

Methylation potential associated with diet, genotype, protein, and metabolite levels in the Delta Obesity Vitamin Study

Jacqueline Pontes Monteiro · Carolyn Wise · Melissa J. Morine · Candee Teitel · Lisa Pence · Anna Williams · Beverly McCabe-Sellers · Catherine Champagne · Jerome Turner · Beatrice Shelby · Baitang Ning · Joan Oguntimein · Lauren Taylor · Terri Toennesen · Corrado Priami · Richard D. Beger · Margaret Bogle · Jim Kaput

Received: 1 February 2014 / Accepted: 6 April 2014 / Published online: 24 April 2014
© Springer-Verlag Berlin Heidelberg 2014

Abstract Micronutrient research typically focuses on analyzing the effects of single or a few nutrients on health by analyzing a limited number of biomarkers. The observational study described here analyzed micronutrients, plasma proteins, dietary intakes, and genotype using a systems approach. Participants attended a community-based summer day program for 6–14 year old in 2 years. Genetic makeup, blood metabolite and protein levels, and dietary differences were measured in each individual. Twenty-four-hour dietary intakes, eight micronutrients

(vitamins A, D, E, thiamin, folic acid, riboflavin, pyridoxal, and pyridoxine) and 3 one-carbon metabolites [homocysteine (Hcy), *S*-adenosylmethionine (SAM), and *S*-adenosylhomocysteine (SAH)], and 1,129 plasma proteins were analyzed as a function of diet at metabolite level, plasma protein level, age, and sex. Cluster analysis identified two groups differing in SAM/SAH and differing in dietary intake patterns indicating that SAM/SAH was a potential marker of nutritional status. The approach used to analyze genetic association with the SAM/SAH metabolites is called middle-out: SNPs in 275 genes involved in the one-carbon pathway (folate, pyridoxal/pyridoxine, thiamin) or were correlated with SAM/SAH (vitamin A, E, Hcy) were analyzed instead of the entire 1M SNP data set. This procedure identified 46 SNPs in 25 genes associated with SAM/SAH demonstrating a genetic contribution to the methylation potential. Individual plasma metabolites correlated with 99 plasma proteins. Fourteen proteins correlated with body mass index, 49 with group age, and 30 with sex. The analytical strategy described here identified subgroups for targeted nutritional interventions.

Jim Kaput affiliated with Division of Personalized Nutrition and Medicine, NCTR/FDA, Jefferson, AR, USA for portions of this study.

Carolyn Wise: Deceased.

Disclaimer: The contents of this manuscript are from the authors and do not necessarily reflect any position of the Government or the Food and Drug Administration or the Nestle Institute of Health Sciences.

Electronic supplementary material The online version of this article (doi:10.1007/s12263-014-0403-9) contains supplementary material, which is available to authorized users.

J. P. Monteiro
Department of Pediatrics, Faculty of Medicine, Faculty of Nutrition and Metabolism, University of São Paulo, Ribeirão Preto, SP, Brazil

C. Wise · C. Teitel · A. Williams · B. Ning · T. Toennesen
Division of Personalized Nutrition and Medicine, National Center for Toxicological Research (NCTR), Food and Drug Administration (FDA), Jefferson, AR, USA

M. J. Morine · C. Priami
Department of Mathematics, University of Trento, Trento, Italy

M. J. Morine · C. Priami
The Microsoft Research, University of Trento Centre for Computational and Systems Biology (COSBI), Rovereto, Italy

L. Pence · R. D. Beger
Division of Systems Biology, NCTR/FDA, Jefferson, AR, USA

B. McCabe-Sellers · M. Bogle
Delta Obesity Prevention Research Unit, United States Department of Agriculture, Agricultural Research Service, Little Rock, AR, USA

C. Champagne
Dietary Assessment and Nutrition Counseling, Pennington Biomedical Research Center, Baton Rouge, LA, USA

J. Turner · B. Shelby
Boys, Girls, Adults Community Development Center & The Phillips County Community Partners, Marvell, AR, USA

Keywords Systems nutrition · Methylation potential · SAM/SAH · Micronutrients · Community-based participatory research

Abbreviations

| | |
|--------|--|
| ANCOVA | Analysis of covariance |
| ANOVA | Analysis of variance |
| BGACDC | Boys, Girls, Adults Community Development Center |
| CBPR | Community-based participatory research |
| GWAS | Genome-wide association studies |
| HEI | Healthy Eating Index |
| Hcy | Homocysteine |
| NHANES | National Health and Nutrition Examination Survey |
| FD | Saturated fats |
| SoFAAS | Solid fats, alcoholic beverages, and added sugars |
| SAH | S-Adenosylhomocysteine |
| SAM | S-Adosylmethionine |
| TDGOVL | Total dark green and orange vegetables and legumes |
| WF | Whole fruit |
| WG | Whole grains |

Introduction

S-Adosylmethionine (SAM) is the methyl donor for 200 methyltransferase reactions of DNA, RNA, proteins, and metabolites (Finkelstein 2007; Lennard 2010) involved in a wide range of metabolic and signaling pathways. The transfer of the methyl group leads to the formation of S-adenosylhomocysteine (SAH), which is subsequently metabolized to homocysteine (Hcy). Plasma levels of SAM, SAH, and Hcy have independently and in combinations been associated with cardiovascular, neurological,

immunological, and obesity pathologies (Poirier et al. 2001c; Guerra-Shinohara et al. 2004; Strain et al. 2004; Selley 2007; Van Hecke et al. 2008; Ming et al. 2009; Obeid et al. 2009; Panza et al. 2009a, b; Linnebank et al. 2010; Muller 2010; Altug Sen et al. 2011). SAM is a product of the one-carbon pathway that requires nutritional cofactors folate, vitamin B6 (pyridoxal), and vitamin B12, and substrates choline, betaine, and methionine all of which are derived from the diet (Mason 2003; Lim et al. 2007; Rubio-Aliaga et al. 2011; Kasperzyk et al. 2011; Dominguez-Salas et al. 2013; Wadhvani et al. 2013). The quality of the diet has been shown to dramatically alter the levels of SAM, SAH, and homocysteine (Poirier et al. 2001a; Dominguez-Salas et al. 2013) and subsequently methylation potential for all substrates including DNA. Variation in genes involved in the one-carbon pathway, which also alter methylation potential, has been intensively studied (Hazra et al. 2007; Wernimont et al. 2011; Kasperzyk et al. 2011; Molloy 2012).

Gene–environment interactions that alter one-carbon pathway production of SAM have both short-term and long-term effects since DNA methylation and other SAM-dependent products contribute to the regulation of gene transcription (Klose and Zhang 2007; Gibney and Nolan 2010; Pu et al. 2010). A growing body of evidence from laboratory animal (Wolff et al. 1998; Cooney et al. 2002; Dolinoy et al. 2006; Waterland et al. 2007) and human studies (Heijmans et al. 2008; Gertz et al. 2011) indicates that changes in DNA methylation in utero or during critical developmental windows contribute to the developmental origins of adult diseases (Barker et al. 1993; Godfrey and Barker 1995; Gluckman et al. 2009; Hochberg et al. 2011). While many studies have understandably focused on methylation potential during fetal and perinatal periods (McGowan and Szyf 2010; Laurent et al. 2010; Beyan et al. 2012), fewer investigations have analyzed methyl substrate pools during childhood and adolescence.

The USDA Delta Nutrition Intervention Research Initiative (Delta NIRI), renamed in 2010 to the Delta Obesity Prevention Research Unit, initiated a community-based participatory research (CBPR) program (O’Fallon et al. 2000) to develop interventions to address childhood obesity in communities of the Lower Mississippi Delta (LMD) region. This project was conducted in conjunction with a Freedom School Program held at the Boys, Girls, Adults Community Development Center (BGACDC). BGACDC has managed and operated a summer day camp for children and teens for 30 years. CBPR simultaneously conducts research while applying existing scientific knowledge to improve prevention practices and healthcare among the participants and their community (Israel et al. 2005; McCabe-Sellers et al. 2008). The USDA program has a long history in the Delta region as evidenced by the many

J. Oguntimein · L. Taylor
Shepherd Program for the Interdisciplinary Study of Poverty and Human Capability, Washington and Lee University, Lexington, VA, USA

J. Oguntimein
Medical School, Drexel University, Philadelphia, PA, USA

L. Taylor
Emory School of Public Health, Atlanta, GA, USA

J. Kaput (✉)
Systems Nutrition and Health Unit, Nestle Institute of Health Sciences, Innovation Square, EPFL Campus, 1015 Lausanne, Switzerland
e-mail: james.kaput@rd.nestle.com

consultations held with the community (Harrison 1997; Smith et al. 1999; Yadrack et al. 2001; Horton et al. 2004; Ndirangu et al. 2007, 2008), assessments of the food insecurity in the region (Stuff et al. 2004a, b; Champagne et al. 2007), local diets (Champagne et al. 2004; McCabe-Sellers et al. 2007; McGee et al. 2008), and the creation of a regional food frequency questionnaire (Tucker et al. 2005). The assessments for efficacy had not previously analyzed biochemical markers to test whether the intervention programs were successful in altering physiological processes. The USDA and the FDA researchers expanded the scope of the CBPR program in consultation with community leaders in Phillips County to include physiological measures from blood analyzed with omics and genomic methodologies, physical activity monitoring, and dietary intakes (McCabe-Sellers et al. 2008).

Levels of 9 plasma metabolites, erythrocyte metabolites *S*-adenosylmethionine and *S*-adenosylhomocysteine, 1,129 plasma proteins, and 1M single-nucleotide polymorphisms were measured. Associations were found between the SAM/SAH and gene variants and between levels of metabolites and plasma proteins, which may provide additional biomarkers of vitamin status. Two SAM/SAH metabolic groups were found by cluster analysis. The clusters significantly differed in terms of multiple dietary variables as well as in genotype within 25 genes involved in SAM/SAH metabolism.

Methods

Study design

This study is an observational *n*-of-1 design, and data were aggregated for population-level statistical analysis (Nikles et al. 2011), to assess levels of metabolites and proteins associated with each other, with dietary intake, and with genetic makeup. Participants in the research studies conducted in the Marvell (AR) School District in 2009 and 2010 were recruited during the youth summer day program at the Boys, Girls, Adults Community Development Center (BGACDC). The camp was held at sites in Marvell and Elaine (Arkansas) and consisted of structured and unstructured physical activities, reading, leadership, and other enrichment activities using material and training from the Children's Defense Fund's Freedom School curriculum (<http://www.childrensdefense.org/programs-campaigns/freedom-schools/>). Children arrived by 7:30 a.m. and the day ended between 3 and 4 p.m. The FDA's Research Involving Human Subjects Committee (RIHSC) and the University of Arkansas for Medical Sciences (UAMS) Institutional Review Board (IRB) approved this research protocol.

Breakfast, lunch, and two healthy snacks were provided per day for 5 weeks of the summer day camp. The snacks offered were fruits and vegetables. Reduced fat milk and water rather than sweetened fruit drinks or other sweetened soft drinks were also provided. Menus were developed in accordance with USDA guidelines for healthy meals for children/teens ages 6–14, and the same foods were offered in both years of the camps although the quantity of food intake per participant was not monitored at the camp.

The assessments were conducted before the beginning of the camp (baseline), at the end of 5 weeks of the camp, and 1 month after camp ended (post-camp). Thirty-six participants were recruited in the first year, and 19 completed all three assessments. In the second year, 72 participants enrolled and a total of 42 completed all three assessments. Hence, data were available for 105 individuals at 2 baseline measures, 72 at the end of the summer day camp, and 61 who had at least three assessments in the 2-year study. Results for the three assessments for each of 2 years are reported, and 15 individuals participated in both years. All participants were healthy African-American children and adolescents. None of the participants were taking prescribed medicines, nor did they have overt malnutrition, suffered from active infection, or any known genetic disease that could alter metabolism. Some participants voluntarily left the study between assessments, and no adverse events were recorded.

Assessments at baseline (time point 1), end of camp (time point 2), and post-camp (time point 3)

Blood was sampled from participants after an overnight fast at each time point. Blood (3 ml) was collected in purple top EDTA vacutainer tubes, kept on ice, and centrifuged within 0.5 h of collection. After centrifugation, plasma, buffy coat, and red blood cells (erythrocytes) were separated, frozen, and stored at -80°C . A second 3-ml blood sample for DNA extraction was collected in a PAX gene tube (Qiagen 761115) at time point 1.

Height and weight were measured according to previously used training protocol modified from Lohman et al. (1998). Body mass index (BMI) was calculated and compared to percentile references of *Centers for Disease Control and Prevention*—CDC 2000 curves (Kuczmarski et al. 2002): 5th percentile \leq BMI $<$ 85th percentile (normal weight), 85th percentile \leq BMI $<$ 95th percentile (overweight), and BMI \geq 95th percentile (obesity) (WHO Expert Committee 1995).

