1.a. **Full Title**: Physical activity and DNA methylation: an epigenome-wide association study

b. **Abbreviated Title (Length 26 characters)**: Physical activity EWAS

2. **Writing Group**:

   Writing group members:
   - Elizabeth Hibler, MPH, PhD, Northwestern University Feinberg School of Medicine, Chicago, IL
   - Yinan Zheng, Northwestern University Feinberg School of Medicine, Chicago, IL, USA, vinanzheng2011@u.northwestern.edu
   - Stella Aslibekyan, University of Alabama-Birmingham, Birmingham, AL, USA, saslibek@uab.edu
   - Sergi Baixeras, University of Alabama-Birmingham, Birmingham, AL, US, Birmssayols@imim.es
   - Qi Dai, Vanderbilt University School of Medicine, Nashville, TN, USA, qi.dai@vanderbilt.edu
   - Ellen W. Demerath, PhD, University of Minnesota, Minneapolis, MN, USA, ewd@umn.edu
   - James S. Pankow, PhD, University of Minnesota, Minneapolis, MN, USA, panko001@umn.edu
   - Steven Nguyen, MPH, University of Minnesota, Minneapolis, MN USA, nguy2295@umn.edu
   - Weihua Guan, PhD, University of Minnesota, Minneapolis, MN, USA, wguan@umn.edu
   - Katie Holliday, PhD, University of North Carolina-Chapel Hill, Chapel Hill, NC, USA, khausman@email.unc.edu
   - Andrea Baccarelli, Columbia University, New York, NY, USA, ab4303@columbia.edu
   - Kiarri Kershaw, Northwestern University Feinberg School of Medicine, Chicago, IL, USA, kershaw@northwestern.edu
   - Lifang Hou, Northwestern University Feinberg School of Medicine, Chicago, IL, USA, lhou@northwestern.edu

**Please note, as described below, this study also incorporates data from multiple cohorts in CHARGE, including ARIC. Additional co-authors from the studies are TBD and will be added as co-authors once the manuscript is ready for review.**
I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. _EH_____ [please confirm with your initials electronically or in writing]

First author: Elizabeth Hibler, MPH, PhD  
Address: Northwestern University Feinberg School of Medicine  
680 N. Lake Shore Dr., Suite 1400  
Chicago, IL 60657  
Phone: 312-503-1178     Fax: 312-908-9588  
E-mail: elizabeth.hibler@northwestern.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: Ellen Demerath  
Address: Division of Epidemiology and Community Health  
University of Minnesota  
1300 S. 2nd St., Suite 300  
Minneapolis, MN 55454  
Phone: 612-624-8231     Fax:  
E-mail: ewd@umn.edu

3. Timeline: We are in the process of conducting the data analysis in individual cohorts. We expect this to be done by March 1, 2017. Then, we expect the meta-analysis to be completed by May 1, 2017 and the manuscript to be out to co-authors by summer 2017 at the latest.

4. Rationale:  
Studies demonstrate that low levels of physical activity are associated with increased mortality and risk of common diseases 1, 2. However, the biological mechanisms driving population-level associations between activity and disease risk are unclear. Changes in patterns of DNA methylation are suggested as a promising mechanism. However, the available data are sparse. No study has examined if physical activity is associated with DNA methylation at CpG loci on an epigenome-wide scale.

