1.a. Full Title: Meta-analysis of GWAS for platelet count

b. Abbreviated Title (Length 26 characters): platelet count GWAS

2. Writing Group:
   Writing group members: CHARGE hemostasis group, including Weihong Tang and Aaron Folsom from ARIC. Final author list is to be determined.

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

First author: Weihong Tang
Address: Division of Epidemiology & Community Health, School of Public Health, University of Minnesota, 1300 South 2nd Street, Suite 300, Minneapolis, MN 55454
Phone: 612-626-9140 Fax: 612-624-0315
E-mail: tang0097@umn.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: Aaron Folsom
Address:
Phone: 612-626-8862 Fax:
E-mail: folso001@umn.edu

3. Timeline:
   Plan to finish GWAS and meta-analysis in 2009, submitting to a journal at the end of 2009.

4. Rationale:
   Hematological traits, including the concentration of haemoglobin (Hb), the numbers of white (WBC) and red blood cells (RBC) and platelets (PLT), are commonly used parameters in the clinic. Deviations outside the normal ranges are indicative of many
different disorders including cancer, infectious and immune diseases. The count and volume of these cellular elements in circulating blood are highly heritable and tightly regulated. Dr. Nicole Soranzo from the Wellcome Trust Sanger Institute conducted an initial meta-analysis of hematological traits in the HaemGen Consortium (including <5000 GWAS samples and 10,000 replications) and identified 12 loci associated with mean platelet volume (MPV) and three with PLT, as well as red- and white-blood cell loci (Soranzo et al Blood 2009; Meisinger et al AJHG 2009; Soranzo et al NG provisionally accepted). Now Dr. Soranzo is inviting other Consortiums including the CHARGE to collaborate with the HaemGen for a larger meta-analysis to further the investigations of genetic associations for PLT and MPV. In CHARGE, ARIC and CHS have measures for PLT that can contribute to the re-meta-analysis. MPV is not available from any cohorts of CHARGE. It is expected that more than 28,000 individuals with GWAS data and blood trait parameters will be assembled with the contribution of the additional cohorts.

5. Main Hypothesis/Study Questions:

Gene variants can be identified that associate with levels of platelet traits.

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).

Participating groups:

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Contact person</th>
<th>Contact person mail</th>
<th>N</th>
<th>Traits available</th>
<th>Agreed to participate</th>
</tr>
</thead>
<tbody>
<tr>
<td>UKBS1</td>
<td>Nicole Soranzo</td>
<td><a href="mailto:ns6@sanger.ac.uk">ns6@sanger.ac.uk</a></td>
<td>1,200</td>
<td>MPV/PLT</td>
<td>Y</td>
</tr>
<tr>
<td>UKBS2</td>
<td>Nicole Soranzo</td>
<td><a href="mailto:ns6@sanger.ac.uk">ns6@sanger.ac.uk</a></td>
<td>1,500</td>
<td>MPV/PLT</td>
<td>Y</td>
</tr>
<tr>
<td>TwinsUK1</td>
<td>Nicole Soranzo</td>
<td><a href="mailto:ns6@sanger.ac.uk">ns6@sanger.ac.uk</a></td>
<td>2,300</td>
<td>MPV/PLT</td>
<td>Y</td>
</tr>
<tr>
<td>TwinsUK2</td>
<td>Nicole Soranzo</td>
<td><a href="mailto:ns6@sanger.ac.uk">ns6@sanger.ac.uk</a></td>
<td>3,500</td>
<td>MPV/PLT</td>
<td>Y</td>
</tr>
<tr>
<td>KORA-F3</td>
<td>Christian Gieger</td>
<td><a href="mailto:christian.gieger@helmholtz-muenchen.de">christian.gieger@helmholtz-muenchen.de</a></td>
<td>1,600</td>
<td>MPV/PLT</td>
<td>Y</td>
</tr>
<tr>
<td>KORA-S4</td>
<td>Christian Gieger</td>
<td><a href="mailto:christian.gieger@helmholtz-muenchen.de">christian.gieger@helmholtz-muenchen.de</a></td>
<td>1,800</td>
<td>MPV/PLT</td>
<td>Y</td>
</tr>
<tr>
<td>SHIP</td>
<td>Alexander Teumer</td>
<td><a href="mailto:ateumer@uni-greifswald.de">ateumer@uni-greifswald.de</a></td>
<td>3,300</td>
<td>MPV/PLT</td>
<td>Y</td>
</tr>
<tr>
<td>ARIC</td>
<td>Weihong Tang</td>
<td><a href="mailto:tang0097@umn.edu">tang0097@umn.edu</a></td>
<td>9,000</td>
<td>PLT</td>
<td>Y</td>
</tr>
<tr>
<td>CHS</td>
<td>Weihong Tang</td>
<td><a href="mailto:tang0097@umn.edu">tang0097@umn.edu</a></td>
<td>3,235</td>
<td>PLT</td>
<td>Y</td>
</tr>
</tbody>
</table>
Three stages are planned (MPV is not available in ARIC but included below for the completeness of the analysis plan):

**Stage 1: GWAS in each cohort.**
1) Traits are:
   a. untransformed PLT values in $1\times10^9/l$
   b. inverse normal transformed PLT values
   c. natural-log transformed MPV in log fl
   d. inverse normal transformed MPV values
2) Use additive genetic model, 1 d.f. trend test
3) Use genotyped and imputed SNPs (where genotyped SNPs available, use them; otherwise, use imputed SNPs)
4) No imputation of missing traits
5) No outlier exclusion (apart for samples with extreme outlier phenotypes (> 6 s.d. from mean for untransformed PTL and MPV)
6) Analyses are age-, gender- and field center-adjusted; In additional analyses, SNPs in already known loci (a list will be provided) will be adjusted for.
   - (transformed) TRAIT ~ age + gender + usual adjustments
   - (transformed) TRAIT ~ age + gender + usual adjustments + SNPs

**Stage 2 (meta-analysis).**
Summary statistics from GWAS of individual cohorts will be used to calculate meta-analysis P-values by using inverse-variance meta-analysis. Leading SNPs with $P < 5\times10^{-8}$ will be selected for follow-up in stage 3. The meta-analysis will be conducted at Dr. Soranzo’s site.

**Stage 3 (on SNPs selected from meta-analysis)**
1) Associations calculated using method above for the correlated hematological traits* (Hb, HCT, RBC, MCH, MCHC, MCV, WBC).
2) Additional analysis carried out for lead traits for additional models, for instance:
   - (transformed) TRAIT ~ age+gender+usual adjustments+MPV

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Trait</th>
<th>Units</th>
<th>Trait transformation</th>
<th>Correction†</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Leading traits</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPV</td>
<td>Mean Platelet Volume</td>
<td>Fl</td>
<td>Natural log and inverse normal</td>
<td>Age, sex</td>
</tr>
<tr>
<td>PLT</td>
<td>Platelet Count</td>
<td>$1\times10^9/l$</td>
<td>None and inverse normal</td>
<td>Age, sex</td>
</tr>
<tr>
<td>*<em>Additional traits</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hb</td>
<td>Haemoglobin concentration (Haemoglobin Count)</td>
<td>g/dl</td>
<td>none</td>
<td>Age, sex</td>
</tr>
<tr>
<td>MCH</td>
<td>Mean Red Blood Cell Haemoglobin Content</td>
<td>Pg</td>
<td>none</td>
<td>Age, sex</td>
</tr>
<tr>
<td>MCHC</td>
<td>Mean Red Blood Cell Haemoglobin Concentration</td>
<td>g/dl</td>
<td>none</td>
<td>Age, sex</td>
</tr>
<tr>
<td>MCV</td>
<td>Mean Red Blood Cell Volume</td>
<td>fl</td>
<td>none</td>
<td>Age, sex</td>
</tr>
<tr>
<td>RBC</td>
<td>Red Blood Cell Count</td>
<td>$1\times10^{12}/l$</td>
<td>natural log</td>
<td>Age, sex</td>
</tr>
<tr>
<td>WBC</td>
<td>White Blood Cell Count</td>
<td>$1\times10^9/l$</td>
<td>natural log</td>
<td>Age, sex</td>
</tr>
</tbody>
</table>
*For evaluating the associations with SNPs that show signals for MPV and PLT only, the CHARGE lead authors for RBC and WBC papers have agreed to look up the SNPs identified for PLT and MPV; †field center will be added in the ARIC analysis.

7.a. Will the data be used for non-CVD analysis in this manuscript? ____ Yes ___xx__ 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 ICTDER03 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? ___xx__ 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”? ___xx__ 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.csc.unc.edu/ARIC/search.php](http://www.csc.unc.edu/ARIC/search.php)

___xx_____ 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.

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

11.b. If yes, is the proposal

___ A. primarily the result of an ancillary study (list number* ____________)

___xx__ B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* 2006.03, 2007.02)

*ancillary studies are listed by number at [http://www.csc.unc.edu/aric/forms/](http://www.csc.unc.edu/aric/forms/)
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.