Dietary intake

Twenty-four-hour dietary recall interviews were done using the USDA automated multiple-pass method (AMPM)

for all children and adolescents at each time point with the assistance of a parent/guardian (Tucker et al. 2005). The FNDDS—USDA Food and Nutrient Database for Dietary Studies software [version 2.0 (2006) and version 3.0 (2008)—Beltsville, MD: Agricultural Research Service, Food Surveys Research Group] was used for energy and nutrient intake analyses (http://www.ars.usda.gov/SP2UserFiles/Place/12355000/pdf/fndds_doc.pdf).

To measure compliance with dietary guidance, the Healthy Eating Index [HEI (Guenther et al. 2007)] was also analyzed. The HEI has a maximum score of 100 with intakes of (1) total fruit (TF), (2) whole fruit (WF), (3) total vegetables (TV), (4) dark green and orange vegetables and legumes (DGOVL), (5) total grains (TG), (6) whole grains (WG) rated to five as maximum in each category, (7) milk, (8) meat and beans, (9) oils, (10) saturated fat and (11) sodium rated to a maximum of 10 in each category, and (12) solid fat, alcohol, and added sugar (SoFAAS) rated to a maximum of 20. By convention, a total score of more than 80 was considered “good,” scores of 51–80 indicated “needs improvement,” and scores of less than 51 were considered “poor” (Guenther et al. 2007; Miller et al. 2011).

Metabolites, proteomic, and genomic analysis

Metabolites

All samples from each year at each time point were randomized at time of analysis. Plasma Hcy was analyzed on HPLC, and erythrocyte SAM and SAH were analyzed on UPLC as reported previously (Wise et al. 1997). Lipid-soluble and water-soluble vitamins were determined separately using UPLC/MS/MS coupled to mass spectrometers (MS) as reported in Morine, Monteiro et al. (submitted). Complete metabolite analysis was done on 105 samples from both years at time point 1, 72 samples from both years at time point 2, and 61 samples from both years at time point 3. The time point and data points (i.e., samples) available for statistical analysis are reported in each table. Statistical analysis was done using age groups (<9 and ≥9 years of age) based on USDA guidelines for dietary intakes for ages 4–8 and 9–13 (<http://www.cnpp.usda.gov/Publications/DietaryGuidelines/2010/PolicyDoc/Appendices.pdf>).

Proteomics

The plasma proteome was analyzed in 110 samples from 6 different time points from the 2 years of sample collection. However, data from 61 participants at time point 1 (the same participants who completed three assessments) were used in subsequent statistical and computational analyses

due to missing samples at time points 2 and 3. Somaligic Inc. (Boulder, CO) performed all proteomic assessments and was blinded to the clinical characteristics of participants in this study. Samples were analyzed as previously described (Gold 1995; Brody and Gold 2000; Gold et al. 2010; Ostroff et al. 2010; Brody et al. 2012).

Genomic analysis

Blood for DNA isolation was collected in a 3-ml PAX gene tube. DNA was isolated as per the manufacturer’s protocol (<http://tinyurl.com/ot2ovuc>). Whole-genome genotyping was done with the HumanOmni1-Quad, version 1.0 kits (Illumina, San Diego, CA) following the manufacturer’s protocol as described in Morine, Monteiro et al. (submitted).

Statistical analysis

All statistical analyses were performed with the use of the R programming language version 3.0.1 or SPSS version 15.0[®]. Unless specified otherwise, only those participants who had completed three questionnaires ($n = 61$) were included for ANCOVA (SPSS), adjusted by age, sex, location, year, and energy intake. Nutrient intakes were also adjusted for energy intake. Statistical differences in nutritional intake variables and metabolite levels between SAM/SAH clusters were determined with analysis of covariance (SPSS-ANCOVA), adjusting for age and sex with a significance threshold set at 0.05. To ascertain whether differences existed within participants’ metabolite levels and Healthy Eating Index at baseline (first time point) and at the end of the camp (second time point), paired t tests were performed for variables with normal distribution and Wilcoxon method for those with non-normal distributions using SPSS. The general linear model for repeated measures was also used in longitudinal analysis to adjust for confounding variables.

While 1M SNPs were measured in the study, analysis focused on SNPs within genes with known functional association with the metabolites measured in the study. This was accomplished by mining all genes from the MetaCore database (version 6.10 build 31731) with a direct functional connection to organic micronutrients. This resulted in 275 unique genes (S2_Gene_Annotate_Net), which were represented across 9515 SNPs on the SNP array. These 9515 SNPs were further filtered to remove those with a minor allele frequency less than 0.1, and those that significantly deviated from Hardy–Weinberg equilibrium, resulting in a final starting set of 4122 SNPs. Significant correlations between genotype in these SNPs and SAM/SAH levels (averaged across three time points) were assessed in each SNP using generalized estimating

equations (GEE), as implemented in the *geese* function in the *geepack* R library (Højsgaard et al. 2006). Genetic analysis used one genotyping data set for each individual (16 individuals attended both years of the camp). In fitting the GEE models, SAM/SAH was modeled as a function of genotype at each SNP locus, adjusted by age, gender, mean total grain intake, and controlling for sibling relationships among the participants (the latter being included as a background correlation structure in the GEE models), which were 9 pairs and 1 trio. A significance threshold was set at 0.1 after correcting for multiple testing using the Benjamini and Hochberg method (Benjamini and Hochberg 1995).

Proteomic data were only comprehensively collected at the first time point, and thus longitudinal analysis was not performed on this data set. The presence of outlier samples in many of the plasma proteins necessitated the use of robust linear regression [using the *lmFit* and *eBayes* functions from the *limma* R library (Smyth 2005)] to identify proteins associated with age group, sex, and body mass index (BMI). Fitted linear models controlled for age, sex, year, and location, and a significance threshold was set at 0.1 after correcting for multiple testing using the Benjamini and Hochberg method (Benjamini and Hochberg 1995). Associations between proteins and plasma vitamins, SAM, SAH, and SAM/SAH, were identified using sparse partial least squares, as implemented in the *sPLS* function in the *mixOmics* R library (Lê Cao et al. 2009). While robust regression was used to identify proteins associated with single outcome variables, *sPLS* allowed for broader analysis of correlation patterns between multiple proteins and multiple metabolites. As *sPLS* was used for exploratory analysis of data set covariance patterns, variable selection was performed by keeping the top 50 variables in both dimensions, i.e., a lasso penalty was imposed that shrank all but 50 variable weights in each dimension (Kim-Anh et al. 2008).

Data mining

Proteins correlated with age, sex, or metabolites were annotated based primarily on manual curation using Online Mendelian Inheritance in Man (OMIM—<http://www.omim.org/>) or published reports. Functional analysis of proteins with high similarity in hierarchical clustering dendrograms (“branches” of dendrograms displayed in Figures) was performed using the MetaCore™ system version 6.15 build 62452 (Thomson Reuters GeneGO) using the Autoanalyze algorithm. Although small gene clusters are not ideal for drawing broader biological conclusions from overrepresentation analysis, we used functional overrepresentation analysis to determine whether the genes that clustered together in hierarchical clustering also

Table 1 Demographic and anthropometric data

| Parameters | Total count (%) | <i>p</i> value |
|--|-----------------|----------------|
| Age (years) | | |
| 6–8 & 11 months 31 days | 32 (30.5) | <0.001 |
| 9–14 | 73 (69.5) | |
| Gender | | |
| Male | 42 (40) | 0.003 |
| Female | 63 (60) | |
| Location | | |
| Marvell | 52 (49.5) | 0.89 |
| Elaine | 53 (50.5) | |
| Year | | |
| 2009 | 35 (33.3) | <0.001 |
| 2010 | 70 (66.7) | |
| Nutritional status | | |
| Body mass index | | |
| Underweight: <5th percentile | 0 (0) | < 0.001 |
| Healthy weight: ≥5th to <85th percentile | 63 (60) | |
| Overweight: ≥85th to <95th percentile | 14 (13.3) | |
| Obesity: ≥95th percentile | 28 (26.7) | |

tended to share common functional roles (as evidenced by significantly overrepresented pathways). We observed that SNPs/genes that were members of a cluster often mapped to different regions of the same chromosome or different chromosomes (Fig. 2). We extended the analysis to protein clusters (Fig. 3) for the same reason, to determine whether the proteins that clustered were functionally related. While all such analysis is limited because of knowledge represented in databases (i.e., publication bias), the use of this strategy expands the single-gene/protein analysis to functional pathways. We provide both single-gene/protein annotations and the functional pathway data (see tabs in Supplementary EXCEL files for each figure).

Results

Demographic and anthropometric data

A total of 108 participants attending the summer day camp in 2 years were initially recruited for the baseline analysis, and samples from 105 were analyzed at baseline. Nutritional status based on CDC 2000 Body Mass Index classification showed a high prevalence of overweight and obese participants in this population: 60 % were considered healthy weight, 13.3 % were overweight, and 26.7 % were obese (Table 1). The majority of participants were older than 9 years, and more females than males attended the

Table 2 Health Eating Index scores for all components

| Health Eating Index (HEI) | Scores (mean \pm SD) |
|---|---------------------------|
| Total HEI (100 points) | 48.8 \pm 7.2 |
| Total fruit (5 points) | 2.3 \pm 1.1 |
| Whole fruit (5 points) | 2.0 \pm 1.3 |
| Total vegetables (5 points) | 2.2 \pm 1.1 |
| Total dark green, orange vegetables legumes (5 points) | 0.3 \pm 0.6 |
| Total grains (5 points) | 4.2 \pm 0.7 |
| Total whole grains (5 points) | 1.0 \pm 0.8 |
| Milk (10 points) | 5.8 \pm 2.0 |
| Meat and beans (10 points) | 8.1 \pm 1.9 |
| Oils (10 points) | 6.4 \pm 1.8 |
| Saturated fat (10 points) | 4.5 \pm 2.1 |
| Sodium (10 points) | 3.6 \pm 1.8 |
| Solid fats, alcoholic beverages, and added sugars (20 points) | 8.4 \pm 3.2 |

camp. A higher percentage of participants completed all three assessments in 2010 (66.7 %) compared with participants who attended the 2009 summer camp (33.3 %).

Dietary intake assessments

Food intake results were based on the mean of three 24-h recalls done at the assessment periods (one 24-h recall per assessment period) since single food intake measurements are less reliable than repeated measures. Food intake during the camp was not monitored. No significant differences in energy and nutrient intake were observed based on sex. Younger (<9 years old) participants reported eating higher amounts of riboflavin ($p = 0.014$), folate ($p = 0.026$), vitamin B12 ($p = 0.008$), iron ($p = 0.042$), and vitamin D ($p = 0.025$) when compared to the participants who were ≥ 9 year old.

Healthy Eating Index (Guenther et al. 2007) scores were also determined from the 24-h dietary intake data (Table 2). Mean Healthy Eating Index (HEI) score from those who completed three 24-h recalls was compared between age and sex, and no statistically different results were found (data not shown). Total HEI for this population was below 51, which is indicative of a poor diet (McCabe-Sellers et al. 2007). Total dark green and orange vegetables and legumes (TDGOVL), whole fruit (WF), and whole grains (WG) had the worst pattern of scores in participants in our study (Table 2) compared to the 2005 Dietary Guidelines for Americans. The data showed negative associations between BMI and total fruit, whole fruit and milk, as well as a positive association between saturated fat and BMI, although after adjusting for age and sex, only the milk component was correlated with BMI ($p = 0.04$).

In separate longitudinal analyses to assess whether the food provided in the camp had any impact on HEI components, 24-h dietary intake data at baseline (time point 1) versus end of camp (time point 2) were analyzed with paired t test and the Wilcoxon method. Seventy-three participants could be paired in these two time points. Total fruit ($p < 0.001$), whole fruit ($p = 0.001$), and whole grain scores ($p = 0.02$) improved after the camp in both years. No other component presented statistically significant results.

