Population Architecture using Genomics and Epidemiology (PAGE)
Ver. 06/14/10

PAGE Manuscript Proposal Template
Submit proposals by email to the PAGE Coordinating Center at Rwilliams@biology.rutgers.edu

All sections must be completed; incomplete applications will be returned. Do not exceed 3 pages in length (not including references).

PAGE Ms. Number: 104 Submission Date: 9/30/2016 [Approval Date: ___]

Title of Proposed MS: Common and Rare Genetic Variants Associated with White Blood Cell Traits in US Minorities

Abbreviated Title of Proposed MS: MEGA Analysis of WBC Traits

I. INVESTIGATOR INFORMATION:

Name of Lead Author: Katherine Nishimura  |  Junior Investigator? Y
Email Address: knishimu@fredhutch.org
Telephone Number: 206.667.1782

Name of Corresponding Author (if different): NA
Email Address:
Telephone Number:

Names, affiliations and email address of PAGE Investigators proposed as co-authors:

*We welcome the inclusion of additional authors and will reassess the author list based on contribution and involvement, as the scope of the paper becomes more defined.

<table>
<thead>
<tr>
<th>N, N</th>
<th>Affiliation in PAGE</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nishimura, Katherine</td>
<td>WHI</td>
<td><a href="mailto:knishimu@fredhutch.org">knishimu@fredhutch.org</a></td>
</tr>
<tr>
<td>Kooperberg, Charles</td>
<td>WHI</td>
<td><a href="mailto:clk@fredhutch.org">clk@fredhutch.org</a></td>
</tr>
<tr>
<td>Reiner, Alex</td>
<td>WHI</td>
<td><a href="mailto:apreiner@uw.edu">apreiner@uw.edu</a></td>
</tr>
<tr>
<td>Shadyab, Aladdin</td>
<td>WHI</td>
<td><a href="mailto:ahshadya@ucsd.edu">ahshadya@ucsd.edu</a></td>
</tr>
<tr>
<td>Avery, Christy</td>
<td>Calico</td>
<td><a href="mailto:christy_avery@unc.edu">christy_avery@unc.edu</a></td>
</tr>
<tr>
<td>Hodonsky, Chani</td>
<td>Calico</td>
<td><a href="mailto:chani_hodonsky@unc.edu">chani_hodonsky@unc.edu</a></td>
</tr>
<tr>
<td>Wojcik, Genevieve</td>
<td>Stanford / CC</td>
<td><a href="mailto:gwojcik@stanford.edu">gwojcik@stanford.edu</a></td>
</tr>
<tr>
<td>Best, Lyle</td>
<td>SHS-FAM</td>
<td><a href="mailto:lbest@restel.com">lbest@restel.com</a></td>
</tr>
</tbody>
</table>
II. SCIENTIFIC RATIONALE  (Please be specific and concise)

Total white blood cell (WBC) counts are often used by clinicians to understand the nature of an ongoing inflammatory process. Abnormally high (leukocytosis) or low (leukopenia) total WBC counts may provide evidence that supports or argues against a diagnosis or informs clinical care. Leukocytosis is usually an indication of an ongoing inflammatory process, and can be observed in a normal (eg. pregnancy, exercise) or abnormal (eg. infection, bone tumor, leukemia) health state. Similarly, leukopenia occurs in a number of different conditions that damage or suppress production of lymphocytes, such as lupus, HIV infection, and cancer chemotherapies.

A WBC differential describes the counts of the five WBC subtypes (neutrophils, lymphocytes, monocytes, basophils, and eosinophils) and often sheds light on the nature of the inflammatory process. For example, high neutrophil counts are often seen in bacterial infections when neutrophils attack invading pathogens. Neutrophil counts may also be used to assess bone marrow health following chemotherapy administration, and for adjusting the dosage and timing of subsequent cycles of therapy. CD4+ T-lymphocyte counts are often monitored in AIDS patients as a proxy for viral load, since active HIV replication destroys these cells. Eosinophils are involved in regulating allergic responses and combatting parasites, thus are often elevated in patients with allergic asthma or parasitic infections. Given that abnormal WBC counts help physicians diagnose and monitor an inflammatory process associated with disease, it is equally important to identify genetic traits that determine normal WBC counts.

