two studies, while participants from the Navy study were more likely to be male, and participants from the Shanghai study had lower BMIs and the men were more likely to currently smoke than men from other studies.

We identified 37 metabolites associated with BMI at the Bonferroni adjusted significance level (Table 2). The distribution of the BMI associated metabolites according to chemical class was as follows: 19 amino acids, three carbohydrates, two peptides, 12 lipids, one xenobiotic, but no cofactors and vitamins, energy metabolites, or nucleotide metabolites. The specific metabolites and their effect sizes and P-values are shown in Table 3 and Fig. 1. The largest effect size and the most significant result—$P = 2.51 \times$

Table 1 Baseline characteristics of study participants

Baseline characteristics	PLCO[a] N = 505 (%)	Navy[b] N = 258 (%)	Shanghai[c] N = 184 (%)
Biospecimen	Serum	Serum	Plasma
Fasting status	Non-fasting	Non-fasting	Fasting[d]
Age (years)			
<60	19.5	59.9	52.7
60–64	31.6	17.2	14.7
65–69	30.4	13.0	14.7
70+	18.5	9.9	17.9
Gender			
Women	43.8	22.9	57.6
Men	56.2	77.1	42.4
Smoking			
Women			
Current	7.2	8.6	4.7
Former	39.8	36.2	0
Never	52.9	55.2	95.3
Men			
Current	10.2	7.0	53.8
Former	52.8	46.0	0
Never	37.0	47.0	46.2
BMI (kg/m^2)			
<20	2.6	1.9	13.6
20–24.9	30.2	32.1	47.8
25.0–29.9	42.7	46.9	35.3
30.0+	24.5	19.1	3.3

[a] The Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial

[b] The Navy Colon Adenoma Study

[c] The Shanghai Physical Activity Study

[d] 84 % adhered to fasting protocol

Table 2 Number of metabolites detected by chemical class, and the number with statistically significant associations with BMI

Chemical class	Total number of metabolites	Number (and %) of metabolites associated with BMI (alpha = 0.05/317)
All	317	37 (12 %)
Amino acids[a]	83	19 (23 %)
Carbohydrates	17	3 (18 %)
Peptides	11	2 (18 %)
Lipids	138	12 (9 %)
Xenobiotics	36	1 (3 %)
Cofactors and vitamins	16	0 (0 %)
Energy metabolites	7	0 (0 %)
Nucleotide metabolites	9	0 (0 %)

[a] The category of amino acids includes both standard amino acids, i.e. the 20 amino acids that are directly encoded by the universal genetic code and are precursors to proteins, as well as non-standard amino acids, i.e. amino acids that are not incorporated into proteins or not directly encoded for by the universal genetic code

10^{-13}—were observed for glutamate. For glutamate, each 1 standard deviation increase was associated with an increase of 1.39 units of BMI. For a person who is 1.7 meters in height, this would be an increase of 4.1 kg in body weight. At $P < 0.05$ level, there were 110 metabolites associated with BMI overall. Complete results for all 317 metabolites are shown in Table I of the online data supplement.

The associations had minimal heterogeneity by study. All tests for heterogeneity were non-significant ($P_{\text{heterogeneity}} > 0.05$). Effect sizes were in the same direction and of similar magnitude between studies. Out of 37 metabolites, only two (3-hydroxyisobutyrate and palmitoyl sphingomyelin) had study-specific estimates that differed substantially, i.e. by more than 50 %, from the meta-analysis estimate. Of the 37 metabolites, 25 had a P-value <0.05 for the metabolite-BMI association in all three studies (by chance alone, no metabolites would be expected to replicate at this level).

Of the 37 metabolites associated with BMI, there were 18 that have never, to our knowledge, been previously linked with BMI in the literature. Seven of these 18 metabolites were highly correlated ($r \geq 0.5$) with the known BMI-associated metabolites of valine, tyrosine, phenylalanine, leucine, or isoleucine (Fig. 2). These seven metabolites were 3-4-hydroxyphenyl-lactate, gamma-glutamyltyrosine, 3-methyl-2-oxobutyrate, gamma-glutamylisoleucine, 3-methyl-2-oxovalerate, 4-methyl-2-oxopentanoate, and 3-hydroxyisobutyrate. Adjusting for the known BMI-related metabolites in sensitivity analyses eliminated the associations for these seven metabolites, suggesting that they provide potentially redundant information.

The 11 remaining metabolites with novel BMI associations were not strongly correlated with known BMI-related metabolites and thus likely provide new information (Fig. 2). Three of these were intercorrelated lipids (1-oleoylglycerophosphocholine, 1-eicosadienoylglycerophosphocholine, 2-linoleoylglycerophosphocholine). The remaining 8 metabolites (butyrylcarnitine, 2-hydroxybutyrate, 7-alpha-hydroxy-3-oxo-4-cholestenoate, alpha-hydroxyisovalerate, benzoate, n-acetlyglycine, palmitoyl sphingomyelin, and histidine) were not strongly correlated with one another.

In addition to the novel associations, there were 19 additional metabolite-BMI associations that had been previously reported upon, including branched chain amino acids like valine, aromatic amino acids like tyrosine, markers of glucose control like mannose and 1-5-anhydroglucitol, and markers of lipid metabolism like glycerol and lathosterol.