Plasma and erythrocyte metabolite levels

Based on normality of the data, the mean and standard deviation or the median and range of for each plasma metabolite are provided in Table 3. ANCOVA was used to model the effect of sex, age, year, and/or location on plasma vitamin levels. Results of the analysis of individual metabolites and comparison with published literature are provided in Supplement 1. The key results are:

- **Vitamin A.** Levels of vitamin A in this population were significantly above that reported in NHANES [36.4 (35.6–37.2)] (CDC 2012). A negative association was observed with HEI ($r = -0.3$; $p = 0.01$) and a positive correlation between plasma vitamin A levels and homocysteine ($r = 0.51$; $p < 0.001$).
- **Vitamin D.** The population average level of vitamin D in the participants in this study was 21 ng/ml, which is just below the 50th percentile of the 2012 NHANES data (CDC 2012). However, 55 % (39 of 70) had vitamin D plasma levels below 20 ng/ml and one participant had a value below 7 ng/ml. The youngest participants (<9 years old) in the present study had higher mean vitamin D plasma levels compared with older participants ($p = 0.005$).
- **Folate.** The standard cutoff for low plasma folate is <3 ng/mL, which is based on microbiological assays (Raiten and Fisher 1995). In our study using LC/MS methods, 96.3 % ($n = 26$ out of 27) of the participants measured were below this cutoff (Table 4). In addition, younger participants had higher plasma levels when compared to the older participants ($p = 0.025$).
- **Thiamin.** 70 % ($n = 49$ out of 70) of the participants in this study (Table 3) had values lower than the reference value of 1.6 μ g/dl (Lynch and Young 2000). Thiamin plasma levels were positively correlated with intakes of total dark green and orange vegetables and legumes ($r = 0.3$; $p = 0.018$). Additional longitudinal analyses are in Supplement 1.
- **Riboflavin.** Younger participants had higher levels of riboflavin compared with older participants ($p < 0.001$). Riboflavin plasma levels also were higher

- in females when compared to males [0.34 $\mu\text{g}/\text{dl}$ (0.04–3.67) versus 0.18 $\mu\text{g}/\text{dl}$ (0.05–0.69; $p = 0.002$)].
- **Pyridoxine and pyridoxal.** High plasma levels of pyridoxal and pyridoxine were found in these participants (Table 3) compared to the limited number of reports measuring these metabolites in children and teens (Midttun et al. 2005; Footitt et al. 2012). Pyridoxal plasma levels were correlated with total fruit intake ($r = 0.35$; $p = 0.006$). Pyridoxal levels were inversely correlated with homocysteine levels (Pearson: $r = -0.46$; $p < 0.001$).
 - **Vitamin E.** The participants in this study (ages 6–14) had higher levels of vitamin E compared to NHANES 2012 data (Table 3). Plasma vitamin E was not associated with any HEI components. Average vitamin E plasma levels from years 1 and 2 decreased between time point 1 and time point 2 (Supplement 1).
 - **Homocysteine.** One hundred percent of the participants had Hcy levels under 15 $\mu\text{mol}/\text{l}$ (reference values are $<15 \mu\text{mol}/\text{l}$) and corroborate the results found by Pfeiffer et al. (2005). For those who had two or three data points (total of 76), positive correlations were found between Hcy and (1) mean plasma vitamin A (Pearson: $r = 0.51$; $p < 0.001$), (2) mean riboflavin plasma levels (Spearman $r = 0.28$; $p = 0.020$), (3) mean erythrocyte SAM (Pearson $r = 0.59$; $p < 0.001$), and (4) erythrocyte SAM/SAH ratio (Spearman $r = 0.49$; $p < 0.001$). Negative correlations were found between Hcy and (1) mean plasma vitamin E (Spearman: $r = -0.36$; $p = 0.001$) and (2) mean pyridoxal plasma levels (Pearson: $r = -0.46$; $p < 0.001$). After adjustment for age and sex, these metabolites were statistically associated with Hcy. Bates et al. (2007) also found a negative correlation between B6 and Hcy.
 - **Erythrocyte SAM, SAH, and SAM/SAH.** Although SAM and SAH concentrations are often measured from plasma (Poirier et al. 2001c; van Driel et al. 2009), levels of these metabolites are higher in erythrocytes (Poirier et al. 2001b; Smulders et al. 2007; Hirsch et al. 2008) and may be less affected by physiological processes that induce cell death or turnover. The mean SAM/SAH level was determined since the ratio did not change with or without controlling for confounding variables such as improved whole fruit, total fruit, and whole grain intake at time point 2. Mean concentrations of erythrocyte SAM and SAH were 0.97 and 0.88 $\mu\text{mol}/\text{L}$, respectively, in 61 participants with at least three assessments per year. The mean erythrocyte SAM/SAH ratio for participants of this study was also considerably lower (0.98) than the ratios of 2–8 reported in the literature (Poirier et al. 2001a, c; Smulders et al. 2007; Hirsch et al. 2008; Dominguez-Salas et al. 2013). These differences may be due to measurement techniques, age, ancestral background, or environment factors that influence SAM/SAH ratio.

Table 3 Plasma vitamin levels

| Plasma metabolite | This study (mean \pm SD) | Reference value median (range) | References |
|---|----------------------------|---|--|
| Vitamin A ($\mu\text{g}/\text{dl}$) | 64.4 \pm 28.9 | 36.4 (35.6–37.2) | (CDC 2012) |
| Vitamin E (mg/dl) | 2.1 \pm 1.5 | 0.82 (0.80–0.84) | (CDC 2012) |
| Vitamin D (ng/ml) | 21.7 \pm 8.8 | 25.5 (24.6–26.4) ^a | (CDC 2012) |
| Thiamin ($\mu\text{g}/\text{dl}$) | 0.6 (0.1–10.6) | 1.6–4.8 ^b | (Leite 2011) |
| Pyridoxal (ng/ml) | 9.3 (0.07–27.3) | 0.77–3.06 ^c 1.7 (0.96–4.78) ^d | (Footitt et al. 2012) (Midttun et al. 2005) |
| Pyridoxine (ng/ml) | 4.7 (1.3–36.9) | Nd–0.15 ^c | (Footitt et al. 2012) |
| Folate (ng/ml) | 1.0 (0.25–16.1) | 6.1 (15.6–16.6) | (CDC 2012) |
| Riboflavin ($\mu\text{g}/\text{dl}$) | 0.2 (0.04–3.7) | 0.78 \pm 0.11 ^c 0.39 (0.18–1.44) ^d | (Capo-chichi et al. 2000) (Midttun et al. 2005) |
| Homocysteine ($\mu\text{mol}/\text{l}$) | 7.0 \pm 3.0 | 7.14 (7.04–7.24) ^f | (CDC 2012) |

Mean and standard deviation were performed for variables with normal distribution and median and range for variables with non-normal distributions

^a 50th percentile values (CDC 2012)

^b Adults

^c Ages 4.3–16

^d Adults

^e Study participants from Benin

^f Age 2

Table 4 Metabolite concentrations and nutrient intakes between SAM/SAH clusters

| | Cluster 1 | Cluster 2 | <i>p</i> value |
|---------------------------------|---|---|----------------|
| Metabolite | | | |
| SAM/SAH | 1.81 ± 0.48 1.61 (1.41–2.95) (<i>n</i> = 10) | 0.81 ± 0.28 0.79 (0.31–1.31) (<i>n</i> = 51) | <0.001 |
| SAM (nmol/ml) | 1.44 ± 0.31 1.34 (1.1–1.95) (<i>n</i> = 10) | 0.77 ± 0.28 0.73 (0.34–1.59) (<i>n</i> = 51) | <0.001 |
| SAH (nmol/ml) | 0.88 ± 0.14 0.87 (0.63–1.15) (<i>n</i> = 10) | 0.96 ± 0.11 0.95 (0.7–1.31) (<i>n</i> = 51) | 0.022 |
| Homocysteine (μmol/l) | 11.1 ± 4.0 11.9 (5.6–17.5) (<i>n</i> = 10) | 7.3 ± 2.8 6.2 (3.5–14) (<i>n</i> = 51) | 0.053 |
| Vitamin A (μg/dl) | 98 ± 35 93 (36–156) (<i>n</i> = 8) | 61 ± 27 56 (11.5–134) (<i>n</i> = 42) | 0.038 |
| Intake variable | | | |
| Energy intake/BW (kcal/kg) | 67.1 ± 29.5 64.4 (22–112) (<i>n</i> = 10) | 46.0 ± 18.7 41.4 (16.0–102.9) (<i>n</i> = 51) | 0.009 |
| Thiamin intake (mg/day) | 1.50 ± 0.59 1.4 (0.83–2.5) (<i>n</i> = 10) | 1.30 ± 0.40 1.3 (0.59–2.2) (<i>n</i> = 51) | 0.013 |
| Iron intake (mg/day) | 12.9 ± 3.4 11.6 (9.5–19) (<i>n</i> = 10) | 10.9 ± 3.4 10.5 (5.4–23.2) (<i>n</i> = 51) | 0.041 |
| β-Carotene intake (mcg/d) | 1,539 ± 2,718 456 (139–8,676) (<i>n</i> = 10) | 787 ± 1,072 381 (30–4,339) (<i>n</i> = 51) | 0.027 |
| Vitamin A intake (RE) (mcg/day) | 518 ± 340 438(200–1,328) (<i>n</i> = 10) | 420 ± 170 370 (125–885) (<i>n</i> = 51) | 0.034 |

Analyses as per methods. Differences in other metabolites between clusters did not reach statistical significance

SAM/SAH ratio: groups and diet influence

The distribution of mean SAM/SAH ratios in 61 participants was further analyzed using the K-means clustering algorithm. Significantly distinct clusters of SAM/SAH (center to center distance 0.991, Fig. 1) were found at $k = 2$, but not at $k = 3, 4$, or 5. Cluster 1 (C1) consisted of

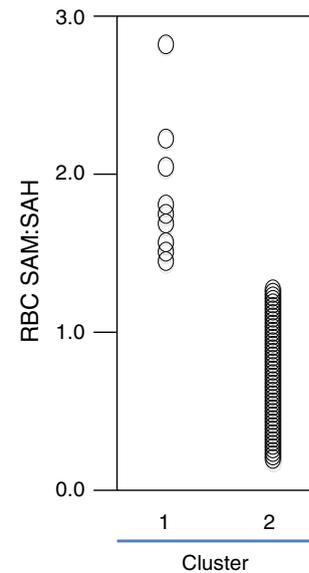


Fig. 1 K means cluster results of SAM/SAH ratios. SAM/SAH was analyzed using SPSS K-means clustering program. Only $K = 2$ yielded significant differences between the groups with center cluster 1 to cluster 2 distance = 0.991

10 participants with higher values for SAM and SAM/SAH ratio compared with 51 participants in cluster 2 (C2) (Fig. 1; Table 4). The proportion of males to females differed ($p = 0.021$) between C1 (7/3, respectively) versus C2 (16/35, respectively), a finding consistent with gender differences in SAM/SAH ratios found in adults (Poirier et al. 2001a; Smulders et al. 2007). C1 had the youngest participants (age 8.2 ± 1.39 vs. 9.7 ± 2.3 ; $p = 0.049$) and more participants from the first year of the study. C2 had more participants from the second year of the study ($p = 0.004$). Body mass index and weight (normal, overweight, obese) were similar between clusters (not shown).

Plasma vitamin A levels also differed significantly between C1 and C2 (Table 4) with high vitamin A correlated with high SAM/SAH ($r = 0.30$; $p = 0.02$). Differences in homocysteine levels between the two groups approached significance ($p = 0.053$) (Table 4) with higher homocysteine (but still normal) plasma levels and higher SAM/SAH ($r = 0.42$, $p = 0.001$). Pyridoxal plasma levels ($r = -0.37$, $p = 0.003$) and vitamin E plasma levels ($r = -0.26$, $p = 0.046$) were negatively correlated with SAM/SAH ratio. Other observed differences in metabolite levels between C1 and C2 were found but they failed to reach significance at $\alpha = 0.05$.

Analysis of the mean 24-h dietary intakes showed that participants in C1 were eating, on average, significantly more vitamin A (retinol equivalents), thiamin, iron, β-carotene, and energy compared with participants in cluster 2 (Table 4). Sixteen individuals attended the camp in both years. Of these, 14 had decreased SAM/SAH ratios

Table 5 Variables in 14 individuals that changed from higher SAM/SAH in year 1 to lower SAM/SAH in year 2

| Variable | Year 1 | Year 2 | p value |
|--------------------------------|----------------|----------------|---------|
| Anthro | | | |
| Height | 139.15 ± 13.08 | 143.86 ± 11.49 | 0.043 |
| Weight | 37.44 ± 13.63 | 43.88 ± 13.63 | 0.016 |
| BMI | 18.75 ± 3.38 | 20.53 ± 4.37 | 0.034 |
| Diet | | | |
| Saturated fat | 24.34 ± 6.28 | 24.58 ± 8.62 | 0.061 |
| Vit B12 | 3.20 ± 1.15 | 4.47 ± 2.15 | 0.052 |
| Plasma metabolites | | | |
| Homocysteine | 11.82 ± 2.27 | 5.68 ± 1.19 | <0.001 |
| Pyridoxine | 1.15 ± 3.23 | 12.33 ± 5.96 | <0.001 |
| Thiamin | 0.37 ± 0.50 | 1.30 ± 0.95 | 0.009 |
| Vitamin A | 88.60 ± 35.42 | 46.76 ± 18.95 | 0.001 |
| Vitamin B2 | 3.20 ± 1.15 | 4.47 ± 2.14 | 0.003 |
| Vitamin D | 24.07 ± 8.69 | 20.46 ± 6.98 | 0.163 |
| Vitamin E | 5.92 ± 183 | 5.76 ± 1.88 | 0.030 |
| Erythrocyte metabolites | | | |
| SAM | 1.08 ± 0.36 | 0.62 ± 0.27 | <0.001 |
| SAH | 0.97 ± 0.16 | 0.99 ± 0.09 | 0.653 |
| SAM/SAH | 1.12 ± 0.33 | 0.65 ± 0.30 | <0.001 |

between year 1 and year 2 (Table 5). In addition, levels of all metabolites except SAH and vitamin D changed statistically in these participants between year 1 and year 2. Vitamin B12 and saturated fat intake changed between year 1 and year 2 but only as a trend. Changes in SAM/SAH ratio due to changes in diet and environment have been noted by others (Dominguez-Salas et al. 2013).