WBC counts are moderately heritable, with most estimates around 0.41-3. However, normal counts can be highly variable between healthy individuals, with heterogeneity by age, sex and race/ethnicity thoroughly documented4-8. Some of these differences are likely associated with the prevalence of genetic traits in certain racial/ethnic populations. For example, a regulatory variant in the Duffy Antigen Receptor of Chemokine (DARC) gene is believed to be responsible for “benign ethnic neutropenia,” a term describing the persistently low neutrophil counts often observed in healthy individuals with African ancestry9-11.

Previous GWAS studies have identified over 30 genome-wide significant associations with total WBC count or a WBC subtype, in European12-18, Asian10;18-22, Hispanic10 and African10;15;17;18;23 ancestry populations. In this project, we aim to perform GWAS and fine-mapping studies of total WBC count and WBC differential counts among African Americans, Native Americans, Asian/Pacific Islanders and Hispanic/Latinos from the PAGE II consortium. Given that differences in WBC counts exist across racial/ethnic groups, we anticipate that the diverse study population of PAGE II coupled with genotyping on the Multi-Ethnic Genotyping Array (MEGA), which was designed for improved variant coverage across...
multiple US minorities, will increase our ability to detect population-specific traits that may be driving racial/ethnic differences in our cohort. Additionally, to the best of our knowledge, this analysis will be the first GWAS of WBC traits to include Native American participants, and will also include the largest sample of Hispanic/Latinos, who can have high levels of Native American ancestry. The ultimate goal of this analysis is to use an ancestrally diverse study population to identify novel WBC loci and expand our understanding of genetic traits that regulate hematopoiesis or contribute to circulating cell counts in healthy non-European individuals.

III. OBJECTIVES AND PLAN (Please be specific and concise)

a. Study Questions/Hypotheses.
1. Discovery of novel common and rare genetic variants influencing total WBC and WBC differential counts by testing all variants individually, and grouped for rarer variants, on MEGA as well as imputed variants.
2. Fine-mapping and replication of variants found in Aim 1 and known WBC loci previously identified in other studies and populations.
3. Identify pleiotropic associations across various hematological traits by comparing results from the 5 WBC differential counts, characterize variants that may be clinically relevant with regards to ethnicity-specific reference ranges, and use PrediXcan to infer gene expression in statistically significant variants.

b. Study populations, study design for each
This analysis will include all African Americans, Native Americans, Asian/Pacific Islanders and Hispanic from the PAGE II that had total WBC and WBC differentials and were genotyped on the MEGA array or are planning to impute variants from GWAS data. The anticipated studies include:

<table>
<thead>
<tr>
<th>Study</th>
<th>on_MEGA == TRUE &amp; age != NA</th>
<th>on_MEGA == FALSE &amp; age != NA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (median (min-max), in 10^9 cells/L)</td>
<td>n (median (min-max), in 10^9 cells/L)</td>
</tr>
<tr>
<td>All Studies</td>
<td>Total WBC count: 26,860 (6.0 (1.2-408))</td>
<td>Total WBC count: 60,460 (5.8 (1.2-801))</td>
</tr>
<tr>
<td></td>
<td>Neutrophil count: 18,059 (3.5 (0.0-26.7))</td>
<td>Neutrophil count: 22,155 (3.4 (0.0-24.4))</td>
</tr>
<tr>
<td>ARIC</td>
<td>Total WBC count:</td>
<td>Total WBC count:</td>
</tr>
<tr>
<td></td>
<td>EU: 27 (7.3 (4.1-18.7))</td>
<td>EU: 11,273 (6.0 (1.2-70.5))</td>
</tr>
<tr>
<td></td>
<td>AA: 3 (5.1 (4.7-12.0))</td>
<td>AA: 3,941 (5.3 (1.9-42.0))</td>
</tr>
<tr>
<td></td>
<td>Neutrophil count:</td>
<td>Neutrophils count:</td>
</tr>
<tr>
<td></td>
<td>EU: 24 (5.1 (2.3-15.5))</td>
<td>EU: 7,901 (3.6 (0.0-19.6))</td>
</tr>
<tr>
<td></td>
<td>AA: 3 (2.2 (1.8-9.1))</td>
<td>AA: 3,929 (2.5 (0.2-23.1))</td>
</tr>
<tr>
<td>BioMeTM</td>
<td>Total WBC count:</td>
<td>Total WBC count:</td>
</tr>
<tr>
<td></td>
<td>EU: 3 (6.0 (5.4-9.1))</td>
<td>EU: 609 (6.7 (2.0-31.0))</td>
</tr>
<tr>
<td></td>
<td>AA: 2,170 (6.1 (1.5-20.4))</td>
<td>AA: 941 (5.8 (1.7-23.0))</td>
</tr>
<tr>
<td></td>
<td>HA: 2,598 (6.8 (1.3-24.1))</td>
<td>HA: 2,217 (6.9 (1.8-25.4))</td>
</tr>
<tr>
<td></td>
<td>AS: 130 (6.3 (3.1-16.5))</td>
<td>AS: 3 (5.9 (4.3-7.4))</td>
</tr>
<tr>
<td></td>
<td>NA: 20 (7.6 (4.1-9.3))</td>
<td>NA: 9 (6.8 (3.6-8.6))</td>
</tr>
<tr>
<td></td>
<td>OT: 390 (6.8 (3.0-17.7))</td>
<td>OT: 73 (6.5 (2.5-16.5))</td>
</tr>
<tr>
<td></td>
<td>Neutrophil count:</td>
<td>Neutrophils count:</td>
</tr>
<tr>
<td></td>
<td>EU: 3 (4.4 (3.3-6.2))</td>
<td>EU: 263 (4.2 (1.1-24.0))</td>
</tr>
<tr>
<td></td>
<td>AA: 1,634 (3.5 (0.5-17.2))</td>
<td>AA: 689 (3.3 (0.2-17.1))</td>
</tr>
<tr>
<td></td>
<td>HA: 1,979 (4.1 (0.0-20.2))</td>
<td>HA: 1,444 (4.2 (0.2-21.4))</td>
</tr>
<tr>
<td></td>
<td>AS: 89 (3.8 (1.1-15.3))</td>
<td>AS: 3 (3.4 (2.4-5.5))</td>
</tr>
<tr>
<td></td>
<td>NA: 11 (3.9 (1.3-6.2))</td>
<td>NA: 7 (3.3 (1.4-5.1))</td>
</tr>
<tr>
<td></td>
<td>OT: 205 (4.0 (1.1-12.8))</td>
<td>OT: 39 (3.9 (0.9-7.5))</td>
</tr>
<tr>
<td>CARDIA</td>
<td>Total WBC count:</td>
<td>Total WBC count:</td>
</tr>
<tr>
<td></td>
<td>EU: -</td>
<td>EU: 2,463 (6.0 (2.1-16.8))</td>
</tr>
<tr>
<td></td>
<td>AA: -</td>
<td>AA: 2,590 (5.6 (1.8-22.5))</td>
</tr>
<tr>
<td></td>
<td>Neutrophil count:</td>
<td>Neutrophils count:</td>
</tr>
<tr>
<td></td>
<td>EU: -</td>
<td>EU: 1,739 (3.2 (0.1-10.7))</td>
</tr>
<tr>
<td></td>
<td>AA: -</td>
<td>AA: 1,671 (2.9 (0.0-14.2))</td>
</tr>
<tr>
<td>HCHS/SOL</td>
<td>Total WBC count:</td>
<td>Total WBC count:</td>
</tr>
<tr>
<td></td>
<td>HA: 11,202 (6.3 (1.2-31.8))</td>
<td>HA: 44 (6.1 (3.1-13.6))</td>
</tr>
<tr>
<td></td>
<td>Neutrophil count:</td>
<td>Neutrophils count:</td>
</tr>
<tr>
<td></td>
<td>HA: 11,199 (3.5 (0.0-26.7))</td>
<td>HA: 44 (3.4 (0.2-9.0))</td>
</tr>
</tbody>
</table>
c. Variant/SNPs (Specify)
All MEGA variants that pass quality control filters (to be performed at PAGE CC/UW) will potentially be included. The MEGA array includes nearly 2 million SNPs, including 1.4 million tag SNPs and 400,000 exomic SNPs, specifically selected to assess genetic variation in US minority groups (African Americans, Hispanic/Latinos, Asians, and Native Americans). Genotyping on the Metabochip, Exomechip, and various GWAS arrays are available for many studies participating in PAGE and will be included in this analysis as appropriate (e.g., replication, meta-analysis). For samples with GWAS data but no MEGA data, variants will be imputed into 1000 Genomes, with all successfully (r^2>0.3) imputed variants eligible for inclusion.