Over the entire set of 317 metabolites, no associations were heterogeneous by study after Bonferroni correction (the lowest P-values were $P = 0.0004$ for serotonin 5-HT, $P = 0.0017$ for dimethylarginine-SDMA-ADMA, and $P = 0.0034$ for adrenate-22-4n6). There were also no associations with statistically significant heterogeneity by age or gender after Bonferroni correction, though some results for gender were of borderline significance. We included gender-specific results in Tables II and III of the online data supplement. The lowest p-values for heterogeneity by gender were $P = 0.0002$ for threitol, $P = 0.0005$ for histidine, and $P = 0.0005$ for benzoate.

In the stepwise regression model, 19 metabolites were associated with BMI at the $P < 0.05$ level (Table IV of the online data supplement). In this model, the associations are mutually adjusted and thus results are statistically independent from one another.

In analyses of heterogeneity by fasting status, there were no interactions after Bonferroni correction. Forty five metabolites met the nominal level of statistical significance ($P < 0.05$) but differences in effect size were modest, generally differing by 0.05–0.07 (or −0.05 to −0.07) for fasted vs. non-fasted samples. Regarding the specific interactions observed, none of the 37 metabolite-BMI associations were weaker in fasting only samples; six were slightly stronger (for the metabolites of gamma glutamyltyrosine, lathosterol, andro steroid monosulfate 2, isovalerylcarnitine, leucine, and carnitine). For metabolites not associated with BMI, only a few could potentially have reached statistical significance in fasting only samples based on the interactions observed (kynurenate, gamma glutamylvaline, gamma glutmylleucine, urate, 1′arachidonoylglycerophosinositol, dihomo-linoleate, and serine). In addition, fasting status was not associated with BMI in either the Navy study (Pearson $r = -0.09$, $P = 0.15$) or the Shanghai study (Pearson $r = 0.06$, $P = 0.40$), suggesting against major confounding by fasting status.

Table 3 Association of metabolites with BMI in the combined random effects meta-analysis and by specific study