Genotype analysis

While the SAM/SAH ratio correlated with different dietary patterns (Table 4) and with indications of specific nutrients affecting the methylation potential, genetic variation may also contribute to the observed levels in plasma and erythrocyte metabolites. Bottom-up analysis of single-gene variants (e.g., SNPs) is unlikely to explain a complex phenotype such as the relationships between SAM and SAH. Genome-wide association analysis, a top-down approach, typically uses sample sizes much larger than the number of participants in this study. Hence, the genotype association with SAM/SAH ratio in this study was done with a middle-out approach. Middle-out is an emerging approach that uses a predetermined subset of high-dimensional data that are limited to a system of interest

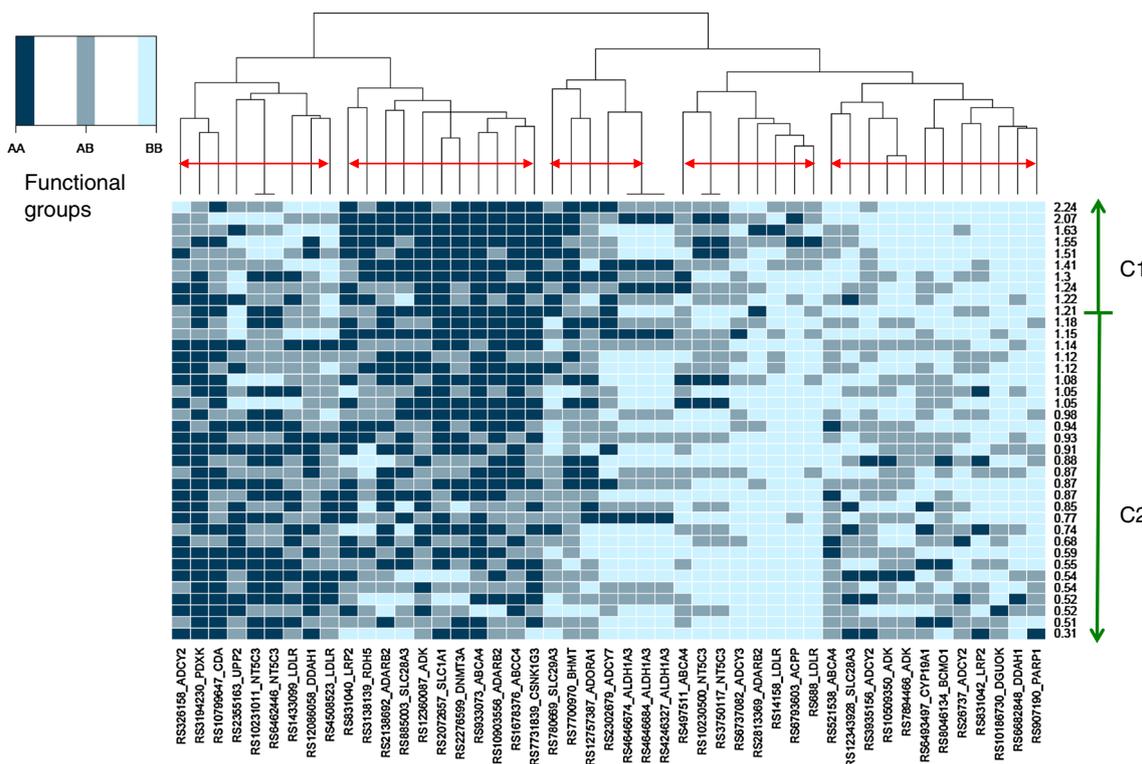


Fig. 2 Heatmap of significant SNPs associated with SAM/SAH ratios. SNPs statistically associated with SAM/SAH ratios (left axis, displayed high SAM/SAH to low) corrected for multiple comparisons

were identified using procedures described in “Methods.” Two-hundred and sixty-seven (267) genes were used for genetic analysis (Supplements 3 and 5)

(Radulescu et al. 2008; de Graaf et al. 2009; Pantelev et al. 2010; Secomb and Pries 2011). Specifically, we matched the genetic system to be analyzed to the plasma and erythrocyte metabolites measured in this study since many are involved in the one-carbon pathway (folate, pyridoxal/pyridoxine, thiamin) or were correlated with SAM/SAH (vitamin A, E, homocysteine). A commercial software and database (GeneGO MetaCore) were used to identify 275 genes in pathways and networks involved in metabolism of the 11 metabolites measured in this study (Fig. 2 and gene list tab in Supplement 2). These are referred to as the micronutrient neighborhood genes.

The genotyping platform used in this study included 9515 SNPs in 268 of these 275 genes. Of the 9515 SNPs, 4122 were selected for analysis based on preprocessing criteria (see “Methods”). Significant correlations between genotype and ratio of SAM/SAH levels were assessed in each SNP using generalized estimating equations (GEE), wherein SAM/SAH was modeled as a function of genotype at each SNP locus, adjusted by age, gender, mean total grain intake, and sibling relationships among the participants (the latter being included as a background correlation structure in the GEE models). The raw ratio of SAM/SAH was used instead of the SAM/SAH cluster memberships due to the higher information content of continuous variables. Resulting *p* values were corrected for multiple testing using the procedure proposed by Benjamini and Hochberg (2000). Forty-six SNPs in 25 genes (annotate tab in Supplement 2) were found to be associated with the SAM/SAH ratio. Two SNPs were not assigned to genes in public databases.

The statistically significant genes identified were involved in organic ion (e.g., *ABCC4*), and other transport systems (e.g., *SCL1A1*, *SCL28A3*, *SCL29A3*) and micronutrient metabolism pathways (e.g., *ALDH1A3*, *BHMT*) including genes involved in vitamin A (*BCMO1*, *RDH5*) and vitamin B6 (*PDXXK*) metabolism. Others have analyzed associations of plasma Hcy levels and Alu and LINE-1 DNA methylation status with 330 polymorphisms in 52 genes directly involved in SAM/SAH metabolism in an elderly population (Wernimont et al. 2011). None of the statistically significant SNPs found in that study overlapped with the study reported here which may be due to differences between genetic makeups, diet intakes, and ages of participants involved in the two studies.

The MetaCore™ functional analysis tools were used to determine gene ontology processes of genes that clustered in branches of the heatmap (red arrows, Fig. 2). The statistically significant SNPs/genes within the major 5 clusters identified cell signaling, energy metabolism, negative insulin regulation, ion and lipid transport, and cell adhesion processes as the main functions of the 25 identified genes (Net tab in Supplement 2). Hence, the middle-out approach

identified genes within a wider SAM/SAH system than defined solely by the one-carbon and methylation pathways.

Proteins correlated with plasma metabolite levels

DNA aptamer technology (Zichi et al. 2008; Gold et al. 2010) was used to measure 1,129 proteins in plasma of 61 participants at baseline. Sparse partial least squares analysis identified 100 protein aptamers (99 unique proteins) associated with baseline plasma Hcy, vitamin A, riboflavin, vitamin E, thiamin, and pyridoxal and erythrocyte SAM, SAH, and SAM/SAH levels (Fig. 3, Annotate tab in Supplement 3). Although the Pearson correlation coefficients for each protein–metabolite pair were modest (between -0.5 and $+0.5$), a large number of proteins showed similar correlation coefficients with Hcy, vitamin A, and riboflavin and inversely with thiamin and pyridoxal. These proteins participate in a wide variety of metabolic, neuronal, immune, growth, and development processes (Fig. 3), and each is annotated in Supplement 3 (Net tab).

Cluster analysis showed that $\sim 85\%$ of the proteins that strongly correlated with plasma metabolites (as opposed to erythrocyte metabolites) were intracellular proteins released from damaged cells (right branch of heat map). In contrast, 72% of the proteins that are predicted to be in the plasma were in the left branch. The significance of the cluster results will require further studies but may be related to developmental reprogramming of tissues expected for individuals in this age range. Cole et al. (2013) analyzed 4705 proteins using iTRAQ mass spectroscopy methods and identified proteins strongly correlated with vitamin A (retinol binding protein 4), 25-hydroxyvitamin D (vitamin D binding protein), α -tocopherol (apolipoprotein C-III), copper (ceruloplasmin), and selenium (selenoprotein P isoform 1) in Nepalese children ages 6–8. None of these proteins were correlated in the study reported here, a difference that may be due to genetic makeup and environment dissimilarities between the two study locations.

MetaCore™ data mining of proteins with closely correlated coefficients (i.e., with similar correlations and designated by red arrows above the heatmap) identified functional subnetworks by mapping to gene ontology (GO) terms (net tab in Supplement 3). For example, the first branch (NCR2, MAP2K1, HBA/HBB, MAP2K4, MAP2K1) was associated with lower levels of erythrocyte SAM and plasma Hcy and was significantly associated with response to hormone stimulus ($p = 8.992 \times 10^{-48}$) as well as response to organic nitrogen ($p = 1.009 \times 10^{-43}$, net tab in Supplement 3). A second branch (KLK6, HIBADH, NR1D1, NCK1, PDGFC, FGF7, DLL4, LGALS2, PTEN) correlated with low plasma vitamin A and riboflavin and was also associated with responses to hormones and

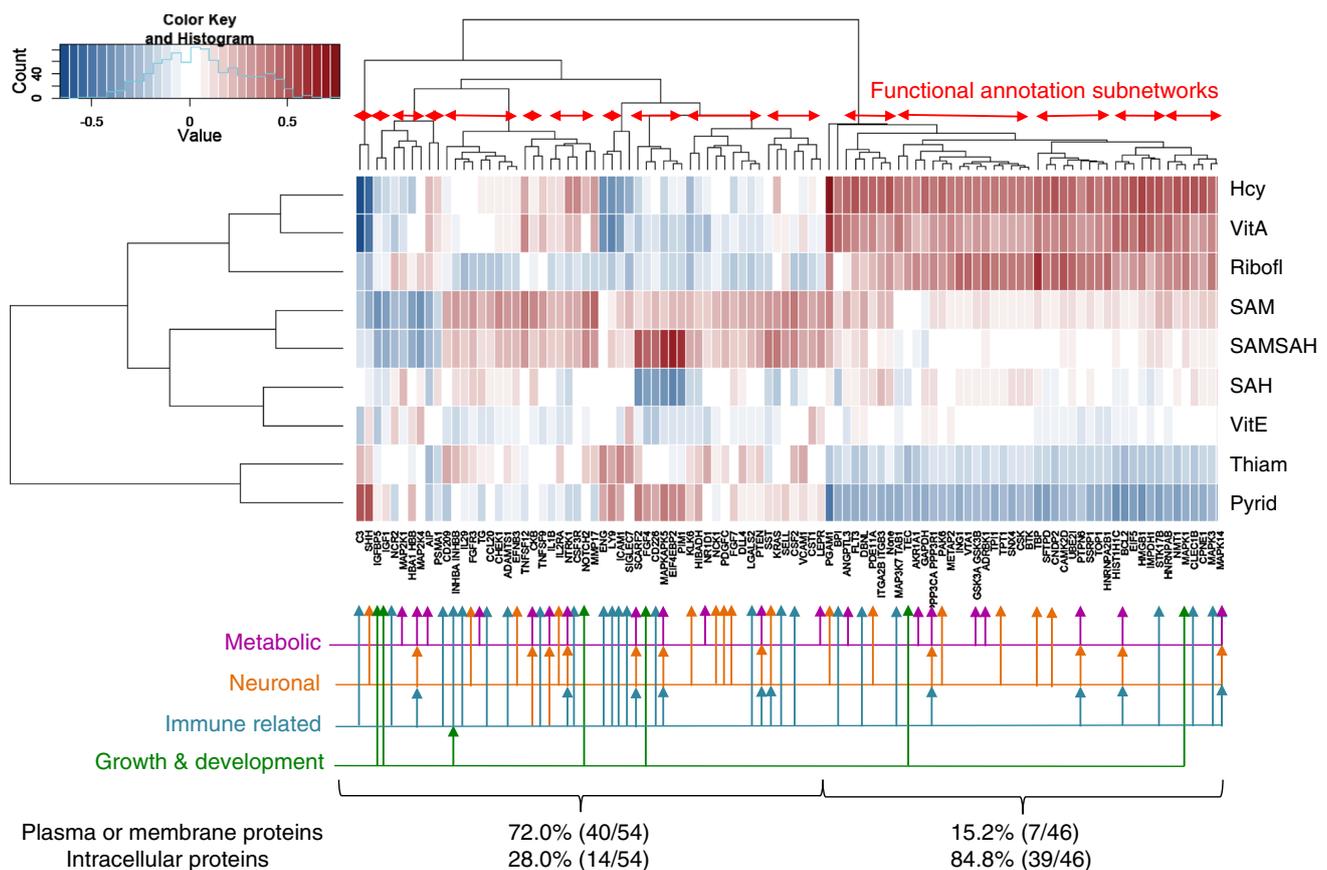


Fig. 3 Heatmap of proteins most strongly correlated with metabolites. DNA aptamer technology was used to analyze 1,129 proteins in plasma of 107 samples from all time points. Analysis was done at Somalogic (Boulder, CO). A complete set of data was available for participants time point 1. See “Methods” for statistical analysis procedures. Annotation of each protein was done with OMIM or

published reports and is provided in Supplement 1. Plasma membrane and soluble proteins refer to proteins expected to be in the plasma. Functional analysis of proteins with similar correlation coefficients (branches marked by red arrows above the heatmap) was performed using GeneGO MetaCore™ programs for identifying gene ontologies (GO) (S2_Gene_Annotate_Net)

organic nitrogen. Results for the remaining branches are provided in Supplement 3. Many individual soluble and membrane proteins have relatively strong positive associations with erythrocyte SAM levels (Fig. 3), consistent with the central role of SAM in methylation reactions in diverse pathways and functional processes. The breadth of biological processes associated with vitamins and metabolites (Hcy, SAM, SAH) is not unexpected since vitamins are cofactors for a large and diverse set of enzymatic functions (Ames et al. 2002).