d. Phenotype(s) (Specify)
1. Total White Blood Cell (WBC) count
2. Neutrophil count
3. Basophil count
4. Eosinophil count
5. Monocyte count
6. Lymphocyte count
   - Should all be reported in 10^9 cells / L (continuous)
   - Will be log(x+1) transformed to account for skewed distribution, and the large number of measured cell count values of 0. Shapiro-Wilks tests of normality will be used to evaluate normality of pre- and post-transformed data. Square root transformations are another potential strategy for correcting skewed distributions.
   - Preliminary examination of the distribution of counts may warrant additional investigation on outliers which may need to be excluded as potential data entry errors or evidence of an ongoing inflammatory event. Depending on the phenotype, subjects with values > 2 standard deviations from the mean may be excluded, or a sensitivity analysis (excluding the top and bottom 5% of observations) may be conducted.
   - Preliminary examination of the basophil and eosinophil counts show a large number of participants with cell count values of 0. We may consider dichotomizing the variable (0 cells detected vs > 0 cells detected) and analyzing with a logistic regression instead.
e. Covariates (Specify)
   1. Age (in years, continuous)
   2. Age^2 (to account for non-linear associations with the phenotype)
   3. Sex (M/F, binary)
   4. BMI (in kg/m^2, continuous), natural log transformed
   5. Current smoker at time of sample collection (Y/N, binary)
   6. Self-identified race/ethnicity (indicator/dummy variables)
   7. Global ancestry (estimated via top 8 principal components)
   8. Study center (indicators/dummy variables, if applicable)
   9. Socioeconomic status indicators (education, income), if available

Global ancestry: Principal components analysis with Eigensoft will be conducted centrally for all studies using unrelated individuals. The top 8 ancestry informative principal components (PCs) will be used in regression models to control for population substructure. We will include HapMap, 1000 Genomes, and other groups to improve interpretation.

For rare variants, there is concern that methods for controlling for population stratification using common variants is not appropriate. It is uncertain how applicable this concern is to study of complex traits in modern-day outbred populations as opposed to composites of highly-stratified endogamous groups localized geographically. Keeping this concern in mind, we will keep up with the literature and be careful of over-interpreting the results for very rare variants. We may also consider modifying our use of PCs, (perhaps increasing the number of PCs used in rare variant versus common variant analyses) or modify linear mixed models (LMMs), or generalized estimating equations (GEEs), possibly by including rarer SNPs in the calculation of the genetic relatedness matrix.

f. Main statistical analysis methods

Aim 1 analysis: Discovery of novel common and rare genetic variants influencing WBC and WBC differential counts

Common variants:

Association testing: Common variants will be defined as MEGA autosomal variants and imputed 1000 Genomes variants with MAF ≥ 1% in one or more of our PAGE racial/ethnic groups. However, if the 1% cut-off yields a race/ethnic subgroup with a small sample size, with fewer than 100 total observations, or fewer than 20 carrying the minor allele, the SNP will be excluded from that subgroup analysis. Single-variant linear regressions will be conducted on each common variant, with all race/ethnicities pooled and adjusted by both self-identified race/ethnicity and the top 8 ancestry PCs, in addition to recognized confounders (age, sex, BMI, smoking status). Confounders, such as BMI, that have a skewed distribution will the natural log transformed. The statistical significance of an association will be the standard GWAS cut-off of 5x10^{-8}, and a p-value of 5x10^{-6} representing suggestive associations warranting examination in the replication and combined discovery + replication analysis.

Analyses stratified by race/ethnicity (and adjusted by study), and then meta-analyzed, will be considered as a supplementary analysis. As observed in previous PAGE meetings, the variation in genetic ancestry (as assessed by PCs) within some racial/ethnic subgroups (Hispanic/Latinos, in particular), is nearly as extensive
when compared to a sample where all racial/ethnic subgroups are pooled together. Based on this observation, it has been posited that stratification by self-identified race/ethnicity does not always create homogeneous groups with similar patterns of genetic ancestry and therefore is not an effective strategy for controlling confounding, while PCs capture population substructure regardless of self-identified race/ethnicity. We do acknowledge that ethnicity-specific social and behavioral traits may still represent a potential source of confounding, and further note that previous studies on variation of WBC were typically based on self-identified race/ethnicity. Therefore, we will include self-identified race/ethnicity as a covariate in the pooled models, and report results from race/ethnicity stratified meta-analyses to facilitate comparison with prior findings. The race/ethnicity subgroups will include:

1) African Americans (AA)  
2) Native Americans (NA)  
3) Asian/Pacific Islanders (AS)  
4) Hispanic (HA)  
5) European Americans (EA)

The European Americans available within the PAGE cohorts will be used to replicate any specific findings. Naturally we will need other replication cohorts. Jointly with writing groups for other PAGE phenotypes, we are currently exploring collaborations with other cohorts. Which cohorts are best for such replication depends largely on for which ethnicities we identify new associations. Overall associations by combining results from the discovery and replication samples will also be reported. We will consider replication to be successful if the association p-value in the replication population meets the Bonferroni correction of 0.5 / (# of SNPs identified for replication), and/or if the discovery+replication p-value is genome-wide significant (p<5x10^{-8}). Alternatively, we may also consider using a Benjamini-Hochberg adjustment to control the false discovery rate, as a less conservative approach for accounting for multiple comparisons.

Other studies that have been used to replicate / compare WBC findings may include:

1) CHARGE consortium: Cohorts for Health and Aging Research in Genomic Epidemiology. Includes 5 cohort studies from US and Europe, n=38,000. (a) Age, Gene/Environment Susceptibility – Reykjavik Study, (b) ARIC, (c) CHS, (d) Framingham, (e) Rotterdam Study.  
2) RIKEN: Japanese, n>17,000 from the BioBank Japan Projects  
3) COGENT: Continental Origins and Genetic Epidemiology Network, n>16,000 self-identified African Americans from 7 cohorts: Atherosclerosis Risk in Communities (ARIC), Coronary Artery Risk Development in Young Adults (CARDIA), John Hopkins Genetic Study of Atherosclerosis Risk (GeneSTAR), Healthy Aging in Neighborhoods of Diversity Across the Life Span (HANDLS), Healthy Aging and Body Composition (Health ABC), Jackson Heart Study (JHS), and Women’s Health Initiative (WHI).  
4) Blood Cell Consortium (BCX): ExomeChip genotyping for ~157,622 multi-cultural individuals from 25 different studies (some overlap with PAGE studies).  
http://www.mhi-humangenetics.org/en/resources

Models will be minimally adjusted for age, sex, BMI, smoking status, top 8 principal components of genetic ancestry, and any necessary study-specific variables (such as study center or region). Each variant will be coded assuming an additive model (with 0, 1, or 2 copies of the coded allele). The coded allele will be defined by the primary analyst and sent to all analysts to prevent strand flipping issues for analyses conducted by studies where centralized data-sharing cannot occur.
Model: \( \log(1 + \text{WBC count}) \sim \text{genotype} + \text{covariates} + \text{global PCs} \)

To account for unequal inclusion probabilities and complex patterns of relatedness in HCHS/SOL, we will use the MMAAPS program developed by Conomos et al.\(^{24}\), which utilizes generalized linear mixed models. We will include a variance component for household membership.

Trans-Ethnic Meta-analysis: The overall association with WBC or WBC differential counts for each variant will be obtained by combining the beta-estimate and standard errors from each race/ethnicity within each study using an inverse-variance weighted fixed-effect meta-analyses. Cochran’s Q statistic and \( I^2 \) will be calculated to measure between-study heterogeneity. Any cross-race/ethnicity analyses may utilize additional methodologies. Excessive heterogeneity in the trans-ethnic meta-analyses may be indications of unique, population-specific SNP-WBC associations.

Rare variants

Rare variants will be defined within each racial/ethnic population, where a variant is considered rare if the MAF is reported as less than 1% in the corresponding PAGE racial/ethnic group. Single variant tests based on score statistics will be conducted for rare variants where the A2 minor allele is observed in a sufficient number (n>100) of participants (score statistic tests can be better powered than the Wald test).

We will also aggregate rare autosomal variants by gene or loci to conduct burden and/or non-burden association tests. Rare variants will be mapped to genes or loci based on the appropriate build using annotations from the UCSC Table Browser. As tissue specific enhancer definitions (and gene-enhancer links using gene expression) will become available, we will also expand aggregated analyses to those in relevant tissues. We will use the Sequence Kernel Association Test (SKAT),\(^{25}\) the optimal SKAT test (SKAT-O)\(^{26}\) or alternative methods to perform this rare-variant association testing on unrelated individuals. We will run the models separately for each race/ethnic group. Equal weights will be applied to all variants or variants may be weighted by MAF or functional scores (such as CADD). A Bonferroni correction for the number of genes tested will be applied.

Aim 2 analysis: Fine-mapping and replication of known WBC loci previously identified in other studies and populations.

Fine mapping: Variants that demonstrate a (genome-wide) statistically significant association with WBC will be utilized as an index variant for fine-mapping. It is expected that variants associated with WBC in other racial/ethnic groups will be correlated with the index variant found in previous studies. Therefore, for each of the index variants identified above, we will identify all SNPs that are correlated (\( r^2 > 0.2 \)) with the original index variant in that region, using the CEU (or whatever population is relevant for the original GWAS identifying the locus) population information from the 1000 Genome Project. Results for these regions will be graphically displayed using LocusZoom. To test for multiple independent signals within a single locus, we will perform a series of sequential conditional analyses at loci that showed evidence of association at \( p < 0.05/(\# \text{ of variants at the locus}) \), where the most significant variant and a correlated variant of interest will be included in the same model. P-values and changes in effect estimates will be evaluated to investigate which variant shows a stronger signal and to comment on the more likely functional variant.
Second signals: We will search for independent second signals among variants that are in the same fine-mapping region but are not correlated with the index variant ($r^2 < 0.2$). Within each region, the statistical significance of potential second signals will be determined by a region-specific Bonferroni-correction (0.05 / # correlated variants in that region). Conditional analyses will again be performed to identify whether the top signals in the region appear to be independent. Conditional analyses will be performed adjusting for successive variants until no variants with p-values lower than the Bonferroni-corrected threshold remain.

Aim 3 analysis: Identify pleiotropic associations across various hematological traits by comparing results from the 5 WBC differential counts, characterize variants that may be clinically relevant with regards to ethnicity-specific reference ranges, and use PrediXcan to infer gene expression in statistically significant variants.

Previously, several genetic loci associated with hematological traits have been observed to have pleiotropic associations with other hematological traits (Nalls, 2011 #114). For example, 38 SNPs in 17q21 were associated with both total WBC count and neutrophil count independently based on conditional analyses. Other SNPs were independently associated with both monocyte and basophil counts (3q21), and total WBC count and lymphocyte counts (6p21). We will compare findings from the 5 WBC subtypes (neutrophil, basophil, eosinophil, lymphocyte, and monocyte counts) to replicate previously identified pleiotropic associations, and identify novel pleiotropic associations. Should we find evidence of pleiotropic associations, we will survey the literature and use an appropriate analysis to verify the association, and assess the relative impact a SNP has on various WBC phenotypes.

Over 17,000 variants on the MEGA are relevant to clinical and pharmacogenetic studies. We will pay close attention to variants that may be associated with population-specific total WBC or WBC differential counts. For example, we hope to replicated the association in the DARC gene, which has been theorized to cause the persistently low neutrophil counts among health African Americans9; 10. Since other race/ethnicity differences in WBC subtype counts have previously been reported, we hope to uncover additional loci that may help to explain differences seen in other groups.

Finally, we will use PrediXcan27 to impute gene expression among PAGE II participants. Briefly, the PrediXcan software predicts gene expression based on each participant’s genotypes by using weights previously developed from various databases that include studies which have measured both genotypes and gene expression in various tissue types. The predicted levels of gene expression are assessed for their associations with the outcome of interest, in this case WBC and WBC differential counts. It is our hope that PrediXcan results, coupled with information from the functional annotation and clinically relevant variant databases currently being assembled, will provide additional information about how WBC loci impact biological pathways that regulate WBC counts.

We also plan to work closely with the WBC admixture analysis, currently being led by Genevieve Wojcik.

g. Ancestry information used? No ___ Yes x_ How is it used in the analyses?
We will adjust for population substructure with global ancestry estimates (i.e. PCs).

h. Anticipated date of draft manuscript to P&P: 2017
i. What manuscript proposals listed on www.pagestudy.org/index.php/manuscripts/ are most related to the work proposed here? Approved PAGE ms. numbers: __M08__

   If any: Have the lead authors of these proposals been contacted for comments and/or collaboration? Yes ☒ No __

IV. SOURCE OF DATA TO BE USED (Provide rationale for any data whose relevance to this manuscript is not obvious): Check all that apply:

Genotype data
We will use ~1.7 million variants on the MEGA array. Genotyping on the MetaboChip, ExomeChip, and various GWAS arrays are available for many studies participating in PAGE and will be included in this analysis as appropriate (e.g., replication, meta-analysis).

Aggregate/summary data to be generated by investigators of the study(ies) mentioned:
[X] ISMMS; [x] CALiCO; [x] MEC; [x] WHI; [ ] CC; [ ] Other: ________________
If CALiCO, specify [ ] ARIC; [x] CARDIA; [x] SHS-Fam; [ ] SHS-Cohort; [x] SOL

I, _(KN)_, affirm that this proposal has been reviewed and approved by all listed investigators.

V. REFERENCES
people of African descent is due to a regulatory variant in the Duffy antigen receptor for chemokines gene. PLoS Genet 5, e1000360.
nine novel loci associated with white blood cell subtypes in a Japanese population.

22. Kong, M., and Lee, C. (2013). Genetic associations with C-reactive protein level and

23. Ramsuran, V., Kulkarni, H., He, W., Mlisana, K., Wright, E.J., Werner, L., Castiblanco,
counts influence HIV-1 susceptibility in high-risk South African black women. Clin
Infect Dis 52, 1248-1256.

Diversity and Association Studies in US Hispanic/Latino Populations: Applications in
the Hispanic Community Health Study/Study of Latinos. Am J Hum Genet 98, 165-
184.

association testing for sequencing data with the sequence kernel association test.
Am J Hum Genet 89, 82-93.

26. Lee, S., Emond, M.J., Bamshad, M.J., Barnes, K.C., Rieder, M.J., Nickerson, D.A.,
unified approach for rare-variant association testing with application to small-sample

A gene-based association method for mapping traits using reference transcriptome
data. Nat Genet 47, 1091-1098.

VI. IF USING SOL DATA (Please provide the information below)
   a. Keywords: White Blood Cell (WBC) counts
   b. Using biomarker data? Yes _x_ No __
   c. Where will the SOL data analyses be performed? On the cloud, unless this is not allowed by the
      study. If so, we will run these at the Fred Hutchinson Cancer Research Center, a PAGE data sharing
      agreement is in place for this.