Metabolite	Chemical class	Overall ES[a]	95 % CI	Overall r	Meta-analysis (P-value)	$P_{heterogeneity}$	PLCO ES	Navy ES	Shanghai ES	PLCO (P-value)	Navy (P-value)	Shanghai (P-value)
Novel associations												
3-(4-Hydroxyphenyl) lactate	Amino acid	0.86	0.58, 1.15	0.19	3.10×10^{-09}	0.95	0.84	0.83	0.93	0.0001	0.0048	0.0007
Gamma-glutamyltyrosine	Peptide	0.85	0.58, 1.12	0.19	6.27×10^{-10}	0.83	0.83	0.74	0.96	6.03×10^{-05}	0.0112	0.0001
Butyrylcarnitine	Lipid	0.83	0.56, 1.09	0.20	9.95×10^{-10}	0.91	0.87	0.84	0.74	1.36×10^{-05}	0.0034	0.0024
2-Hydroxybutyrate (AHB)	Amino acid	0.82	0.55, 1.09	0.19	2.67×10^{-09}	0.54	0.96	0.57	0.78	2.40×10^{-06}	0.0601	0.0016
3-Methyl-2-oxobutyrate	Amino acid	0.76	0.49, 1.03	0.14	2.87×10^{-08}	0.69	0.88	0.73	0.61	1.75×10^{-05}	0.0115	0.0133
Gamma-glutamylisoleucine	Peptide	0.73	0.39, 1.06	0.16	2.03×10^{-05}	0.26	0.66	0.45	1.08	0.0021	0.1205	0.0001
7-HOCA[b]	Lipid	0.68	0.41, 0.95	0.16	9.66×10^{-07}	0.41	0.88	0.51	0.51	2.26×10^{-05}	0.0760	0.0394
Alpha-hydroxyisovalerate	Amino acid	0.62	0.34, 0.90	0.14	1.19×10^{-05}	0.49	0.46	0.65	0.87	0.0246	0.0273	0.0013
3-Methyl-2-oxovalerate	Amino acid	0.61	0.33, 0.89	0.14	1.68×10^{-05}	0.57	0.51	0.50	0.83	0.0165	0.0847	0.0012
4-Methyl-2-oxopentanoate	Amino acid	0.57	0.29, 0.85	0.13	6.05×10^{-05}	0.38	0.40	0.51	0.85	0.0659	0.0855	0.0008
3-Hydroxyisobutyrate	Amino acid	0.56	0.28, 0.83	0.13	6.88×10^{-05}	0.52	0.65	0.27	0.64	0.0019	0.3636	0.0111
1-Oleoylglycerophosphocholine	Lipid	−0.52	−0.79, −0.25	−0.13	0.00015	0.49	−0.59	−0.71	−0.29	0.0039	0.0146	0.2517
1-Eicosadienoylglycerophosphocholine	Lipid	−0.53	−0.80, −0.27	−0.13	9.56×10^{-05}	0.51	−0.65	−0.62	−0.30	0.0014	0.0296	0.2338
2-Linoleoylglycerophosphocholine	Lipid	−0.61	−0.88, −0.33	−0.14	1.36×10^{-05}	0.87	−0.57	−0.74	−0.57	0.0066	0.0103	0.0255
Benzoate	Xenobiotics	−0.63	−0.93, −0.33	−0.15	3.41×10^{-05}	0.31	−0.86	−0.54	−0.38	3.06×10^{-05}	0.0605	0.1384
N-Acetylglycine	Amino acid	−0.67	−0.94, −0.39	−0.16	1.58×10^{-06}	0.40	−0.87	−0.53	−0.47	2.60×10^{-05}	0.0731	0.0594
Palmitoyl sphingomyelin	Lipid	−0.70	−0.99, −0.41	−0.16	1.98×10^{-06}	0.34	−0.86	−0.34	−0.74	3.30×10^{-05}	0.2397	0.0043
Histidine	Amino acid	−0.89	−1.16, −0.63	−0.21	7.61×10^{-11}	0.57	−1.05	−0.79	−0.74	2.85×10^{-07}	0.0064	0.0034
Previously reported associations												
Glutamate	Amino acid	1.39	1.02, 1.77	0.33	2.51×10^{-13}	0.14	1.70	1.33	1.09	5.09×10^{-16}	2.36×10^{-06}	8.03×10^{-06}
Valine	Amino acid	1.01	0.74, 1.28	0.25	4.17×10^{-13}	0.35	0.78	1.10	1.22	0.0002	0.0001	4.16×10^{-07}
Mannose	Carbohydrate	0.95	0.56, 1.33	0.23	1.43×10^{-06}	0.14	1.15	1.14	0.55	2.27×10^{-08}	7.80×10^{-05}	0.0266
Isoleucine	Amino acid	0.93	0.65, 1.20	0.22	8.15×10^{-11}	0.35	0.73	0.89	1.19	0.0006	0.0019	1.32×10^{-06}
Isovalerylcarnitine	Amino acid	0.93	0.60, 1.25	0.22	2.58×10^{-08}	0.24	0.87	0.62	1.22	4.21×10^{-05}	0.0320	5.09×10^{-07}
Tyrosine	Amino acid	0.91	0.65, 1.18	0.21	1.97×10^{-11}	0.91	0.86	0.91	0.99	3.48×10^{-05}	0.0014	5.12×10^{-05}
Leucine	Amino acid	0.88	0.51, 1.25	0.22	2.64×10^{-06}	0.16	0.65	0.77	1.24	0.0019	0.0076	5.62×10^{-07}
Andro steroid monosulfate 2	Lipid	0.82	0.54, 1.11	0.18	9.91×10^{-09}	0.87	0.91	0.78	0.75	4.48×10^{-05}	0.0103	0.0024
Lathosterol	Lipid	0.78	0.51, 1.05	0.18	9.85×10^{-09}	0.65	0.81	0.55	0.89	5.24×10^{-05}	0.0545	0.0003
Kynurenine	Amino acid	0.74	0.47, 1.01	0.17	9.42×10^{-08}	0.65	0.82	0.50	0.79	6.51×10^{-05}	0.0907	0.0016
Glycerol	Lipid	0.67	0.38, 0.97	0.14	8.34×10^{-06}	0.87	0.76	0.60	0.60	0.0009	0.0593	0.0212
Phenylalanine	Amino acid	0.66	0.39, 0.93	0.16	1.30×10^{-06}	0.65	0.57	0.89	0.63	0.0052	0.0019	0.0113
Hexanoylcarnitine	Lipid	0.63	0.35, 0.91	0.14	8.53×10^{-06}	0.77	0.67	0.45	0.71	0.0018	0.1216	0.0049
Propionylcarnitine	Lipid	0.62	0.34, 0.89	0.14	1.22×10^{-05}	0.77	0.70	0.45	0.62	0.0009	0.1295	0.0145
Carnitine	Lipid	0.58	0.31, 0.85	0.13	2.76×10^{-05}	0.82	0.48	0.68	0.63	0.0183	0.0176	0.0107
Lactate	Carbohydrate	0.54	0.27, 0.82	0.12	0.00010	0.99	0.53	0.54	0.56	0.0125	0.0604	0.0231
1,5-AG[c]	Carbohydrate	−0.69	−0.96, −0.42	−0.16	4.42×10^{-07}	0.64	−0.79	−0.45	−0.71	9.06×10^{-05}	0.1239	0.0042

Table 3 continued

Metabolite	Chemical class	Overall ES[a]	95 % CI	Overall r	Meta-analysis (P-value)	$P_{heterogeneity}$	PLCO ES	Navy ES	Shanghai ES	PLCO (P-value)	Navy (P-value)	Shanghai (P-value)
Asparagine	Amino acid	−0.80	−1.10, −0.51	−0.19	9.18×10^{-08}	0.30	−1.04	−0.62	−0.61	3.07×10^{-07}	0.0310	0.0132
Glycine	Amino acid	−0.91	−1.18, −0.61	−0.21	2.53×10^{-11}	0.89	−0.91	−0.80	−0.98	8.02×10^{-06}	0.0065	6.15×10^{-05}

Results are based on a population of 947 participants (505 from PLCO, 258 from Navy, and 184 from Shanghai). Within each specific study, the association between metabolite levels and BMI was modeled using linear regression, adjusted for age, gender, current smoking status. PLCO models were additionally adjusted for study center and case status and Navy models were also adjusted for case status. The combined model was done using random effects meta-analysis. Only the metabolites that met the Bonferroni corrected threshold of statistical significance in the combined models, i.e. alpha = 0.05/317 = 0.000158, are shown here. Complete results are shown in Table I of the online data supplement

[a] Effect Size (ES) indicates the increase in units of BMI (kg/m^2) per one standard deviation increase in metabolite level (log-scale)

[b] 7-alpha-hydroxy-3-oxo-4-cholestenoate

[c] 1-5-anhydroglucitol

In analyses that excluded participants who later developed colorectal cancer (PLCO study) or had a prevalent adenoma (Navy study), results remained materially the same as in the main analysis (see Table V of the online data supplement). For the 37 BMI-associated metabolites, changes in effect size were almost uniformly less than 10 %. Also, we examined heterogeneity by colorectal adenoma status (Navy study), and future colorectal cancer status (PLCO study), but found no heterogeneity at the Bonferroni corrected level.

4 Discussion

Through meta-analysis of metabolomics data from three study populations, we identified 37 metabolites associated with BMI at the Bonferroni corrected level of statistical significance, including 19 lipids, 12 amino acids, and 6 other metabolites. The associations were strong, had highly significant P-values, and replicated across the studies at the $P < 0.05$ level. Of the 37 metabolites, 18 metabolites, to our knowledge, had never before been associated with BMI. Of these 18 metabolites, 11 had no strong correlations with known BMI-related metabolites, and thus are considered to be "new" BMI-associated metabolites. We also replicated previously observed associations for 19 metabolites. Overall, the strongest associations were observed for amino acids.

Six of the 11 "new" BMI-associated metabolites were lipids. Prior research suggests that fatty acid oxidation is higher among those with insulin-resistance (Newgard 2012), and our results suggest a similar result for high BMI. Butyrylcarnitine and 2-hydroxybutyrate (an amino acid), for example, are both known markers of excessive fatty acid oxidation (van Maldegem et al. 2006; Gall et al. 2010) and both were positively associated with BMI. The other five lipids (7-HOCA, 1-oleoylglycerophosphocholine, 1-eicosadienoylglycerophosphocholine, 2-linoleoylglycerophosphocholine, and palmitoyl sphingomyelin) have not been well-characterized with respect to health and could be interesting targets for future research.

Lipids also figure prominently among the replicated associations, comprising 6 of the 19 metabolites with a previously observed BMI association. Of particular interest were associations for glycerol and lathosterol because prior studies had been small and replication was needed (Puhakainen et al. 1992; Paramsothy et al. 2011). Glycerol is a three-carbon molecule that forms the backbone of all triglycerides and blood levels of glycerol reflect the rate of fatty acid breakdown (Puhakainen et al. 1992; Venables et al. 2008). Lathosterol is a cholesterol precursor and blood levels reflect the rate of cholesterol synthesis (Pihlajamaki et al. 2004; Miettinen et al. 1990). Interestingly,
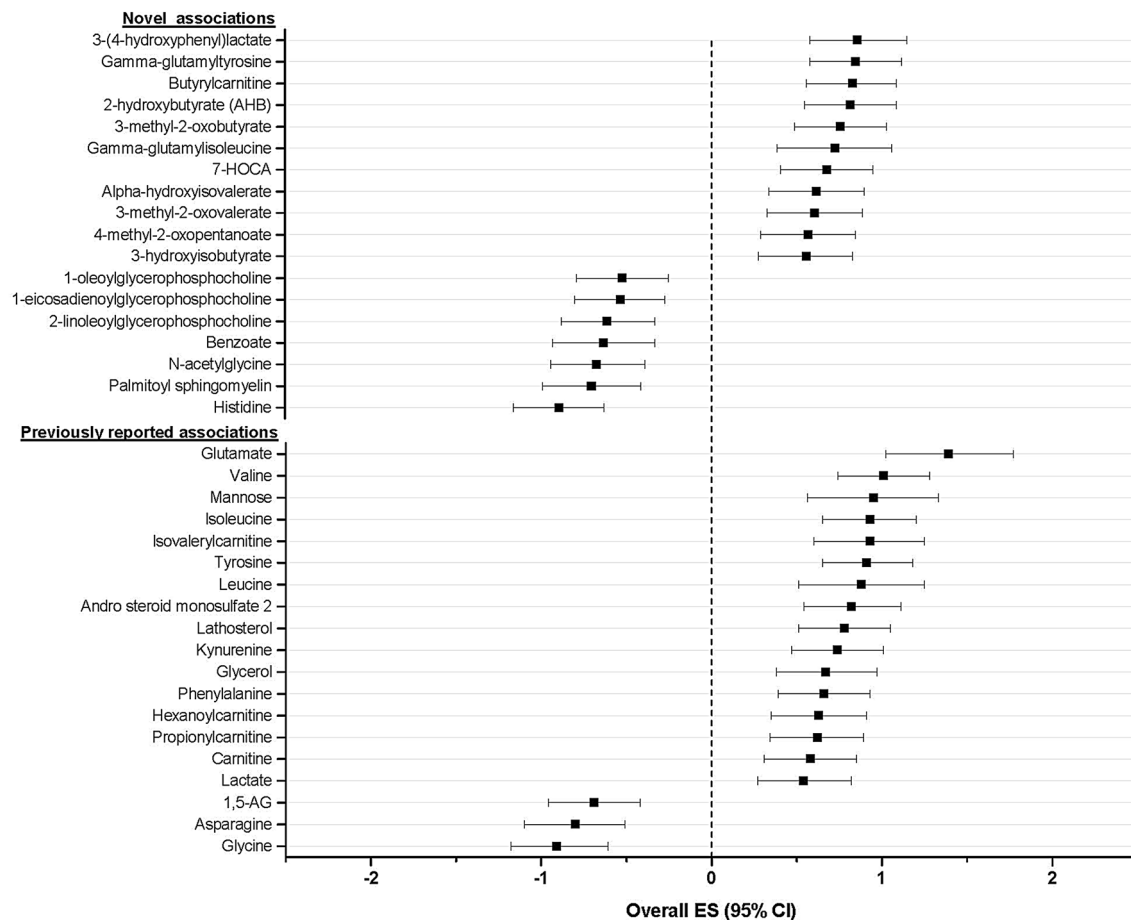
Fig. 1 Effect sizes and 95 percent confidence intervals for combined associations between metabolite levels and body mass index. Results are based on a population of 947 participants (505 from PLCO, 258 from Navy, and 184 from Shanghai). Within each specific study, the association between metabolite levels and BMI was modeled using linear regression, adjusted for age, gender, current smoking status. PLCO models were additionally adjusted for study center and case status and Navy models were also adjusted for case status. The combined model was done using random effects meta-analysis. Only the metabolites that met the Bonferroni corrected threshold of statistical significance in the combined models, i.e. alpha = 0.05/317 = 0.000158, are shown here. The overall effect size (ES)—the *black square*—indicates the change in units BMI (kg/m^2) per one standard deviation increase in metabolite level (log scale). The *horizontal line* indicates the 95 percent confidence interval of the estimate. The *dotted line* at zero overall ES is the line of no effect

lathosterol was much more strongly associated with BMI than was cholesterol, possibly indicating that lathosterol reflects an aspect of cholesterol metabolism that is especially sensitive to an elevated BMI.

Four of the 11 "new" BMI-associated metabolites were amino acids: 2-hydroxybutyrate (discussed above), alpha-hydroxyisovalerate, n-acetylglycine, and histidine. Alpha-hydroxyisovalerate has been implicated in several metabolic disorders, including phenylketonuria. N-acetylglycine is the acetylation product of glycine, but the significance of this metabolite for health is otherwise unknown. Histidine is a proteinogenic amino acid and a precursor of the neurotransmitter histamine. Our study is the first to identify a statistically significant association between histidine and BMI although three prior studies did examine this association (Cheng et al. 2012; Newgard et al. 2009; Felig et al. 1969).

Our analysis also replicated associations already established between amino acids and BMI, including associations for valine, isoleucine, leucine, tyrosine, phenylalanine, glutamate, kynurenine, and glycine (Gaudet et al. 2012; Newgard et al. 2009; Felig et al. 1969; Cheng et al. 2012). Although a link between amino acids and BMI was first noted four decades ago in a clinical study by Felig et al. (1969), this link is being reexamined due to their recent replication in metabolic profiling studies (Gaudet et al. 2012; Newgard et al. 2009; Cheng et al. 2012; Floegel et al. 2012), and the additional finding that these amino acids are associated with future diabetes risk, even after controlling for BMI (Wang et al. 2011; Floegel et al. 2012). Possibly, these amino acids are biologically important for diabetes risk, above and beyond their association with BMI.

Why amino acid levels are high among those with high BMIs is not well-understood. One theory is that insulin-
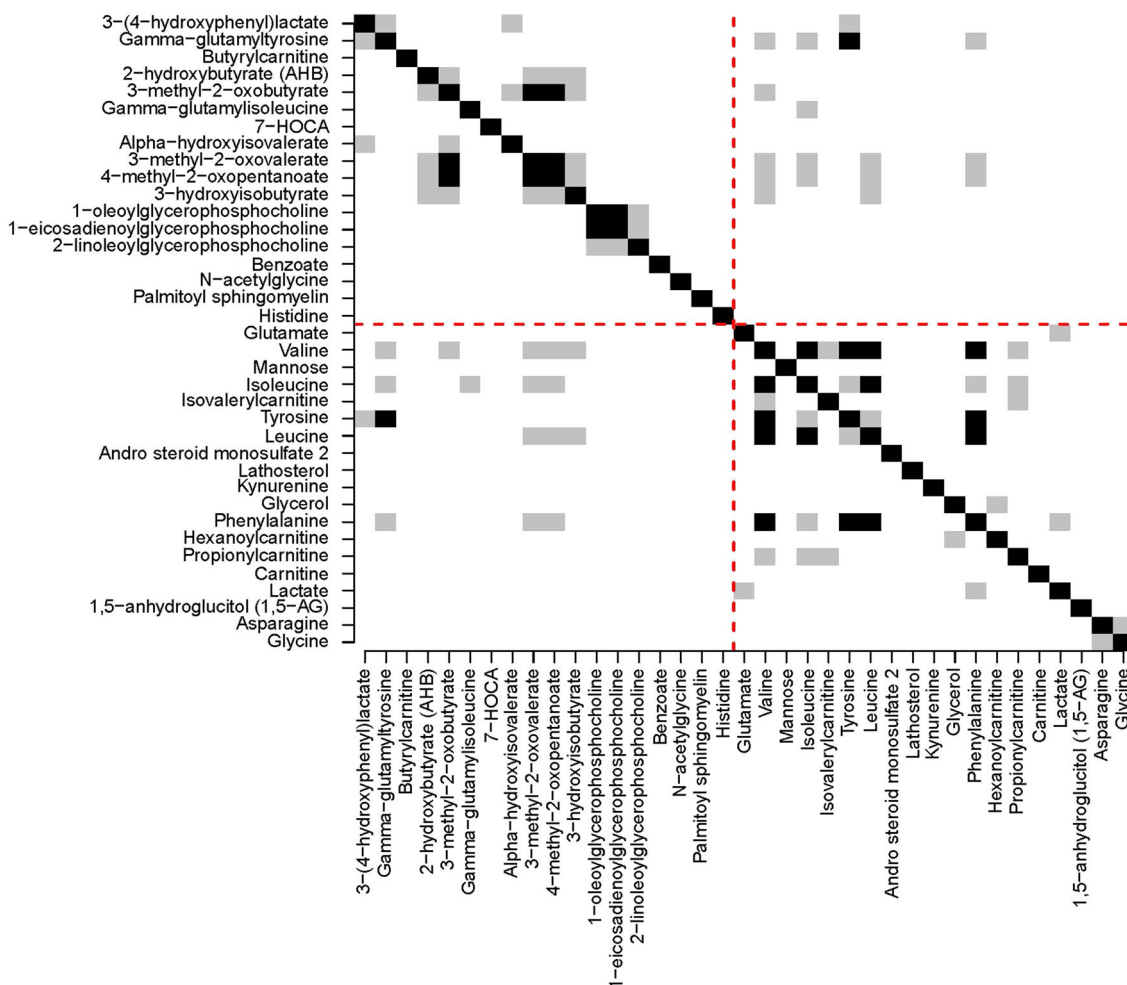
Fig. 2 Pairwise correlations in the PLCO dataset for the 37 metabolites associated with BMI. A *white square* represents a pairwise correlation of less than 0.50, a *grey square* represents a correlation 0.50–0.69, and *a black square* indicates a correlation of 0.70 or greater. The *red dashed line* separates metabolites that had been previously reported to be associated with BMI (bottom and/or right) vs. metabolites that had never before been reported to be associated with BMI (Color figure online)

resistance (which tends to co-occur with a high BMI) causes excess protein breakdown in skeletal muscle and amino acid breakdown products are released into the blood (Felig et al. 1969; Forlani et al. 1984; Tremblay et al. 2007). Alternatively, the high fat free mass among those with high BMIs may be related to higher protein turnover, which may, in turn, cause amino acid breakdown products to be released into the blood. Finally, high amino acid levels may themselves affect eating behavior through their neuroactive properties. The amino acids associated with BMI include two neurotransmitters (glutamate, and glycine), three neurotransmitter precursors (histidine, tyrosine, and phenylalanine), and kynurenine, which has neuroactive properties. Several of these have been implicated in appetite regulation (Haas et al. 2008; Wu et al. 2012; Johnson and Kenny 2010). A limitation to this theory is that, due to interference by the blood–brain barrier, it is not certain that amino acid blood levels will correlate with their level of utilization in the brain. For one amino acid neurotransmitter—tyrosine—there is at least preliminary evidence indicating a link between blood levels and brain activity (Montgomery et al. 2003).

The final "new" BMI-associated metabolite was benzoate, a common preservative added to food and personal care products. In exploratory analyses, benzoate levels were modestly higher among women and highly correlated ($r > 0.5$) with levels of metabolites found in personal care products, including 2-ethylhexanoate, heptanoate, and pelargonate. Possibly, the benzoate-BMI association is driven by behavioral patterns related to personal care product use. Alternately, there may be greater sequestration of these metabolites among people of a higher BMI.

For carbohydrates, we replicated known associations between BMI and mannose, lactate, and 1-5-anhydroglucitol. Mannose is a C-2 epimer of glucose and blood mannose levels correlate highly with those of glucose (Sone

et al. 2003; Pitkanen et al. 1999), but are less sensitive to food intake and are more tightly linked with diabetes (Sone et al. 2003). Lactate is a product of glycolysis when oxygen supply is limited, a state that may occur more frequently among those with high BMIs due to poorer circulation (Trayhurn et al. 2008). 1-5-anhydroglucitol is an established biomarker for short-term glucose control (McGill et al. 2004).

A major strength of our study was the large number of metabolites assayed, which allowed us to discover new associations, and to confirm associations from small studies in the literature that might otherwise have been neglected. Additional strengths were its comparatively large sample size, and the replication of associations across three ethnically diverse study populations. This means that our findings are unlikely to be due to chance, and that they should have a reasonably high level of generalizability in different studies and ethnic groups.

There were several important limitations. Our study is cross-sectional, which complicates efforts to determine the causal nature of associations. BMI does not discriminate between fat and lean mass, and does not measure body fat distribution. However, BMI, body fat percent, and trunk fat are typically highly correlated and, in at least one study (Sun et al. 2010), they have been similarly associated with obesity-related biomarkers and metabolic syndrome. BMI also was determined using self-reported weight and height in our largest two studies (PLCO and Navy) but correlations between measured and self-reported weight and height typically exceed 0.9 (Willett 2012) and so errors were likely modest. The metabolite levels in our study were based on blood samples from a single point in time and, if metabolite levels vary over time, the associations could be attenuated relative to the true associations (Sampson et al. 2013). Our blood samples also differed between studies in type (serum for PLCO and Navy, plasma for Shanghai) and fasting status (non-fasting for PLCO and Navy, primarily fasting for Shanghai) which would add heterogeneity and could result in some false negatives. On the other hand, the findings that we did report are likely to be robust across a variety of sample collection designs typical for epidemiologic studies. Finally, our metabolomics analysis was limited to metabolites that could be detected by the platforms used and our data were measured in terms of relative concentrations rather than actual concentrations.

In summary, we used metabolomics to identify a large and diverse set of metabolites associated with BMI. Our results provide a critical baseline for the establishment of the "metabolome" of BMI. They also suggest a large number of candidate markers that can be explored by future prospective studies as potential mediators of the associations between obesity and disease risk.

References

Chan, J. M., Rimm, E. B., Colditz, G. A., Stampfer, M. J., & Willett, W. C. (1994). Obesity, fat distribution, and weight gain as risk factors for clinical diabetes in men. *Diabetes Care, 17*(9), 961–969.

Cheng, S., Rhee, E. P., Larson, M. G., et al. (2012). Metabolite profiling identifies pathways associated with metabolic risk in humans. *Circulation, 125*(18), 2222–2231.

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7*(3), 177–188.

Evans, A. M., Dehaven, C. D., Barrett, T., Mitchell, M., & Milgram, E. (2009). Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. *Analytical Chemistry, 81*(16), 6656–6667.

Felig, P., Marliss, E., & Cahill, G. F், Jr. (1969). Plasma amino acid levels and insulin secretion in obesity. *New England Journal of Medicine, 281*(15), 811–816.

Field, A. P. (2001). Meta-analysis of correlation coefficients: a Monte Carol comparison of fixed- and random-effects methods. *Psychological Methods, 6*(2), 161–180.

Floegel, A., Stefan, N., Yu, Z., et al. (2012). Identification of Serum Metabolites Associated With Risk of Type 2 Diabetes Using a Targeted Metabolomic Approach. *Diabetes, 62*(2), 639–648.

Forlani, G., Vannini, P., Marchesini, G., et al. (1984). Insulin-dependent metabolism of branched-chain amino acids in obesity. *Metabolism, 33*(2), 147–150.

Gall, W. E., Beebe, K., Lawton, K. A., et al. (2010). alpha-hydroxybutyrate is an early biomarker of insulin resistance and glucose intolerance in a nondiabetic population. *PLoS One, 5*(5), e10883.

Gaudet, M. M., Falk, R. T., Stevens, R. D., et al. (2012). Analysis of serum metabolic profiles in women with endometrial cancer and controls in a population-based case-control study. *Journal of Clinical Endocrinology and Metabolism, 97*(9), 3216–3223.

Haas, H. L., Sergeeva, O. A., & Selbach, O. (2008). Histamine in the nervous system. *Physiological Reviews, 88*(3), 1183–1241.

Johnson, P. M., & Kenny, P. J. (2010). Dopamine D2 receptors in addiction-like reward dysfunction and compulsive eating in obese rats. *Nature Neuroscience, 13*(5), 635–641.

Jurj, A. L., Wen, W., Xiang, Y. B., et al. (2007). Reproducibility and validity of the Shanghai Men's Health Study physical activity questionnaire. *American Journal of Epidemiology, 165*(10), 1124–1133.

McGill, J. B., Cole, T. G., Nowatzke, W., et al. (2004). Circulating 1,5-anhydroglucitol levels in adult patients with diabetes reflect longitudinal changes of glycemia: a U.S. trial of the GlycoMark assay. *Diabetes Care, 27*(8), 1859–1865.

Miettinen, T. A., Tilvis, R. S., & Kesaniemi, Y. A. (1990). Serum plant sterols and cholesterol precursors reflect cholesterol absorption and synthesis in volunteers of a randomly selected male population. *American Journal of Epidemiology, 131*(1), 20–31.

Montgomery, A. J., McTavish, S. F., Cowen, P. J., & Grasby, P. M. (2003). Reduction of brain dopamine concentration with dietary tyrosine plus phenylalanine depletion: an [11C] raclopride PET study. *American Journal of Psychiatry, 160*(10), 1887–1889.

National Cancer Institute. *Provocative Questions Project.* http://provocativequestions.nci.nih.gov/rfa. Accessed March 11, 2013.

Newgard, C. B. (2012). Interplay between lipids and branched-chain amino acids in development of insulin resistance. *Cell Metabolism, 15*(5), 606–614.

Newgard, C. B., An, J., Bain, J. R., et al. (2009). A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell Metabolism, 9*(4), 311–326.

Ogata, H., Goto, S., Sato, K., et al. (1999). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research, 27*(1), 29–34.

Paramsothy, P., Knopp, R. H., Kahn, S. E., et al. (2011). Plasma sterol evidence for decreased absorption and increased synthesis of cholesterol in insulin resistance and obesity. *American Journal of Clinical Nutrition, 94*(5), 1182–1188.

Pihlajamaki, J., Gylling, H., Miettinen, T. A., & Laakso, M. (2004). Insulin resistance is associated with increased cholesterol synthesis and decreased cholesterol absorption in normoglycemic men. *Journal of Lipid Research, 45*(3), 507–512.

Pitkanen, O. M., Vanhanen, H., & Pitkanen, E. (1999). Metabolic syndrome is associated with changes in D-mannose metabolism. *Scandinavian Journal of Clinical and Laboratory Investigation, 59*(8), 607–612.

Poirier, P., Giles, T. D., Bray, G. A., et al. (2006). Obesity and cardiovascular disease: pathophysiology, evaluation, and effect of weight loss: an update of the 1997 American Heart Association Scientific Statement on Obesity and Heart Disease from the Obesity Committee of the Council on Nutrition, Physical Activity, and Metabolism. *Circulation, 113*(6), 898–918.

Prorok, P. C., Andriole, G. L., Bresalier, R. S., et al. (2000). Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. *Controlled Clinical Trials, 21*(6 Suppl), 273S–309S.

Puhakainen, I., Koivisto, V. A., & Yki-Jarvinen, H. (1992). Lipolysis and gluconeogenesis from glycerol are increased in patients with noninsulin-dependent diabetes mellitus. *Journal of Clinical Endocrinology and Metabolism, 75*(3), 789–794.

Renehan, A. G., Tyson, M., Egger, M., Heller, R. F., & Zwahlen, M. (2008). Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet, 371*(9612), 569–578.

Roberts, D. L., Dive, C., & Renehan, A. G. (2010). Biological mechanisms linking obesity and cancer risk: new perspectives. *Annual Review of Medicine, 61*, 301–316.

Sampson, J. N., Boca, S. M., Shu, X. O., et al. (2013). Metabolomics in Epidemiology: Sources of Variability in Metabolite Measurements and Implications. *Cancer Epidemiol Biomarkers Prev, 22*(4), 631–640.

Sinha, R., Chow, W. H., Kulldorff, M., et al. (1999). Well-done, grilled red meat increases the risk of colorectal adenomas. *Cancer Research, 59*(17), 4320–4324.

Sone, H., Shimano, H., Ebinuma, H., et al. (2003). Physiological changes in circulating mannose levels in normal, glucose-intolerant, and diabetic subjects. *Metabolism, 52*(8), 1019–1027.

Sun, Q., van Dam, R. M., Spiegelman, D., et al. (2010). Comparison of dual-energy x-ray absorptiometric and anthropometric measures of adiposity in relation to adiposity-related biologic factors. *American Journal of Epidemiology, 172*(12), 1442–1454.

Takkouche, B., Cadarso-Suarez, C., & Spiegelman, D. (1999). Evaluation of old and new tests of heterogeneity in epidemiologic meta-analysis. *American Journal of Epidemiology, 150*(2), 206–215.

Trayhurn, P., Wang, B., & Wood, I. S. (2008). Hypoxia in adipose tissue: a basis for the dysregulation of tissue function in obesity? *British Journal of Nutrition, 100*(2), 227–235.

Tremblay, F., Brule, S., Hee, U. S., et al. (2007). Identification of IRS-1 Ser-1101 as a target of S6K1 in nutrient- and obesity-induced insulin resistance. *Proc Natl Acad Sci U S A, 104*(35), 14056–14061.

van Maldegem, B. T., Duran, M., Wanders, R. J., et al. (2006). Clinical, biochemical, and genetic heterogeneity in short-chain acyl-coenzyme A dehydrogenase deficiency. *JAMA, 296*(8), 943–952.

Venables, M. C., Hulston, C. J., Cox, H. R., & Jeukendrup, A. E. (2008). Green tea extract ingestion, fat oxidation, and glucose tolerance in healthy humans. *American Journal of Clinical Nutrition, 87*(3), 778–784.

Wang, T. J., Larson, M. G., Vasan, R. S., et al. (2011). Metabolite profiles and the risk of developing diabetes. *Nature Medicine, 17*(4), 448–453.

Willett, W. C. (2012). *Nutritional epidemiology* (3rd ed.). New York: Oxford Univ.

Wu, Q., Clark, M. S., & Palmiter, R. D. (2012). Deciphering a neuronal circuit that mediates appetite. *Nature, 483*(7391), 594–597.

Zheng, W., Chow, W. H., Yang, G., et al. (2005). The Shanghai Women's Health Study: rationale, study design, and baseline characteristics. *American Journal of Epidemiology, 162*(11), 1123–1131.