1.a. **Full Title:** The Metabolome of BMI: A CONsortium of METabolomics Studies (COMETS) Meta-analysis of 85,000 adults

b. **Abbreviated Title (Length 26 characters):** The Metabolome of BMI

2. **Writing Group:**
   Writing group members: Rachel S Kelly, Kristin Young, Kari North, Steven Moore, Krista Zanetti, Ella Temprosa, Jessica Lasky-Su on behalf of the COMETS BMI working group

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. _RSK_____ [please confirm with your initials electronically or in writing]

**First author:** Rachel S Kelly
Address: Channing Division of Network Medicine, Brigham and Womens Hospital, Harvard Medical School, 181 Longwood Ave, Boston, MA 02115

Phone: +16175250065  
E-mail: hprke@channing.harvard.edu

**ARIC author** to be contacted if there are questions about the manuscript and the first author does not respond or cannot be located (this must be an ARIC investigator).
Name: Kari North
Address: Department of Epidemiology, University of North Carolina at Chapel Hill

Phone: (919) 966-2148  
Fax: (919) 966-9800  
E-mail: kari.north@unc.edu

3. **Timeline:** Manuscript to be completed <1 year after approval

4. **Rationale:**

More than one third of US adults are obese \(^1\). Obesity is associated with significant adverse
health outcomes including cardiovascular disease, diabetes and many cancers, posing a huge public health burden \(^2,3\). In the majority of cases, obesity emerges from a complex interaction between lifestyle, environmental factors and underlying genetic susceptibility with heritability estimates of up to 70\% \(^4,5\). The metabolome represents a dynamic functional readout of the state of a biological system; encompassing both genetic and environmental influences. Consequently, metabolomics is ideally suited to explore the drivers of BMI and the manifestation of obesity on a mechanistic and metabolic level.

Obesity is a whole-system disorder eliciting a coordinated metabolic response. Adipocytes are organs with multiple autocrine, paracrine, and endocrine roles in the regulation of the physiology of the entire body \(^6,7\). These roles are altered and dysregulated as individuals become obese and their adipocytes enlarge \(^8\). Therefore, it can hypothesized that obesity has a biologically informative metabolomic ‘fingerprint’ that can be identified via metabolomic profiling in a variety of biosamples.

This hypothesis is supported by evidence from the literature; which has identified a number of metabolites and metabolite classes as being associated with obesity, including branched-chain and aromatic amino acids, long-chain polyunsaturated fatty acids, phospholipids, lysophosphatidylcholines, sphingomyelins, carbohydrates, nucleotides and carnitines \(^9,10\). It is thought that the majority of these pathways are dysregulated as part of a compensatory response against the insulin resistance, altered glucose and lipid metabolism, and low-grade inflammatory state that accompany obesity \(^11\). Metabolomics and the characterization of the obese metabolome offers the opportunity to elucidate these underlying pathophysiological mechanisms and to potentially identify novel therapeutic targets that could help in the management of this condition. We propose to identify the metabolome of obesity and to explore how the metabolic alterations associated with increased BMI may be associated with dysregulated biological and pathogenic processes.

5. Main Hypothesis/Study Questions:

**Aim 1:** To evaluate relationships between metabolite concentrations and BMI across multiple cohorts, and to meta-analyze cohort-specific results.

**Aim 2:** To evaluate heterogeneity of associations by participant characteristics (gender, BMI, smoking status, prevalent disease, race, HRT use, age range) and by study characteristics (metabolomics platform, extraction, serum vs. plasma, fasting status, and region).

6. Design and analysis (study design, inclusion/exclusion, outcome and other variables of interest with specific reference to the time of their collection, summary of data analysis, and any anticipated methodologic limitations or challenges if present).

First, the analyses, comprising 15 potential models, will be conducted by ARIC investigators through the COMETs Analytics Portal [https://www.comets-analytics.org/](https://www.comets-analytics.org/)

Second, the results will be meta-analyzed with the results from all other participating cohorts to compute pooled effect estimates for each metabolite under each model. Then the results will be compared, summarized and interpreted.
**Study specific analyses:**

Data will be processed and normalized according to the methods chosen by each individual cohort. We will exclude participants in whom BMI is not available, or in whom it cannot be computed based on weight and height (kg/m²).