A growing number of studies demonstrate that methylation at CpG sites in disease-associated gene loci (i.e. candidate genes) varies with environmental and lifestyle exposures 3-7. Studies found that physical activity may also influence gene expression by changing DNA methylation patterns in genes associated with disease risk 6, 8, 9. For example, Nakajima et al. identified changes in methylation signature for the apoptosis-associated speck-like protein containing a caspase recruitment domain (ASC) gene 10. In addition, physical activity is associated with methylation at the promoters of peroxisome proliferator-activated receptor 1 (PPAR-γ) and pyruvate dehydrogenase kinase isoenzyme 4 (PDK4), which are involved in carcinogenesis pathways such as inflammation and cellular metabolism 9, 11-13. Studies have also evaluated changes in global methylation status with mixed results. Among cancer-free adults
(n=161), Zhang et al. found significantly increased global DNA methylation with increasing physical activity, as demonstrated by increased global white blood cell methylation (β = 2.54, 95% CI 0.67-4.42) among those participating in 26-30 minutes of daily physical activity compared to ≤ 5 minutes/day. In contrast, in the largest study identified, Luttropp et al. reported decreased global methylation in peripheral white blood cells following exercise among 1016 older adults. Other studies also reported decreased global methylation with increasing physical activity in peripheral white blood cells and skeletal muscle. However, no studies evaluated the influence of physical activity on an epigenome-wide scale in white blood cells.

In summary, for this paper, we will evaluate the relationship between physical activity and epigenome-wide patterns of DNA methylation using data from multiple studies, including ARIC. As described below, the methods for harmonizing physical activity data across studies will be the biggest challenge. However, we have developed a plan with alternative approaches that will be employed as needed. Furthermore, no studies have evaluated whether the relationship between activity and DNA methylation differs by sex or race, though there is growing evidence to suggest that patterns of DNA methylation may differ by these factors. This will be the first EWAS to evaluate the relationship between physical activity and DNA methylation across multiple, large-scale cohort studies.

5. Main Hypothesis/Study Questions:

The study questions include:
1. Which gene regions are differentially methylated between individuals who report high vs. low physical activity?
2. Does the association between physical activity and DNA methylation level differ by race/ethnicity or by sex?

Hypotheses
1. Higher physical activity is associated with hypomethylation in genes related to reduced disease risk and hypermethylation at loci linked to increased disease risk.
2. Higher physical activity has a greater influence on DNA methylation among women compared to men. Patterns of DNA methylation associated with higher physical activity will differ by race.

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).

Overview. The primary aims of the CHARGE consortium is to facilitate meta-analyses of epigenome-wide studies among an international group of large population-based cohort studies. The proposed study will be a large-scale study and novel in that we will evaluate associations between healthy behaviors and epigenome-wide variation in DNA methylation. We will determine if patterns of DNA methylation vary at CpG in association with physical activity. To complete the study questions, we have developed detailed data analysis plans and R code for
each study, with study specific instructions as necessary. After each cohort reports the results of their multivariate regression analysis to us, we will run a meta-analysis across studies to estimate summary measures of association. This unique study will potentially identify biomarkers that vary with self-reported physical activity level, which may have future applications towards studies of chronic disease and cancer risk.

**Study Population:** Within the ARIC study, there are a total of \( n = 2,097 \) African American participants with available DNA methylation and activity data. In addition, to-date, other cohorts included in this study include the Women’s Health Initiative (WHI), Coronary Artery Risk Development in Young Adults (CARDIA), Normative Aging Study (NAS), Genetics of Lipid Lowering Drugs and Diet Network (GOLDN), and Registre Gironí del COR, the Girona Heart Registry (REGICOR).

**Exclusions:** Participants affected by leukemia and on chemotherapy will be excluded because it is possible that these factors may alter DNA methylation in whole blood cells.

**Primary Outcomes.** DNA methylation (DNAm) at CpG sites as determined by the methods described below. The ARIC study used the Illumina Human Methylation 450 Bead Chips and followed the Illumina Infinium HD Methylation protocol (Illumina, San Diego, CA, USA).

**Main Exposure.** Physical activity will be evaluated as a dichotomous variable (Yes/No) for meeting the American Cancer Society guidelines for 150 minutes of moderate-vigorous physical activity per week. Further, metabolic-equivalent hours (MET-hours) per week will be evaluated as a continuous variable. We will also assess physical activity intensity variables.