Proteins correlated with physiological and anthropometric variables

The availability of the quantitative levels of blood proteins provided an opportunity to discover correlations with other physiological or anthropometric variables. Using robust linear regression, fourteen proteins were found to be associated with BMI (Table 6) including proteins involved in appetite regulation (leptin and ghrelin) and inflammatory

processes (CRP, LGALS3BP, CD70, and APCS). Leptin levels have been positively associated (Chu et al. 2001; Hansen et al. 2010) and ghrelin is negatively associated with BMI (Stylianou et al. 2007) in adolescents. However, ghrelin levels are high before meals (Moran 2009), and the samples in this study were taken in the fasting state, suggesting that the acute appetite stimulus was highest in the high-BMI individuals in the study.

Correlation analysis identified 49 proteins associated with age group and 30 proteins correlated with sex (Supplement 5). Mapping sets of proteins to function revealed a difference between females and males in glucose transport (through FASLG, PGK1, IDS) and insulin metabolic processes (through IDE). Protein levels indicate that these subsystems might be more active in males than females, a finding consistent with different metabolic trajectories induced by puberty (DiVall and Radovick 2009; Lewis and Lee 2009) including processes involved in lipid metabolism and fat deposition [e.g., (Staiano and Katzmarzyk 2012; Shapira 2013)]. Due to space limitations, the

Table 6 Proteins associated with BMI

| Protein_ID | Protein | OMIM ID | logFC | <i>p</i> value | Adj <i>p</i> val | Function |
|------------|---|---------|----------|----------------|------------------|--|
| LEP | Leptin | 164160 | 1,447.51 | 8.15E−12 | 9.20E−09 | Role in the regulation of body weight by inhibiting food intake and stimulating energy expenditure |
| GHR | Ghrelin | 605353 | 43.80 | 3.63E−08 | 2.05E−05 | Involved in regulating growth hormone release. Derived from a preproghrelin, which also generates a second peptide called obestatin involved in satiety and decreased food intake |
| CD70 | TNF ligand member 7 | 602840 | 37.96 | 3.79E−05 | 0.010705 | Potent activator of T cells in vivo |
| CRP | C-reactive protein | 123260 | 1,894.64 | 1.18E−05 | 0.004431 | Marker of low-grade inflammation—may have a role in the pathogenesis of atherosclerotic lesions in humans |
| APCS | Amyloid P component | 104770 | 751.46 | 5.37E−05 | 0.012127 | Participates in pathogenesis of amyloidosis in vivo |
| LGALS4 | Lectin, galactoside binding | 602518 | 9.278 | 0.000398 | 0.053566 | Antibacterial lysis factor expressed in the intestine |
| CCDC80 | Upregulated in Bombesin-deficient mice | 608298 | 37.11 | 0.000494 | 0.053566 | May play a role in energy metabolism and the regulation of body weight |
| F9 | Coagulation factor IX | 300746 | 122.52 | 0.000569 | 0.053566 | Factor IXa activates factor X as part of an intrinsic activating complex that also consists of factor VIIIa |
| DSC3 | Desmocollin 3 | 600271 | 7.18 | 0.000523 | 0.053566 | Intercellular adhesion molecules belonging to the cadherin superfamily |
| APOE | Apolipoprotein E | 107741 | 663.39 | 0.000519 | 0.053566 | Apolipoprotein E mediates the presentation of serum-borne lipid antigens. Association with multiple chronic diseases |
| LGALS3BP | Lectin, galactoside binding, soluble, binding protein | 600626 | 30.36 | 0.000215 | 0.040546 | Stimulates the host immune system via induction of IL2 and possibly other cytokines |
| HSD17B1 | 17-beta-hydroxysteroid dehydrogenase | 109684 | 21.05 | 0.000616 | 0.053566 | Interconversion of estrone (E1) and estradiol (E2) as well as the interconversion of androstenedione and testosterone |
| PDE11A | Phosphodiesterase 11A | 604961 | 27.00 | 0.000575 | 0.053566 | Catalyze the hydrolysis of 3',5'-cyclic nucleotides to the nucleoside 5'-monophosphates. Pde11a knockout mice show subtle behavioral deficits, hyperactivity in an open field, impaired social odor recognition memory, and social avoidance |
| RTN4R | Reticulon 4 receptor | 605566 | 28.38 | 0.001119 | 0.090252 | Central roles in limiting axonal regeneration in CNS injury |

OMIM = Online Mendelian Inheritance in Man unique ID

LogFC = log of the fluorescence, a measure of relative abundance (see “Methods”)

Adj *p* value = corrected for multiple comparisons

figures and analysis of these proteins are provided in Supplement 4.

Discussion

Experimental design and main results

Data from the sequencing of the human genome were to enable a more comprehensive analysis of physiology, responses of individuals to drugs, nutrition, and lifestyle factors, and personalized healthcare. However, many

human experiments continue to use case–control designs, which implicitly assume that individuals randomized to the case (intervention) group and to the control group are genetically identical with similar, if not identical, diets and lifestyles. The HapMap (The International HapMap 2005; Frazer et al. 2007), Human Genetic Diversity (<http://www.hagsc.org/hgdp/>), 1000 Genomes Projects (Durbin et al. 2010), and every published whole-genome sequence [references in (Olson 2012)] demonstrated that individuals are genetically unique. Metabolomic, proteomic, and clinical data also demonstrate biochemical individuality (Williams 1956; Robinette et al. 2012).

The experimental design employed in the observational study reported here provided the same dietary and physical environment to all participants in a community-based summer day program for 6–14 year olds. This design accounts for heterogeneity in genetic makeup, individuality in metabolites and proteins, and dietary differences by measuring these parameters and aggregating the data for population-level analysis (this report) and for group level analysis (SAM/SAH cluster in this report and Morine, Monteiro et al. submitted).

Discovery-based methods were used to identify two groups of participants that were statistically different in SAM/SAH ratio. A difference in SAM/SAH levels based on gender was also found consistent with published results of others (Poirier et al. 2001c; Van Hecke et al. 2008). Average intakes of energy, thiamin, iron, β -carotene, and vitamin A (RE) were higher in the participants in the high SAM/SAH group compared with the low SAM/SAH group. Longitudinal analysis of 14 participants who attended 2 years of the summer day camp indicated that saturated fat intake, vitamin B12 intake, and metabolites changes might have contributed to differences in SAM/SAH ratio. Hence, SAM/SAH may be a marker of nutritional status, a conclusion consistent with previous published reports with larger numbers of participants (Poirier et al. 2001a; Barbosa et al. 2008; Dominguez-Salas et al. 2013). Others showed that the SAM/SAH ratio correlated with differences in methylation at metastable epialleles based on season and food availability (Waterland et al. 2010). Changes in epigenetic programming at critical developmental windows such as in utero, early childhood, or during puberty have been associated with developmental plasticity, health, and susceptibility to chronic diseases in adults (Barker et al. 1993; Gluckman et al. 2009; Kussmann et al. 2010). The ability to identify plasma proteins involved in inflammatory or other metabolic processes that respond to nutritional interventions may lead to diets that improve the SAM/SAH ratio and associated physiological processes.

To determine whether genotypic differences were also associated with the SAM/SAH levels, we used a middle-out genetic analysis. Specifically, existing pathway and network knowledge was mined to identify a micronutrient-related neighborhood of 275 genes whose protein products interact with, regulate, or metabolize micronutrients measured in the study. Forty-six SNPs in 25 genes were found to be significantly associated with differences in the SAM/SAH ratio after correction for multiple comparisons. Expanding single-gene [e.g., (Lee et al. 2011)] or single pathway [e.g., (Kiyohara et al. 2006; Kelemen et al. 2008; Wernimont et al. 2011; Signorello et al. 2011)] analysis allows for a broader interrogation of the system interacting with the measured metabolites. The discovery-based

middle-out strategy also avoided a pitfall evident in genome-wide association studies that contain a large multiple-testing burden due to the presence of many SNPs (and hence statistical tests) that may not be core to the study.

In addition, a phenotype may be caused by genetic contributions to many different metabolic processes (Kaput 2008). For example, individuals of the same age or sex had different proteomic profiles (Supplement 4 and 5) and individuals with identical SAM/SAH differed in genotype in one or more of the statistically significant SNPs identified in this study (Fig. 2). The genetic contribution to a complex phenotype is therefore dispersed among many genes and most likely with different population average effect sizes (Peltonen and McKusick 2001). Effect sizes of the same variant at each locus may differ between individuals due to epistatic interactions and gene–environment interactions (Williams 1956; Olson 2012). Any given protein or DNA marker needs to be interpreted in conjunction with other variants in the genome, and the similarity of genotype (and/or metabolite and/or proteomic) patterns may reveal more than a single genetic variant, even when corrected for multiple comparisons and even if the study were highly powered. The effects of diet on SAM/SAH were analyzed separately from effects of genotype on SAM/SAH although the ratio was the common variable in the analysis. The sample size is too small to calculate the interaction term and the effect size for each nutrient–SNP combination.

These data and those of others (Waterland et al. 2010; Dominguez-Salas et al. 2013) are examples for genetic association and metabolomics studies: using metabolite levels without knowledge of usual dietary intakes may result in misclassification of individuals, thereby affecting association analysis. Lack of dietary intake data (and the genetic heterogeneity added in large sample sizes) may be among many reasons for the “missing heritability” (Hardy and Singleton 2009; Manolio et al. 2009; Hebebrand et al. 2010; Ober and Vercelli 2011) in genetic association studies. The interpretation of the differences in metabolite levels was made possible by assessing nutrient intakes, in this case by using 24-h dietary recalls. While dietary assessments have been criticized (Thompson et al. 2010) and are not routinely used in genome-wide studies, any validated intake data are better than having no information or data about environmental conditions.

Plasma metabolite levels—population-based data aggregation

In addition to the main finding of SAM/SAH groups with different dietary intakes, metabolite levels, and genetic differences, data were also aggregated for “population”-level analysis. The average diet of the participants was

classified as poor, which may affect the observed levels of individual plasma metabolites and protein measured in this study relative to studies in populations with different nutritional intakes. For example, total dark green and orange vegetables and legumes (TDGOVL), whole fruit (WF), and whole grains (WG) had the worst pattern of scores in participants in our study compared to the 2005 Dietary Guidelines for Americans. Intake of WG in children and adolescents is not well documented and subject to methodological questions (Newby et al. 2007; Garden et al. 2011); however, studies conducted before (Forshee and Storey 2003) and after (O'Neil et al. 2011) the release of the 2005 Dietary Guidelines for Americans (DGA) indicated that WG consumption is low in participants below 18 years of age. Others have found inverse associations between TDGOVL, WF, WG, and central obesity among adolescents (Bradlee et al. 2010), and overweight girls ate fewer servings of fruit than the non-overweight girls from low socioeconomic status [SES—(Wilson et al. 2009)]. Lower intakes of milk were also found to be associated with being overweight (Wilson et al. 2009) and having a high BMI (Garden et al. 2011). Other studies are focused on adults, making it difficult to compare results (French et al. 1999; Ello-Martin et al. 2007).

Over 50 % of the participants in this study had low levels of vitamin D and 70 % were below the recommended range of thiamin. Metabolite–metabolite interactions were detected since positive correlations were found between Hcy and (1) mean plasma vitamin A, (2) mean riboflavin plasma levels, (3) mean erythrocyte SAM, and (4) erythrocyte SAM/SAH ratio. The correlation between higher levels of vitamin A and higher levels of homocysteine is novel but will require further research to understand its significance. Negative correlations were found between Hcy and (1) mean plasma vitamin E and (2) mean pyridoxal plasma levels. Bates et al. (2007) also found a negative correlation between vitamin B6 and homocysteine. These associations were further analyzed by discovery-based methods and will be reported elsewhere (Morine, Monteiro et al. submitted).

Correlations of plasma proteins with vitamins, BMI, age, and sex

Robust linear regression analysis identified proteins associated with plasma and erythrocyte metabolites, BMI, age, and sex. These analyses were limited since sufficient samples were available only for time point 1 in both years of the study. The 99 plasma proteins were marginally associated with levels of individual or combinations of plasma and erythrocyte metabolites. The results indicate that no one protein could be a marker for a plasma metabolite and specifically for a micronutrient. Additional

studies would be needed to test whether combinations of the proteins associated with metabolite levels may be markers of micronutrient status. Post-analysis data mining indicated the many biological processes that might be influenced by or correlated with changes in plasma levels of these metabolites including basic metabolic pathways, hormonal responses, immune, neuronal, and growth processes.

The prevalence of obesity and overweight in the study participants was 26.7 and 13.3 %, respectively, while the national average in the United States for children ages 6–11 is 33 % overweight and obese (<http://www.cdc.gov/healthyouth/obesity/facts.htm>). Proteomic analysis indicated that leptin and inflammatory markers CRP, LGALS3BP, CD70, and APCS were associated with BMI in this population and in other published reports (Chu et al. 2001; Tam et al. 2010; Hansen et al. 2010). Ghrelin was also positively associated with BMI which differs from the work of others (Stylianou et al. 2007). Since ghrelin levels are high before meals (Moran 2009), the results reported here may be due to sampling in the fasted state. Different combinations of proinflammatory markers have been found in overweight or obese children in studies, e.g., high CRP levels not only associated with obesity but BMI can predict CRP levels [rev in (Tam et al. 2010)]. Additional studies will be needed to determine whether the other proteins correlated with BMI (Table 6) are specific to the gene–environment interactions that occurred in this population.

Forty proteins differed between younger (<9) and older (≥ 9) participants and 30 between males and females. A noteworthy outcome of this study is that proteomic analysis may allow for the determination of biological age as opposed to chronological age since some younger children clustered with the older children and vice versa (Supplement 4). Manual annotation showed that plasma and membrane proteins were involved in neuronal-, immune-, and growth-related processes which associate vitamin homeostasis with basic cellular and physiological processes (Ma et al. 2009; Liu et al. 2013; Swartz et al. 2013; Kedishvili 2013). Age and sex differences in proteins involved glucagon and insulin signaling and glucose transport suggesting that these processes change during aging and by sex. These results need to be tested using metabolomic and proteomic analysis targeted to these processes in other populations with more participants than used in this study. Many of the proteins that differed between males and females would be expected (CGA/FSHB), and others may provide stimulus to examine early changes in neuronal, metabolic, and immune system functions that differ between the sexes. Other studies have also shown differences in proteomic profiles in plasma between males and females [e.g., (Silliman et al. 2013)]. Changes in basic metabolic processes during aging

(Staiano and Katzmarzyk 2012; Shapira 2013) and puberty (DiVall and Radovick 2009; Lewis and Lee 2009) make it challenging for creating dietary recommendations to maintain health.

Community-based participatory research

The translational study described here used community-based participatory research principles, one of which emphasizes “conducting research that is beneficial to the communities involved” (Kannan et al. 2009), a form of socially engaged nutrition science (Beauman et al. 2005; Cannon and Leitzmann 2005; Schubert et al. 2012). Consultations with community members and frequent meetings with community leaders ensured active participation of the community and participants. The “standard” CBPR methods (Israel et al. 2005) were extended (McCabe-Sellers et al. 2008) using a discovery-based analytical approach that classifies participants by similar homeostatic profiles and allows for analyses of individuals. While difficult to quantify, this research study raised the awareness of nutrition in health of participants and their families through the research activities and meetings with the community. Further dialogs with the community are planned to provide a report of the findings of this research to continue the dialog on improving health through nutrition and lifestyle choices.

Limitations and reproducibility

Although the sample size of this study was small, significant results were obtained for diet intakes, metabolite levels, proteins, and genetic associations. Proteomic and genomic data were corrected for multiple comparisons. Nevertheless, human genetic, cultural, and lifestyle (including dietary intakes) variability will make it challenging to replicate experimental results: the specific genotype (see Fig. 2) and diets (averages shown in Table 2) produce different physiological readouts as shown for proteomic analysis (e.g., Supplements 3 and 5). Hence, the results reported here are specific to the genetic makeup of individuals in the study, their dietary patterns at the time of the study, geographical location, built environment, and socioeconomic factors that alter their physiology (Hochberg et al. 2011; United Nations Standing Committee on Nutrition 2012). Nevertheless, the results presented can be integrated into other experimental findings assuming that all reports analyze genetic makeup, diet, and metadata associated with the experimental setting. Progress in nutrition and health research would be more rapid with the development of harmonized protocols that would allow for the integration of high-dimensional data sets from different genetic, cultural, and environmental backgrounds.

Acknowledgments The research team gratefully acknowledges the contributions of the Phillips County community members in the Marvell (AR) School District for their participation in this research. We also are indebted to the Helena Regional Medical Center Phlebotomists for their participation in this research. Members of the Divisions of Personalized Nutrition and Medicine and Systems Biology generously assisted with activities during the assessment periods. Lauren Taylor, Alyshia Phillips, and Lauren Shinholster of the Shepard Poverty Program were interns during the summer day camp. This research was funded by the Division of Personalized Nutrition and Medicine at the U.S. FDA National Center for Toxicological Research (Jefferson AR), the USDA Agricultural Research Service Delta Obesity Prevention Research Unit (Little Rock, AR), and the Nestlé Institute of Health Sciences (Lausanne, Switzerland).

Conflict of interest JK is employed by the Nestle Institute of Health Sciences, a for-profit company. No other authors declared a conflict of interest.

References

- Altug Sen T, Koken R, Narci A, Yilmazer M (2011) Homocysteine and ghrelin link with polycystic ovary syndrome in relation to obesity. *J Pediatr Adolesc Gynecol* 24:211–217. doi:10.1016/j.jpag.2011.02.007
- Ames BN, Elson-Schwab I, Silver EA (2002) High-dose vitamin therapy stimulates variant enzymes with decreased coenzyme binding affinity (increased K(m)): relevance to genetic disease and polymorphisms. *Am J Clin Nutr* 75:616–658
- Barbosa PR, Stabler SP, Trentin R et al (2008) Evaluation of nutritional and genetic determinants of total homocysteine, methylmalonic acid and S-adenosylmethionine/S-adenosylhomocysteine values in Brazilian childbearing-age women. *Clin Chim Acta* 388:139–147. doi:10.1016/j.cca.2007.10.023
- Barker DJ, Gluckman PD, Godfrey KM et al (1993) Fetal nutrition and cardiovascular disease in adult life. *Lancet* 341:938–941
- Bates CJ, Mansoor MA, Gregory J et al (2007) Correlates of plasma homocysteine, cysteine and cysteinyl-glycine in respondents in the British national diet and nutrition survey of young people aged 4–18 years, and a comparison with the survey of people aged 65 years and over. *Br J Nutr* 87:71. doi:10.1079/BJN2001479
- Beauman C et al (2005) The giessen declaration. *Public Health Nutr* 8:783–786
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* 57:289–300
- Benjamini Y, Hochberg Y (2000) The adaptive control of the false discovery rate in multiple hypotheses testing. *J Behav Educ Stat* 25:60–83
- Beyan H, Down TA, Ramagopalan SV et al (2012) Guthrie card methylomics identifies temporally stable epialleles that are present at birth in humans. *Genome Res*. doi:10.1101/gr.134304.111
- Bradlee ML, Singer MR, Qureshi MM, Moore LL (2010) Food group intake and central obesity among children and adolescents in the Third National Health and Nutrition Examination Survey (NHANES III). *Public Health Nutr* 13:797–805. doi:10.1017/S136898009991546
- Brody EN, Gold L (2000) Aptamers as therapeutic and diagnostic agents. *J Biotechnol* 74:5–13
- Brody E, Gold L, Mehan M et al (2012) Life’s simple measures: unlocking the proteome. *J Mol Biol* 422:595–606. doi:10.1016/j.jmb.2012.06.021

- Cannon G, Leitzmann C (2005) The new nutrition science project. *Public Health Nutr* 8:673–694
- Capo-chichi CD, Feillet F, Guéant J et al (2000) Concentrations of riboflavin and related organic acids in children. *Am J Clin Nutr* 71:978–986
- CDC (2012) Second national report on biochemical indicators of diet and nutrition in the U.S. population, p 495. <http://www.cdc.gov/nutritionreport/>
- Champagne CM, Bogle ML, McGee BB et al (2004) Dietary intake in the lower Mississippi delta region: results from the foods of our delta study. *J Am Diet Assoc* 104:199–207. doi:10.1016/j.jada.2003.11.011
- Champagne CM, Casey PH, Connell CL et al (2007) Poverty and food intake in rural America: diet quality is lower in food insecure adults in the Mississippi delta. *J Am Diet Assoc* 107:1886–1894. doi:10.1016/j.jada.2007.08.003
- Chu NF, Stampfer MJ, Spiegelman D et al (2001) Dietary and lifestyle factors in relation to plasma leptin concentrations among normal weight and overweight men. *Int J Obes Relat Metab Disord* 25:106–114
- Cole RN, Ruczinski I, Schulze K et al (2013) The plasma proteome identifies expected and novel proteins correlated with micronutrient status in undernourished Nepalese children. *J Nutr* 142:1540–1548. doi:10.3945/jn.113.175018. However
- Cooney CA, Dave AA, Wolff GL (2002) Maternal methyl supplements in mice affect epigenetic variation and DNA methylation of offspring. *J Nutr* 132:2393S–2400S
- De Graaf AA, Freidig AP, De Roos B et al (2009) Nutritional systems biology modeling: from molecular mechanisms to physiology. *PLoS Comput Biol* 5:e1000554
- DiVall SA, Radovick S (2009) Endocrinology of female puberty. *Curr Opin Endocrinol Diabetes Obes* 16:1–4. doi:10.1097/MED.0b013e3283207937
- Dolinoy DC, Weidman JR, Waterland RA, Jirtle RL (2006) Maternal genistein alters coat color and protects Avy mouse offspring from obesity by modifying the fetal epigenome. *Environ Health Perspect* 114:567–572
- Dominguez-Salas P, Moore SE, Cole D et al (2013) DNA methylation potential: dietary intake and blood concentrations of one-carbon metabolites and cofactors in rural African women. *Am J Clin Nutr* 97:1217–1227. doi:10.3945/ajcn.112.048462. The
- Durbin RM, Abecasis GR, Altshuler DL et al (2010) A map of human genome variation from population-scale sequencing. *Nature* 467:1061–1073. doi:10.1038/nature09534
- Ello-Martin JA, Roe LS, Ledikwe JH et al (2007) Dietary energy density in the treatment of obesity: a year-long trial comparing 2 weight-loss diets. *Am J Clin Nutr* 85:1465–1477
- Finkelstein JD (2007) Metabolic regulatory properties of S-adenosylmethionine and S-adenosylhomocysteine. *Clin Chem Lab Med* 45:1694–1699. doi:10.1515/CCLM.2007.341
- Footitt EJ, Clayton PT, Mills K et al (2012) Measurement of plasma B(6) vitamers profiles in children with inborn errors of vitamin B(6) metabolism using an LC-MS/MS method. *J Inher Metab Dis*. doi:10.1007/s10545-012-9493-y
- Forshee RA, Storey ML (2003) Total beverage consumption and beverage choices among children and adolescents. *Int J Food Sci Nutr* 54:297–307
- Frazer KA, Ballinger DG, Cox DR et al (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449:851–861. doi:10.1038/nature06258
- French SA, Jeffery RW, Murray D (1999) Is dieting good for you? Prevalence, duration and associated weight and behaviour changes for specific weight loss strategies over four years in US adults. *Int J Obes Relat Metab Disord* 23:320–327
- Garden FL, Marks GB, Almqvist C et al (2011) Infant and early childhood dietary predictors of overweight at age 8 years in the CAPS population. *Eur J Clin Nutr* 65:454–462. doi:10.1038/ejcn.2011.7
- Gertz J, Varley KE, Reddy TE et al (2011) Analysis of DNA methylation in a three-generation family reveals widespread genetic influence on epigenetic regulation. *PLoS Genet* 7:e1002228. doi:10.1371/journal.pgen.1002228
- Gibney ER, Nolan CM (2010) Epigenetics and gene expression. *Heredity* 105:4–13. doi:10.1038/hdy.2010.54
- Gluckman PD, Hanson MA, Buklijas T et al (2009) Epigenetic mechanisms that underpin metabolic and cardiovascular diseases. *Nat Rev Endocrinol* 5:401–408. doi:10.1038/nrendo.2009.102
- Godfrey KM, Barker DJ (1995) Maternal nutrition in relation to fetal and placental growth. *Eur J Obstet Gynecol Reprod Biol* 61:15–22
- Gold L (1995) Oligonucleotides as research diagnostic, and therapeutic agents. *J Biol Chem* 270:13581–13584
- Gold L, Ayers D, Bertino J et al (2010) Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS ONE* 5:e15004. doi:10.1371/journal.pone.0015004
- Guenther PM, Reedy J, Krebs-Smith SM, Reeve BB, Basitotis PP (2007) Development and evaluation of the Healthy Eating Index-2005. Technical report, Washington, DC. <http://www.cnpp.usda.gov/publications/hei/hei-2005/hei-2005technicalreport.pdf>
- Guerra-Shinohara EM, Morita OE, Peres S et al (2004) Low ratio of S-adenosylmethionine to S-adenosylhomocysteine is associated with vitamin deficiency in Brazilian pregnant women and newborns. *Am J Clin Nutr* 80:1312–1321
- Hansen D, Dendale P, Beelen M et al (2010) Plasma adipokine and inflammatory marker concentrations are altered in obese, as opposed to non-obese, type 2 diabetes patients. *Eur J Appl Physiol* 109:397–404. doi:10.1007/s00421-010-1362-5
- Hardy J, Singleton A (2009) Genomewide association studies and human disease. *N Engl J Med* 360:1759–1768. doi:10.1056/NEJMra0808700
- Harrison G for the Lower Mississippi Delta Nutrition Intervention Research Consortium (1997) Nutrition and health status in the Lower Mississippi Delta of Arkansas, Louisiana, and Mississippi: a review of existing data. Westat, Delta NIRI Coordinating Center. p 207
- Hazra A, Wu K, Kraft P et al (2007) Twenty-four non-synonymous polymorphisms in the one-carbon metabolic pathway and risk of colorectal adenoma in the Nurses' Health Study. *Carcinogenesis* 28:1510–1519. doi:10.1093/carcin/bgm062
- Hebebrand J, Volckmar AL, Knoll N, Hinney A (2010) Chipping away the “missing heritability”: GIANT steps forward in the molecular elucidation of obesity—but still lots to go. *Obes Facts* 3:294–303. doi:10.1159/000321537
- Heijmans BT, Tobi EW, Stein AD et al (2008) Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci USA* 105:17046–17049. doi:10.1073/pnas.0806560105
- Hirsch S, Ronco AM, Guerrero-Bosagna C et al (2008) Methylation status in healthy subjects with normal and high serum folate concentration. *Nutrition* 24:1103–1109. doi:10.1016/j.nut.2008.05.018
- Hochberg Z, Feil R, Constanica M et al (2011) Child health, developmental plasticity, and epigenetic programming. *Endocr Rev* 32:159–224. doi:10.1210/er.2009-0039
- Højsgaard S, Halekoh U, Yan J (2006) The R package geeppack for generalized estimating equations. *J Stat Softw* 15:1–11
- Horton JC, Bogle ML, Formby B, Forsythe W, Goolsby S, Gossett J, Hyman E, Kramer TR, Lovejoy J, Rasco M, Robbins J, Simon W, Simpson PP, Bogle M, Santell R et al (2004) Self-reported health of residents of the Mississippi delta. *J Heal Care Poor Underserved* 15:645–662

- Israel BA, Eng E, Schulz AJ, Parker EA (2005) Methods in community-based participatory research for health 479
- Kannan S, Schulz A, Israel B et al (2009) A community-based participatory approach computer-generated nutrition feedback reports: the healthy environments partnership. *Prog Community Health Partnersh* 2:41–53. doi:10.1353/cpr.2008.0004.A
- Kaput J (2008) Nutrigenomics research for personalized nutrition and medicine. *Curr Opin Biotechnol* 19:110–120. doi:10.1016/j.copbio.2008.02.005
- Kasperzyk JL, Chang ET, Birmann BM et al (2011) Nutrients and genetic variation involved in one-carbon metabolism and Hodgkin lymphoma risk: a population-based case-control study. *Am J Epidemiol* 174:816–827. doi:10.1093/aje/kwr190
- Kedishvili NY (2013) Enzymology of retinoic acid biosynthesis and degradation. *J Lipid Res* 54:1744–1760. doi:10.1194/jlr.R037028
- Kelemen LE, Sellers TA, Schildkraut JM et al (2008) Genetic variation in the one-carbon transfer pathway and ovarian cancer risk. *Cancer Res* 68:2498–2506. doi:10.1158/0008-5472.CAN-07-5165
- Kim-Anh LC, Debra R, Christèle R-G, Philippe B (2008) A sparse PLS for variable selection when integrating omics data. *Stat Appl Genet Mol Biol*. doi:10.2202/1544-6115.1390
- Kiyohara C, Takayama K, Nakanishi Y (2006) Association of genetic polymorphisms in the base excision repair pathway with lung cancer risk: a meta-analysis. *Lung Cancer* 54:267–283
- Klose RJ, Zhang Y (2007) Regulation of histone methylation by demethylation and demethylation. *Nat Rev Mol Cell Biol* 8:307–318
- Kuczmariski RJ, Ogden CL, Guo SS et al (2002) 2000 CDC growth charts for the United States: methods and development. *Vital Health Stat* 11:1–190
- Kussmann M, Krause L, Siffert W (2010) Nutrigenomics: where are we with genetic and epigenetic markers for disposition and susceptibility? *Nutr Rev* 68(Suppl 1):S38–S47. doi:10.1111/j.1753-4887.2010.00326.x
- Laurent L, Wong E, Li G et al (2010) Dynamic changes in the human methylome during differentiation. *Genome Res* 20:320–331. doi:10.1101/gr.101907.109
- Lê Cao K-A, González I, Déjean S (2009) integrOmics: an R package to unravel relationships between two omics datasets. *Bioinformatics* 25:2855–2856. doi:10.1093/bioinformatics/btp515
- Lee Y-C, Lai C-Q, Ordovas JM, Parnell LD (2011) A database of gene-environment interactions pertaining to blood lipid traits, cardiovascular disease and type 2 diabetes. *J Data Mining Genomics Proteomics* 2:1–8. doi:10.4172/2153-0602.1000106
- Leite HP (2011) Low blood thiamine concentrations in children upon admission to the intensive care unit: risk factors and prognostic significance. *Am J Clin Nutr* 1–3:57–61. doi:10.3945/ajcn.2009.29078.1
- Lennard L (2010) Methyltransferases. In: Guengerich FP (ed) *Bio-transformation*. Comprehensive toxicology, vol 4, chap 21. Elsevier, Oxford, UK
- Lewis K, Lee PA (2009) Endocrinology of male puberty. *Curr Opin Endocrinol Diabetes Obes* 16:5–9. doi:10.1097/MED.0b013e32832029be
- Lim U, Wang SS, Hartge P et al (2007) Gene-nutrient interactions among determinants of folate and one-carbon metabolism on the risk of non-Hodgkin lymphoma: NCI-SEER case-control study. *Blood* 109:3050–3059. doi:10.1182/blood-2006-07-034330
- Linnebank M, Popp J, Smulders Y et al (2010) S-adenosylmethionine is decreased in the cerebrospinal fluid of patients with Alzheimer's disease. *Neurodegener Dis* 7:373–378. doi:10.1159/000309657
- Liu JJ, Prescott J, Giovannucci E et al (2013) One-carbon metabolism factors and leukocyte telomere length. *Am J Clin Nutr* 97:794–799. doi:10.3945/ajcn.112.051557.794
- Lohman TG, Martorell R, Roche AF (1998) *Anthropometric standardization manual*. Human Kinetics, Inc. Champaign, IL, USA
- Lynch PL, Young IS (2000) Determination of thiamine by high-performance liquid chromatography. *J Chromatogr A* 881:267–284
- Ma E, Iwasaki M, Kobayashi M et al (2009) Dietary intake of folate, vitamin B2, vitamin B6, vitamin B12, genetic polymorphism of related enzymes, and risk of breast cancer: a case-control study in Japan. *Nutr Cancer* 61:447–456. doi:10.1080/01635580802610123
- Manolio TA, Collins FS, Cox NJ et al (2009) Finding the missing heritability of complex diseases. *Nature* 461:747–753. doi:10.1038/nature08494
- Mason JB (2003) Biomarkers of nutrient exposure and status in one-carbon (methyl) metabolism. *J Nutr* 133(Suppl):941S–947S
- McCabe-Sellers BJ, Bowman S, Stuff JE et al (2007) Assessment of the diet quality of US adults in the lower Mississippi delta. *Am J Clin Nutr* 86:697–706
- McCabe-Sellers B, Lovera D, Nuss H et al (2008) Personalizing nutrigenomics research through community based participatory research and omics technologies. *OMICS* 12:263–272. doi:10.1089/omi.2008.0041
- McGee BB, Richardson V, Johnson GS et al (2008) Perceptions of factors influencing healthful food consumption behavior in the lower Mississippi delta: focus group findings. *J Nutr Educ Behav* 40:102–109. doi:10.1016/j.jneb.2006.12.013
- McGowan PO, Szyf M (2010) The epigenetics of social adversity in early life: implications for mental health outcomes. *Neurobiol Dis* 39:66–72. doi:10.1016/j.nbd.2009.12.026
- Midttun O, Hustad S, Solheim E et al (2005) Multianalyte quantification of vitamin B6 and B2 species in the nanomolar range in human plasma by liquid chromatography-tandem mass spectrometry. *Clin Chem* 51:1206–1216. doi:10.1373/clinchem.2005.051169
- Miller PE, Mitchell DC, Harala PL et al (2011) Development and evaluation of a method for calculating the Healthy Eating Index-2005 using the nutrition data system for research. *Public Health Nutr* 14:306–313. doi:10.1017/S136898010001655
- Ming Z, Legare DJ, Lutt WW (2009) Obesity, syndrome X, and diabetes: the role of HISS-dependent insulin resistance altered by sucrose, an antioxidant cocktail, and age. *Can J Physiol Pharmacol* 87:873–882. doi:10.1139/Y09-079
- Molloy AM (2012) Genetic aspects of folate metabolism. *Subcell Biochem* 56:105–130. doi:10.1007/978-94-007-2199-9_7
- Moran TH (2009) Gut peptides in the control of food intake. *Int J Obes* 33(Suppl 1):S7–S10. doi:10.1038/ijo.2009.9
- Muller T (2010) Catechol-O-methyltransferase enzyme: cofactor S-adenosyl-L-methionine and related mechanisms. *Int Rev Neurobiol* 95:49–71. doi:10.1016/B978-0-12-381326-8.00004-1
- Ndirangu PH, Yadrack K, West JR, Bogle ML, Avis-Williams A, Santell R, Connell CLM (2007) Conducting needs assessment using the comprehensive participatory planning and evaluation model to develop nutrition and physical activity interventions in a rural community in the Mississippi delta. *Prog Community Health Partnersh* 1:41–48
- Ndirangu M, Yadrack K, Bogle ML, Graham-Kresge S (2008) Community—academia partnerships to promote nutrition in the lower Mississippi delta: community members' perceptions of effectiveness, barriers, and factors related to success. *Health Promot Pract* 9:237–245. doi:10.1177/1524839907300804
- Newby PK, Maras J, Bakun P et al (2007) Intake of whole grains, refined grains, and cereal fiber measured with 7-d diet records and associations with risk factors for chronic disease. *Am J Clin Nutr* 86:1745–1753

- Nikles J, Mitchell GK, Schluter P et al (2011) Aggregating single patient (n-of-1) trials in populations where recruitment and retention was difficult: the case of palliative care. *J Clin Epidemiol* 64:471–480. doi:[10.1016/j.jclinepi.2010.05.009](https://doi.org/10.1016/j.jclinepi.2010.05.009)
- O'Fallon LR, Tyson FL, Dearry A (2000) Successful models of community based participatory research. *Environ Health Perspect* 110:155–159
- O'Neil CE, Fulgoni VL, Nicklas TA (2011) Association of candy consumption with body weight measures, other health risk factors for cardiovascular disease, and diet quality in US children and adolescents: NHANES 1999–2004. *Food Nutr Res* 55:1–12. doi:[10.3402/fnr.v55i0.5794](https://doi.org/10.3402/fnr.v55i0.5794)
- Obeid R, Schadt A, Dillmann U et al (2009) Methylation status and neurodegenerative markers in Parkinson disease. *Clin Chem* 55:1852–1860. doi:[10.1373/clinchem.2009.125021](https://doi.org/10.1373/clinchem.2009.125021)
- Ober C, Vercelli D (2011) Gene-environment interactions in human disease: nuisance or opportunity? *Trends Genet* 27:107–115. doi:[10.1016/j.tig.2010.12.004](https://doi.org/10.1016/j.tig.2010.12.004)
- Olson MV (2012) Human genetic individuality. *Annu Rev Genomics Hum Genet* 13:1–27. doi:[10.1146/annurev-genom-090711-163825](https://doi.org/10.1146/annurev-genom-090711-163825)
- Ostroff R, Foreman T, Keeney TR et al (2010) The stability of the circulating human proteome to variations in sample collection and handling procedures measured with an aptamer-based proteomics array. *J Proteom* 73:649–666. doi:[10.1016/j.jprot.2009.09.004](https://doi.org/10.1016/j.jprot.2009.09.004)
- Pantelev MA, Balandina AN, Lipets EN et al (2010) Task-oriented modular decomposition of biological networks: trigger mechanism in blood coagulation. *Biophys J* 98:1751–1761
- Panza F, Frisardi V, Capurso C et al (2009a) Polyunsaturated fatty acid and S-adenosylmethionine supplementation in predementia syndromes and Alzheimer's disease: a review. *ScientificWorldJournal* 9:373–389. doi:[10.1100/tsw.2009.48](https://doi.org/10.1100/tsw.2009.48)
- Panza F, Frisardi V, Capurso C et al (2009b) Possible role of S-adenosylmethionine, S-adenosylhomocysteine, and polyunsaturated fatty acids in predementia syndromes and Alzheimer's disease. *J Alzheimers Dis* 16:467–470. doi:[10.3233/JAD-2009-1012](https://doi.org/10.3233/JAD-2009-1012)
- Peltonen L, McKusick VA (2001) Genomics and medicine. Dissecting human disease in the postgenomic era. *Science* 291(80):1224–1229
- Pfeiffer CM, Caudill SP, Gunter EW et al (2005) Biochemical indicators of B vitamin status in the US population after folic acid fortification: results from the National Health and Nutrition Examination Survey 1999–2000. *Am J Clin Nutr* 82:442–450
- Poirier LA, Brown AT, Fink LM et al (2001a) Blood S-adenosylmethionine concentrations and lymphocyte methylenetetrahydrofolate reductase activity in diabetes mellitus and diabetic nephropathy. *Metabolism* 50:1014–1018. doi:[10.1053/meta.2001.25655](https://doi.org/10.1053/meta.2001.25655)
- Poirier LA, Wise CK, Delongchamp RR (2001b) Blood determinations of S-adenosylmethionine, S-adenosylhomocysteine, and homocysteine: correlations with diet blood determinations of S-adenosylmethionine, S-adenosylhomocysteine, and homocysteine. *Cancer Epidemiol Biomarkers Prev* 10:649–655
- Poirier LA, Wise CK, Delongchamp RR, Sinha R (2001c) Blood determinations of S-adenosylmethionine, S-adenosylhomocysteine, and homocysteine: correlations with diet. *Cancer Epidemiol Biomarkers Prev* 10:649–655
- Pu S, Turinsky AL, Vlasblom J et al (2010) Expanding the landscape of chromatin modification (CM)-related functional domains and genes in human. *PLoS ONE* 5:e14122. doi:[10.1371/journal.pone.0014122](https://doi.org/10.1371/journal.pone.0014122)
- Radulescu O, Gorban AN, Zinovyev A, Lilienbaum A (2008) Robust simplifications of multiscale biochemical networks. *BMC Syst Biol* 2:86. doi:[10.1186/1752-0509-2-86](https://doi.org/10.1186/1752-0509-2-86)
- Raiten DJ, Fisher KD (1995) Assessment of folate methodology used in the Third National Health and Nutrition Examination Survey (NHANES III, 1988–1994). *J Nutr* 125:1371S–1398S
- Robinette SL, Holmes E, Nicholson JK, Dumas ME (2012) Genetic determinants of metabolism in health and disease: from biochemical genetics to genome-wide associations. *Genome Med* 4:30. doi:[10.1186/gm329](https://doi.org/10.1186/gm329)
- Rubio-Aliaga I, De Roos B, Sailer M et al (2011) Alterations in hepatic one-carbon metabolism and related pathways following a high-fat dietary intervention. *Physiol Genomics* 43:408–416. doi:[10.1152/physiolgenomics.00179.2010](https://doi.org/10.1152/physiolgenomics.00179.2010)
- Schubert L, Gallegos D, Foley W, Harrison C (2012) Re-imagining the “social” in the nutrition sciences. *Public Health Nutr* 15:352–359. doi:[10.1017/S1368980011001297](https://doi.org/10.1017/S1368980011001297)
- Secomb TW, Pries AR (2011) The microcirculation: physiology at the mesoscale. *J Physiol* 589:1047–1052. doi:[10.1113/jphysiol.2010.201541](https://doi.org/10.1113/jphysiol.2010.201541)
- Selley ML (2007) A metabolic link between S-adenosylhomocysteine and polyunsaturated fatty acid metabolism in Alzheimer's disease. *Neurobiol Aging* 28:1834–1839. doi:[10.1016/j.neurobiolaging.2006.08.003](https://doi.org/10.1016/j.neurobiolaging.2006.08.003)
- Shapira N (2013) Women's higher health risks in the obesogenic environment: a gender nutrition approach to metabolic dimorphism with predictive, preventive, and personalised medicine. *EPMA J* 4:1. doi:[10.1186/1878-5085-4-1](https://doi.org/10.1186/1878-5085-4-1)
- Signorello LB, Shi J, Cai Q et al (2011) Common variation in vitamin D pathway genes predicts circulating 25-hydroxyvitamin D Levels among African Americans. *PLoS ONE* 6:e28623. doi:[10.1371/journal.pone.0028623](https://doi.org/10.1371/journal.pone.0028623)
- Silliman CC, Dzieciatkowska M, Moore EE et al (2013) Proteomic analyses of human plasma: venus versus mars. *Transfusioin* 52:417–424. doi:[10.1111/j.1537-2995.2011.03316.x](https://doi.org/10.1111/j.1537-2995.2011.03316.x). *Proteomic*
- Smith J, Lensing S, Horton JA et al (1999) Prevalence of self-reported nutrition-related health problems in the lower Mississippi delta. *Am J Public Health* 89:1418–1421
- Smulders YM, Smith DEC, Kok RM et al (2007) Red blood cell folate vitamers distribution in healthy subjects is determined by the methylenetetrahydrofolate reductase C677T polymorphism and by the total folate status. *J Nutr Biochem* 18:693–699. doi:[10.1016/j.jnutbio.2006.11.010](https://doi.org/10.1016/j.jnutbio.2006.11.010)
- Smyth G (2005) Limma: linear models for microarray data. In: Gentleman R, Carey V, Dudoit S et al (eds) *Bioinformatics and computational biology solutions using R and Bioconductor*. Springer, New York, pp m397–m420
- Staiano AE, Katzmarzyk PT (2012) Ethnic and sex differences in body fat and visceral and subcutaneous adiposity in children and adolescents. *Int J Obes* 36:1261–1269. doi:[10.1038/ijo.2012.95](https://doi.org/10.1038/ijo.2012.95)
- Strain JJ, Dowe L, Ward M et al (2004) B-vitamins, homocysteine metabolism and CVD. *Proc Nutr Soc* 63:597–603
- Stuff JE, Casey PH, Szeto KL et al (2004a) Household food insecurity is associated with adult health status. *J Nutr* 134:2330–2335
- Stuff JE, Horton JA, Bogle ML et al (2004b) High prevalence of food insecurity and hunger in households in the rural lower Mississippi delta. *J Rural Health* 20:173–180
- Stylianou C, Galli-Tsinopoulou A, Farmakiotis D et al (2007) Ghrelin and leptin levels in obese adolescents. Relationship with body fat and insulin resistance. *Hormones* 6:295–303
- Swartz MD, Peterson CB, Lupo PJ et al (2013) Investigating multiple candidate genes and nutrients in the folate metabolism pathway to detect genetic and nutritional risk factors for lung cancer. *PLoS ONE* 8:e53475. doi:[10.1371/journal.pone.0053475](https://doi.org/10.1371/journal.pone.0053475)
- Tam CS, Clément K, Baur LA, Tordjman J (2010) Obesity and low-grade inflammation: a paediatric perspective. *Obes Rev* 11:118–126. doi:[10.1111/j.1467-789X.2009.00674.x](https://doi.org/10.1111/j.1467-789X.2009.00674.x)
- The International HapMap C (2005) A haplotype map of the human genome. *Nature* 437:1299–1320

- Thompson FE, Subar AF, Loria CM et al (2010) Need for technological innovation in dietary assessment. *J Am Diet Assoc* 110:48–51. doi:[10.1016/j.jada.2009.10.008](https://doi.org/10.1016/j.jada.2009.10.008)
- Tucker KL, Maras J, Champagne C et al (2005) A regional food-frequency questionnaire for the US Mississippi Delta. *Public Health Nutr* 8:87–96
- United Nations Standing Committee on Nutrition (2012) What progress in nutrition? *Nutr Rev* 5:134. doi:[10.1111/j.1753-4887.1947.tb04238.x](https://doi.org/10.1111/j.1753-4887.1947.tb04238.x)
- Van Driel LM, Eijkemans MJ, de Jonge R et al (2009) Body mass index is an important determinant of methylation biomarkers in women of reproductive ages. *J Nutr* 139:2315–2321. doi:[10.3945/jn.109.109710](https://doi.org/10.3945/jn.109.109710)
- Van Hecke MV, Dekker JM, Nijpels G et al (2008) Homocysteine, S-adenosylmethionine and S-adenosylhomocysteine are associated with retinal microvascular abnormalities: the Hoorn Study. *Clin Sci* 114:479–487. doi:[10.1042/CS20070275](https://doi.org/10.1042/CS20070275)
- Wadhvani NS, Pisal HR, Mehendale SS, Joshi SR (2013) A prospective study of maternal fatty acids, micronutrients and homocysteine and their association with birth outcome. *Matern Child Nutr* 1–15. doi:[10.1111/mcn.12062](https://doi.org/10.1111/mcn.12062)
- Waterland RA, Travisano M, Tahiliani KG (2007) Diet-induced hypermethylation at agouti viable yellow is not inherited transgenerationally through the female. *FASEB J* 21:3380–3385. doi:[10.1096/fj.07-8229com](https://doi.org/10.1096/fj.07-8229com)
- Waterland RA, Kellermayer R, Laritsky E et al (2010) Season of conception in rural gambia affects DNA methylation at putative human metastable epialleles. *PLoS Genet* 6:e1001252. doi:[10.1371/journal.pgen.1001252](https://doi.org/10.1371/journal.pgen.1001252)
- Wernimont SM, Clark AG, Stover PJ et al (2011) Folate network genetic variation, plasma homocysteine, and global genomic methylation content: a genetic association study. *BMC Med Genet* 12:150. doi:[10.1186/1471-2350-12-150](https://doi.org/10.1186/1471-2350-12-150)
- WHO Expert Committee (1995) Physical status: the use and interpretation of anthropometry. A report of the WHO Expert Committee 416
- Williams RP (1956) *Biochemical individuality: the basis for the genotrophic concept*. Keats Publishing, New Canaan
- Wilson TA, Adolph AL, Butte NF (2009) Nutrient adequacy and diet quality in non-overweight and overweight Hispanic children of low socioeconomic status: the Viva la Familia Study. *J Am Diet Assoc* 109:1012–1021
- Wise CK, Cooney CA, Ali SF, Poirier LA (1997) Measuring S-adenosylmethionine in whole blood, red blood cells and cultured cells using a fast preparation method and high-performance liquid chromatography. *J Chromatogr B Biomed Sci Appl* 696:145–152
- Wolff GL, Kodell RL, Moore SR, Cooney CA (1998) Maternal epigenetics and methyl supplements affect agouti gene expression in *Avy/a* mice. *FASEB J* 12:949–957
- Yadrick K, Horton J, Stuff J et al (2001) Perceptions of community nutrition and health needs in the lower Mississippi delta: a key informant approach. *J Nutr Educ* 33:266–277
- Zichi D, Eaton B, Singer B, Gold L (2008) Proteomics and diagnostics: let's get specific, again. *Curr Opin Chem Biol* 12:78–85. doi:[10.1016/j.cbpa.2008.01.016](https://doi.org/10.1016/j.cbpa.2008.01.016)