**Aim 1.** In each study, we will estimate Spearman correlations between metabolite and BMI (as a continuous variable) using three models adjusted for

- Model 1 - gender, age, race, nested case control status, study site (if applicable)
- Model 2 - gender, age, race, nested case control status, study site, educational level, smoking status, alcohol consumption, fasting status
- Model 3 - gender, age, race, nested case control status, study site, educational level, smoking status, alcohol consumption, fasting status, diabetes status

The specific coding and categorization of the variables for adjustment are shown in **Table 1** below.

For **aim 2**, in each study, we will stratify upon relevant participant characteristics (omitting the stratifying variable from that model), specifically we will stratify models 2 and 3 by

- Gender
- Race
- Fasting status
- Diabetes status
- Nested case status
- Age group

Furthermore we will run the two age group stratified models additionally adjusting for age status.

The full list of models can be seen in **Table 2**
Table 1: Parameterization of variables for purposes of adjustment and/or stratification

<table>
<thead>
<tr>
<th>Variable reference</th>
<th>Variable definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>Age at Entry</td>
</tr>
<tr>
<td>female</td>
<td>Female (0=male, 1=female)</td>
</tr>
<tr>
<td>smk_grp</td>
<td>Smoking status (0=never smoker, 1=former smoker, 2=current smoker, 3=missing)</td>
</tr>
<tr>
<td>race_grp</td>
<td>Race (0=White/European ancestry, 1=Non-European ancestry, 2=missing)</td>
</tr>
<tr>
<td>educ_grp</td>
<td>Education (0=Did not complete high school, 1=completed high-school, 2=post high-school training / some college, 3=completed college, 4=missing)</td>
</tr>
<tr>
<td>alc_grp</td>
<td>Alcohol (0=zero alcohol intake, 1=&gt;0-&lt;15 g/day, 2=15-&lt;30 g/day, 3=30+g/day, 4=missing)</td>
</tr>
<tr>
<td>fasted</td>
<td>Fasted (0=no, 1=yes--threshold at 6 hours, 2=missing)</td>
</tr>
<tr>
<td>prev_diabetes</td>
<td>Any diabetes at date of blood sample or within one year after blood sample (0=no, 1=yes, 2=missing)</td>
</tr>
<tr>
<td>age_grp</td>
<td>Age (0=0-5, 1=5-&lt;10, 2=10-&lt;20, 3=20-&lt;30, 4=30-&lt;40, 5=&lt;50-&lt;70, 6=70+)*</td>
</tr>
<tr>
<td>nested_case</td>
<td>Selected as a case for a nested case-control study (0=no, 1=yes)</td>
</tr>
</tbody>
</table>

For covariates that have a small proportion of missingness (e.g. <15% of participants), missing indicators will be used, per table 1 above. If a certain covariate is missing entirely from a certain study, analyses should proceed regardless, but the lack of that covariate will be noted, and these cohorts may be excluded from that specific analysis.

For metabolite observations that have listed values that are below the lab’s purported limit of detection (which happens for a few platforms), analyses should proceed regardless, as the values likely have at least some correspondence with concentrations at the low-end range.

For nested case-control studies, participants who later become cases should be retained, in keeping with our initial inclusive eligibility criteria. To evaluate the effect of including these participants, there will be a stratified analysis that excludes these participants and evaluates whether associations differ by eventual case status.

The analysis will not include unidentified metabolites, except for unidentified metabolites from Metabolon Inc. (e.g “X_17185”, “X_12063”) as these metabolites have a track record of reproducing with high biological consistency (X_17185 is a coffee/caffeine metabolite, X_12063 is a sex steroid hormone metabolite associated with BMI).

All data analyses will be done using COMETS-Analitics, where a sample dataset and models has been posted and is available to all. The COMETS-Analytics app is designed to output results in an e-mailable meta-analysis-ready format. It also incorporates a metabolite harmonization scheme on the backend that links metabolite names from one study to another. Once a study completes its analysis, the results are e-mailed centrally for meta-analysis.
### Table 2: BMI-metabolite models to be run in each Cohort

<table>
<thead>
<tr>
<th>Model</th>
<th>Adjustments</th>
<th>Strata</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>age female race_grp nested_case</td>
<td></td>
</tr>
<tr>
<td>Model 2</td>
<td>age female race_grp educ_grp smk_grp alc_grp fasted nested_case</td>
<td></td>
</tr>
<tr>
<td>Model 3</td>
<td>age female race_grp educ_grp smk_grp alc_grp fasted prev_diabetes nested_case</td>
<td></td>
</tr>
<tr>
<td>Model 2: Gender Stratified</td>
<td>age race_grp educ_grp smk_grp alc_grp fasted nested_case</td>
<td>female</td>
</tr>
<tr>
<td>Model 3: Gender Stratified</td>
<td>age race_grp educ_grp smk_grp alc_grp fasted prev_diabetes</td>
<td>female</td>
</tr>
<tr>
<td>Model 2: Race Stratified</td>
<td>age female educ_grp smk_grp alc_grp fasted nested_case</td>
<td>race_grp</td>
</tr>
<tr>
<td>Model 3: Race Stratified</td>
<td>age female educ_grp smk_grp alc_grp fasted prev_diabetes</td>
<td>race_grp</td>
</tr>
<tr>
<td>Model 2: Fasting Status Stratified</td>
<td>age female race_grp educ_grp smk_grp alc_grp fasted nested_case</td>
<td>fasted</td>
</tr>
<tr>
<td>Model 3: Fasting Status Stratified</td>
<td>age female race_grp educ_grp smk_grp alc_grp fasted prev_diabetes nested_case</td>
<td>fasted</td>
</tr>
<tr>
<td>Model 2: Diabetes Stratified</td>
<td>age female race_grp educ_grp smk_grp alc_grp fasted</td>
<td>prev_diabetes</td>
</tr>
<tr>
<td>Model 2: Neasted Case status Stratified</td>
<td>age female race_grp educ_grp smk_grp alc_grp fasted</td>
<td>nested_case</td>
</tr>
<tr>
<td>Model 2: Age Stratified</td>
<td>female race_grp educ_grp smk_grp alc_grp fasted nested_case</td>
<td>age_grp</td>
</tr>
<tr>
<td>Model 3: Age Stratified</td>
<td>female race_grp educ_grp smk_grp alc_grp fasted prev_diabetes</td>
<td>age_grp</td>
</tr>
<tr>
<td>Model 2: Age Stratified-Age adjusted</td>
<td>age female race_grp educ_grp smk_grp alc_grp fasted</td>
<td>age_grp</td>
</tr>
<tr>
<td>Model 3: Age Stratified-Age adjusted</td>
<td>age female race_grp educ_grp smk_grp alc_grp fasted prev_diabetes nested_case</td>
<td>age_grp</td>
</tr>
</tbody>
</table>

**Meta-analysis**

The meta-analyses will be performed centrally. We will use Fisher’s R to Z conversion to obtain effect estimates that can be meta-analyzed. Standard errors will be based on sample size using standard formula. Meta-analysis will be done using fixed effects meta-analysis, with random effects calculated as well and compared with fixed effects estimates. Heterogeneity across studies will be evaluated by Cochran’s Q. Z-values will be back-converted to R values for final presentation. Influence for aim 1 will be tested by omitting each cohort in sequence, and reestimating meta-analysis fixed effects.

**Aim 1**, we will use a Bonferroni-corrected threshold to determine statistical significance of associations, with an expected total of ~2,000 metabolites across all COMETS studies. Based on
the prior studies, we estimate that dozens, and potentially hundreds, of metabolites will have Spearman correlations greater than 0.2. Given our consortium-based approach, we have near 100% power for detecting associations of this magnitude even after Bonferroni correction. For a subset of the top metabolites, we may additionally evaluate dose-response using spline models in locally available studies.