**Co-variables.** Additional covariates include: body mass intake, alcohol intake, smoking status, history of cancer, and sleep. The technical variables of plate and chip number will adjust for any variation that may have been introduced in the lab. We will also adjust for Houseman estimates of cell type proportions (neutrophils; eosinophils; monocytes; lymphocytes; basophils). Covariates were selected based upon prior evidence in the literature and/or a biological basis for confounding the relationship between activity and DNA methylation.

**Primary exposures: Self-reported physical activity**

Our first step will be data harmonization across cohorts to ensure comparability of the physical activity variables. We expect to assess and report results for several types of physical activity variables. Since self-reported physical activity varies greatly between cohorts, we at least plan to create a binomial variable defining dichotomous physically active ("yes/no") variables by simple percentage-based thresholds. This approach is also based upon guidance from University of Copenhagen researchers, who led successful projects to physical activity data across cohorts for a large-scale GWAS. We will assess options for defining dichotomous physically active ("yes/no") variables by simple percentage-based thresholds (25% vs. 75% and 50% vs. 50%). However, we will also explore alternatives including 1) to assess continuous measures of physical activity in time units, METs 2) to assess the relationships based upon existing guidelines for physical activity, such as 150 minutes per week moderate-vigorous or 75 minutes per week vigorous activity.
**Statistical analysis**

We will use multivariate regression with models adjusted for, at minimum, age, sex, race, and technical covariates. Additional covariates will be evaluated including body mass index, education, alcohol intake, dietary intake, smoking status, medical history of chronic disease, and sleep. The technical variables of plate and chip number will adjust for any variation that may have been introduced in the lab. Other covariates were selected based upon prior evidence in the literature and/or a biological basis for confounding the relationship between physical activity and DNA methylation. Adjustment for confounding variables will be the same across studies. Stratified analysis by sex will be examined. Furthermore, using the Houseman method, we will adjust for white blood cell proportion to account for any confounding that may occur related to variation in proportions of cell types found in whole blood. We will then conduct a meta-analysis to determine the overall associations between physical activity and DNA methylation at specific CpG loci across cohorts. We will use fixed or random models depending upon tests of heterogeneity between cohorts. We will use R software for analysis.

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? ____ 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? __X__ Yes   ____ 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”? __X__ 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/search.php

____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)?

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data? _____ Yes   __X__ No
11.b. If yes, is the proposal
   ___ A. primarily the result of an ancillary study (list number* ____________)
   ___ 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 http://www.cscce.unc.edu/aric/forms/

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://publicaccess.nih.gov/submit_process_journals.htm shows you which journals
     automatically upload articles to PubMed central.

13. Per Data Use Agreement Addendum, approved manuscripts using CMS data shall be
     submitted by the Coordinating Center to CMS for informational purposes prior to
     publication. Approved manuscripts should be sent to Pingping Wu at CC, at
     pingping_wu@unc.edu. I will be using CMS data in my manuscript ____ Yes __X__ No.

1. Friedenreich CM, Neilson HK and Lynch BM. State of the epidemiological evidence on physical activity
2. Matthews CE, Cohen SS, Fowke JH, Han X, Xiao Q, Buchowski MS, Hargreaves MK, Signorello LB and
   Blot WJ. Physical activity, sedentary behavior, and cause-specific mortality in black and white adults in the
3. Castillo-Fernandez JE, Spector TD and Bell JT. Epigenetics of discordant monozygotic twins: implications
4. Mathers JC, Strathdee G and Relton CL. Induction of epigenetic alterations by dietary and other
5. White AJ, Sandler DP, Bolick SC, Xu Z, Taylor JA and DeRoo LA. Recreational and household physical
   Santella RM. Physical activity and global genomic DNA methylation in a cancer-free population. Epigenetics.
   Cardarelli R. Dietary patterns are associated with levels of global genomic DNA methylation in a cancer-free
9. Pareja-Galeano H, Sanchis-Gomar F and Garcia-Gimenez JL. Physical exercise and epigenetic modulation:


