1.a. Full Title: Genome-Wide Association Study of Physical Activity

b. Abbreviated Title (Length 26 characters): GWAS of Physical Activity

2. Writing Group:
Marilyn Cornelis, Ph.D., Harvard School of Public Health
Keri Monda, Ph.D., University of North Carolina at Chapel Hill
Kari North, Ph.D., University of North Carolina at Chapel Hill
Marcel den Hoed, MRC Epidemiology Unit, Cambridge, UK
Other authors welcome

Lead Author**
**This will be a multi-first authored (starred) effort, but our working lead is listed here:
Marilyn Cornelis, Ph.D., Department of Nutrition, Harvard School of Public Health

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

First author:
Name: Marilyn Cornelis
Address: Harvard School of Public Health
655 Huntington Avenue
Building 2 - Room 355
Boston, MA 02115
Office Phone: 617.432.1333
Fax: 617.432.1333
mcorneli@hsph.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: Keri Monda
Address: 137 E. Franklin St., Ste 306, Chapel Hill, NC, 27514
Phone: 919-966-8491 Fax: 919-966-9800
E-mail: monda@unc.edu
3. **Timeline:**

*Expected meta-Analysis completion = July 1, 2011*; *Additional analyses (if necessary) complete = August 1, 2011*; *Manuscript drafting / submission = September 15, 2011*

4. **Rationale:**

There is persuasive epidemiological and experimental research evidence on the benefits of regular physical activity in middle-aged and older adults\(^1\text{-}^4\). Interventions designed to promote adoption and maintenance of physical activity have largely failed to produce sustained effects across large segments of the population, as the majority of individuals who initiate an exercise regimen are unable to persist in the long term. Identifying individual characteristics that are predictive of physical activity maintenance is a critical first step to inform the design of behavioral interventions to promote ongoing physical activity engagement.

Evidence suggests that physical activity level is a heritable trait, with estimates ranging from low to moderate levels of heritability (17-62%) due to genetic factors\(^5\text{-}^{11}\). Many of the heritability estimates have been derived from twin studies which have found that genetic factors play a role in adult exercise participation\(^8\text{-}^{12}\), suggesting that the remaining heritability is accounted for by environmental factors. A number of genome-wide linkage studies\(^13\text{-}^{15}\) and candidate gene association studies\(^8\text{-}^{16}\text{-}^{18}\) have been conducted as well as one genome-wide association study of exercise behavior\(^19\); however, although a number of suggestive loci have been discovered, no loci have yet been incontrovertibly associated with either physical activity, the time spent on moderate and vigorous intensity activities, or sedentary behavior.

A better understanding how biology shapes health behaviors can open up new avenues of investigation into more personalized behavioral therapies to increase physical activity in populations. Our aim is to increase our knowledge on the genetic etiology of physical activity related traits by means of a large-scale meta-analysis of genome-wide association studies.

5. **Main Hypothesis/Study Questions:**

We hypothesize that the motivation for habitual physical activity is at least partly genetically determined and driven by an individual’s innate propensity to be physically active. We propose to perform a genome-wide association (GWA) study to identify loci associated with physical activity levels in individuals of European descent.

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).**
Analytic Plan:

<table>
<thead>
<tr>
<th>Exclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Exclude individuals with missing information on physical activity variables of interest or covariates</td>
</tr>
<tr>
<td>• For studies with multiple time points, select baseline/earliest time point</td>
</tr>
<tr>
<td>• Analyses will be restricted to European Americans</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trait Creation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our goal is to capture a dichotomous trait corresponding to any moderate to vigorous leisure activity. Such activity would not include occupation-related activity (i.e. shoveling, heavy-lifting) and/or light leisure activity (i.e. walking, gardening).</td>
</tr>
</tbody>
</table>

Trait is created from questions in the ARIC questionnaire querying hours per week of physical activity as follows:

Moderate/vigorously active ‘1’ vs non-moderate/vigorously active ‘0’.

- **modvig=’1’** if subjects report engaging in ≥1 h/wk of leisure moderate/vigorous activity. These include activities with MET values of 3 or greater (see ‘Compendium of Physical Activity’).
- Otherwise, **modvig=’0’**.

**NOTE**: subjects who specifically report <1 h/wk leisure moderate or vigorous activity will have a trait1 of ‘0’. Do not confuse with subjects with ‘missing information’ whom should be excluded from analyses.

<table>
<thead>
<tr>
<th>Covariate Creation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. <strong>age</strong>: continuous</td>
</tr>
<tr>
<td>2. <strong>sex</strong>: dichotomous, 0=male, 1=female</td>
</tr>
<tr>
<td>3. <strong>study_site</strong></td>
</tr>
<tr>
<td>4. <strong>EV1, EV2, EV3</strong>: top eigenvectors will be included to adjust for population substructure</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Running the Analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td>An additive genetic model will be used. For imputed data, use allele dosages. Otherwise, use measured genotype. Code reference allele as you wish (each study will document the reference allele in the results file).</td>
</tr>
</tbody>
</table>

- **Model 1:**

  Logistic regression of **modvig** on allele dosage/genotype:
  
  \[
  \text{trait} \sim \mu + \text{SNP} + \text{age} + \text{age}^2 \text{study site} + EVs
  \]

- **Model 2:**

  Logistic regression of **modvig** on allele dosage/genotype:
  
  \[
  \text{trait} \sim \mu + \text{SNP} + \text{age} + \text{age}^2 + \text{study site} + EVs + BMI
  \]
### Results Documentation

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>marker</td>
<td>SNP id as rs number.</td>
</tr>
<tr>
<td>chr</td>
<td>Chromosome number. Use symbols X and Y for non-autosomal markers.</td>
</tr>
<tr>
<td>position</td>
<td>Physical position for the reference sequence (indicate build 35/36 in readme file).</td>
</tr>
<tr>
<td>effect_allele</td>
<td>The ‘effect-allele’, also referred to as ‘coded-allele’ or ‘risk-allele’. For example: AA=0, AG=1 and GG=2, the effect_allele is G.</td>
</tr>
<tr>
<td>other_allele</td>
<td>The alternate allele.</td>
</tr>
<tr>
<td>strand</td>
<td>+ or -, representing either the positive/forward strand or the negative/reverse strand of the human genome reference sequence.</td>
</tr>
<tr>
<td>beta</td>
<td>Beta estimate from genotype-phenotype association --‘NA’ if not available.</td>
</tr>
<tr>
<td>stderr</td>
<td>Standard error of beta estimate --‘NA’ if not available.</td>
</tr>
<tr>
<td>pvalue</td>
<td>p-value for test statistic --‘NA’ if not available.</td>
</tr>
<tr>
<td>freq</td>
<td>Allele frequency for the effect_allele --‘NA’ if not available.</td>
</tr>
<tr>
<td>hwe</td>
<td>Exact test Hardy-Weinberg equilibrium p-value—only directly typed SNPs, ‘NA’ for imputed.</td>
</tr>
<tr>
<td>n</td>
<td>Total sample with phenotype and genotype for SNP.</td>
</tr>
<tr>
<td>imputed</td>
<td>1=imputed, 0=if directly typed. Use ‘0’ coding when imputation was used to only ‘fill-in’ the missing genotypes (i.e &lt;10% of subjects/snp).</td>
</tr>
<tr>
<td>oevar_imp</td>
<td>Observed divided by expected variance for imputed allele dosage. Also referred to as ‘Rsq” in MACH. If some other quality measure was used for imputed SNPs, please describe in readme file. ‘NA’ or 1 if directly genotyped.</td>
</tr>
</tbody>
</table>

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?  
   N/A

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.cscn.unc.edu/ARIC/search.php](http://www.cscn.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)?
   
   • None. All extant GWAS manuscript proposals use PA as a covariate or effect measure modifier, not as an outcome.

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 #2006.03 & 2007.02)

12. 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.

   Understood, and we will meet this deadline

References