Aim 2. heterogeneity across groups defined by participant characteristics (e.g. gender) and study characteristics (e.g. platform used) will be meta-analyzed within each group. Group-specific results will be compared using the Wald test for homogeneity, and interaction tests will be corrected for multiple testing.

The first paper will focus on the metabolites identified as significant in the meta-analyses, with a specific focus on those that retain significance across the models. Pathway analysis and metabolite visualization methods will be employed to explore the biological context of these metabolites; including but not limited to MetaboAnalyst \(^2\) and MetPA \(^3\). A number of \textit{a priori} hypotheses suggesting involvement of BCAA catabolism, \(\beta\)-oxidation of fatty acids, inflammatory and immune responses, carnitines, carbohydrate and nucleotide metabolism \(^6,8,9,11\) have been generated based on the literature. Results will be critically evaluated with respect to these stated pathways and metabolites.

**Required outcome data:** BMI (continuous, at least two decimals preferred)

BMI is used here as an inexpensive, non-invasive measure of obesity available in the majority of cohorts (35/43 cohorts) that has been demonstrated to predict the risk of obesity-related complications \(^5\).

**Required exposure data:** Metabolites (for a metabolite to be included for analysis, at least 15 persons must have non-missing value of that metabolite. This includes in stratified analyses.)

Metabolites profiled by both Mass spec and NMR will be eligible for inclusion

**Required covariate data:** See table 1.

Notes: No individual level data is required for this proposal. Cohorts will upload their own data to COMETS Analytics, which will not be accessible or visible to the PIs of this study (Kelly and Lasky-Su). Cohorts will receive their own summary statistics for each of the models.

7.a. Will the data be used for non-CVD analysis in this manuscript? ____ Yes  ____X__ No

b. If Yes, is the author aware that the file ICTDER03 must be used to exclude persons with a value RES OTH = “CVD Research” for non-DNA analysis, and for DNA analysis RES DNA = “CVD Research” would be used?  ____X__ Yes  ____ No

(This file ICTDER has been distributed to ARIC PIs, and contains the responses to consent updates related to stored sample use for research.)

8.a. Will the DNA data be used in this manuscript? ____ Yes  ____X__ No
8.b. If yes, is the author aware that either DNA data distributed by the Coordinating Center must be used, or the file ICTDER03 must be used to exclude those with value RES_DNA = “No use/storage DNA”?  ____ Yes  ____ No

9. The lead author of this manuscript proposal has reviewed the list of existing ARIC Study manuscript proposals and has found no overlap between this proposal and previously approved manuscript proposals either published or still in active status. ARIC Investigators have access to the publications lists under the Study Members Area of the web site at: http://www.cscc.unc.edu/aric/mantrack/maintain/search/dtSearch.html

____X___ Yes  _______ No

10. What are the most related manuscript proposals in ARIC (authors are encouraged to contact lead authors of these proposals for comments on the new proposal or collaboration)?

3066 (Kristin Young and Kari North are writing group members on both proposals, and Dr. Young has a funded R21 to look at metabolite-obesity associations in ARIC).

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data?  ____X___ Yes  ____ No

11.b. If yes, is the proposal

____X_ A. primarily the result of an ancillary study (list number* _2017.10__)  

____ B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* __________  __________  __________)

*ancillary studies are listed by number at https://www2.cscc.unc.edu/aric/approved-ancillary-studies

12a. Manuscript preparation is expected to be completed in one to three years. If a manuscript is not submitted for ARIC review at the end of the 3-years from the date of the approval, the manuscript proposal will expire.

12b. The NIH instituted a Public Access Policy in April, 2008 which ensures that the public has access to the published results of NIH funded research. It is your responsibility to upload manuscripts to PubMed Central whenever the journal does not and be in compliance with this policy. Four files about the public access policy from http://publicaccess.nih.gov/ are posted in http://www.cscc.unc.edu/aric/index.php, under Publications, Policies & Forms. http://publicaccess.nih.gov/submit_process_journals.htm shows you which journals automatically upload articles to PubMed central.
References