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

PAGE Ms. Number: ______ Submission Date: Feb 1, 2016  [Approval Date: ______]

Title of Proposed MS: Investigation of the Genetic Architecture Underlying Kidney Traits Among African-Americans and Hispanics Using the MEGA Array in the Population Architecture Using Genomics and Epidemiology (PAGE) Study

Abbreviated Title of Proposed MS: Genetics of Kidney Traits in PAGE

I. INVESTIGATOR INFORMATION:

Name of Lead Author: Girish Nadkarni
Email Address: girish.nadkarni@mountsinai.org
Telephone Number:

Name of Corresponding Author (if different):
Christina Wassel
Email Address: cwassel@med.uvm.edu
Telephone Number: 802-656-1970

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

<table>
<thead>
<tr>
<th>N, N</th>
<th>Affiliation in PAGE</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kari E North</td>
<td>CALiCo-ARIC</td>
<td><a href="mailto:Kari_north@unc.edu">Kari_north@unc.edu</a></td>
</tr>
<tr>
<td>Myriam Fornage</td>
<td>CALiCo-CARDIA</td>
<td><a href="mailto:Myriam.Fornage@uth.tmc.edu">Myriam.Fornage@uth.tmc.edu</a></td>
</tr>
<tr>
<td>Steve Buyske</td>
<td>CC</td>
<td><a href="mailto:buyske@stat.rutgers.edu">buyske@stat.rutgers.edu</a></td>
</tr>
<tr>
<td>Girish Nadkarni</td>
<td>ISMMS</td>
<td><a href="mailto:girish.nadkarni@mountsinai.org">girish.nadkarni@mountsinai.org</a></td>
</tr>
<tr>
<td>Christina L Wassel</td>
<td>CALiCo-SOL</td>
<td><a href="mailto:cwassel@med.uvm.edu">cwassel@med.uvm.edu</a></td>
</tr>
<tr>
<td>Fred Schumacher</td>
<td>MEC</td>
<td><a href="mailto:frs2@case.edu">frs2@case.edu</a></td>
</tr>
<tr>
<td>Charles Kooperberg</td>
<td>WHI</td>
<td><a href="mailto:clk@fredhutch.org">clk@fredhutch.org</a></td>
</tr>
<tr>
<td>Melissa Richard</td>
<td>CALiCo-CARDIA</td>
<td><a href="mailto:Melissa.A.Lee@uth.tmc.edu">Melissa.A.Lee@uth.tmc.edu</a></td>
</tr>
<tr>
<td>Cynthia Bell</td>
<td>CALiCo-CARDIA</td>
<td><a href="mailto:Cynthia.Bell@uth.tmc.edu">Cynthia.Bell@uth.tmc.edu</a></td>
</tr>
<tr>
<td>Lani Park</td>
<td>MEC</td>
<td><a href="mailto:sungship@med.usc.edu">sungship@med.usc.edu</a></td>
</tr>
</tbody>
</table>
II. SCIENTIFIC RATIONALE

Chronic kidney disease (CKD), defined by a reduced kidney function (GFR) or presence of increased albuminuria, is an emerging public health problem with an estimated prevalence of 14.5% of the adult US population. CKD is a strong risk factor for cardiovascular disease, is associated with increased morbidity and mortality, increased health care resource use and a decreased quality of life. Hypertension and type 2 diabetes are the most common causes of CKD and other known risk factors for kidney disease are metabolic syndrome, hyperlipidemia and smoking. CKD disproportionately affects minorities, who also more often progress to end-stage renal disease (ESRD) than those of European descent. African-Americans have a disproportionate burden of CKD and albuminuria compared to non-Hispanic whites, and Mexican Americans have similar prevalence of CKD and albuminuria compared to non-Hispanic whites. However, little data exists on other Hispanic subgroups such as Puerto Ricans, Dominicans, and Cubans.

Familial aggregation studies suggest a strong genetic susceptibility to CKD. Heritability of eGFR and albumin-creatinine ratio (ACR) is high across race/ethnic groups, ranging from 36-75% for eGFR and 16-46% for ACR. To date, a number of significant genetic loci have been discovered and successfully replicated in GWAS for a variety of kidney related traits in European populations. In African-Americans, fewer loci have been identified and replicated, either in GWAS or admixture mapping studies. However, among Hispanics, few genetic variants or loci have been discovered for kidney traits, and existing studies are primarily from candidate gene and linkage studies. Genetic admixture has been significantly associated with albuminuria among Hispanics, indicating that the ancestral background could be an important indicator of kidney function, and that there may be genetic variants not yet discovered underlying renal function among Hispanic groups.

Recently discovered loci appearing to play a large, significant role in kidney disease, especially among minority race/ethnic groups, include APOL1 and MYH9. APOL1 alleles are strongly associated with CKD progression in African Americans. MYH9, an already well-known locus for kidney disease in African-Americans is also associated with non-diabetic ESRD among Hispanic Americans.

III. OBJECTIVES AND PLAN

a. Study Questions/Hypotheses.

Using the newly developed PAGE II MEGA Array, we will determine whether there are novel genetic variants influencing levels of estimated GFR and ACR, as well as risk for CKD. We will additionally fine-map any known kidney traits loci previously identified in populations primarily of European ancestry.
b. Study populations, study design for each

Primary analyses will be performed in PAGE II African-American and Hispanic study populations with MEGA Array data and measured serum creatinine, urine albumin/urine creatinine ratio (ACR), estimated glomerular filtration rate (GFR), and chronic kidney disease (CKD).

c. Variant/SNPs

The MEGA Array is a custom chip containing ~2M SNPs, of which about 200k are selected to capture genetic variation at established CVD-associated loci across racially and ethnically diverse populations. The tagSNP (1.4M) and exomic (400k) content of the MEGA Array have been selected specifically to capture the genetic variation of African, Hispanic, European, Americas, and Asian populations, while providing whole genome and candidate region coverage. However, genotyping on the Metabochip, Exome Chip, and various GWAS arrays are available for many studies participating in PAGE II and will be included in this analysis as appropriate (e.g., replication, meta-analysis). Additionally, non-PAGE replication studies (i.e. the Multi-Ethnic Study of Atherosclerosis (MESA)), will be sought out as needed, and these studies are also genotyped a variety of arrays.

d. Phenotype(s)

Quantitative traits are estimated GFR (calculated from serum creatinine, age, race and sex by CKD-Epi equation\(^{40}\), urine albumin to urine creatinine ratio (ACR). The CKD-Epi equation has been previously validated in Hispanics by Stevens et al\(^{41}\).

Secondary phenotypes are chronic kidney disease (eGFR≤60 ml/min/1.73m\(^2\) vs eGFR>60 ml/min/1.73m\(^2\) and albuminuria (≥30mg/g vs <30 mg/g). In MEC and ISMMS, CKD will be defined using ICD-9 codes 585.1, 585.2, 585.3, 585.4, 585.5, 585.9 or ICD-10 codes N18.1, N18.2, N18.3, N18.4, N18.5, and N18.9. We will also include eGFR MDRD as a secondary trait, as many previous genetic kidney traits studies have been published using eGFR MDRD, and it may be of interest to compare our PAGE results to these previously published studies.

e. Covariates

Age, Sex, Field site (if necessary), principal components will be the primary covariates. We will also adjust for local ancestry estimates if available. Additional levels of adjustment for other kidney disease risk factors, i.e. smoking, blood pressure, diabetes) may be performed to determine whether any loci identified for kidney traits are actually working through another closely associated trait or disease (i.e. mediation).

f. Main statistical analysis methods

Single variant tests will be conducted on variants with MAF ≥1% and an additive genetic model. Analyses will be conducted stratified by race/ethnic group. For continuous outcomes, i.e. eGFR and ACR, linear regression will be used, while for binary traits, i.e. CKD, logistic regression will be used. Covariates and PCs will be included in these models, as described in section e above. SOL analyses will require adjustment for the complex sampling scheme. SUGEN\(^{42}\), a generalized estimating equation (GEE) approach, has been developed that uses sampling weights and robust variance estimators to control for sampling design and partially control for relatedness due to household effects. To additionally control for relatedness due to endogamous mating, SUGEN
allows for a kinship working correlation matrix to be incorporated. Alternatively, generalized linear mixed models (GLMM) may be used to account for the analytic challenges addressed by SUGEN. GEE or GLMM will be used to analyze SOL data, depending on the recommendations of the SOL analysis committee at the time of analysis, adjusted as described for the models above.

Finally, we will perform trans-ethnic inverse variance-weighted fixed effects meta-analysis in METAL to identify loci with comparable effects across the PAGE-II studies and races. For SNPs with an agnostic a priori hypothesis, the standard GWAS cut-off of $5 \times 10^{-8}$ will be used to define statistical significance. When loci in candidate regions for eGFR or CKD are tested, a Bonferroni correction for the number of candidate loci will be applied.

Rare variants tests (MAF<1% will be conducted in unrelated individuals using SKAT\(^43\) to assess the association of regions in genome-wide and exome-wide genotyping. In SOL, FBAT\(^44\) will be used to account for relatedness in the cohort. A SUGEN implementation of SKAT is under development and will be used if available at the time of analysis. Variants with MAFs <1% will be mapped to genes based on the appropriate build using annotations from the UCSC Table Browser or MySQL server (https://genome.ucsc.edu/cgi-bin/hgTables). Since primary analyses will be restricted to rare variants, equal weights will be applied to all variants. A Bonferroni correction for the number of regions tested will be applied.

In addition to analyses of only rare variants, intragenic common variants identified in single-variant analyses will be tested for joint gene-based effects with rare variants. To account for the incorporation of more common variants, Madsen-Browning weighting and the default SKAT upweighting of low frequency variants will be applied. Burden tests and variable threshold (VT) tests will be conducted for each PAGE-II race/cohort in addition to the primary rare variant analyses. The number of rare variant minor alleles at loci with MAF <1% and 5% will summed for each individual and regressed on the phenotype for the T1/T5 test. The VT test will be conducted assuming a fixed but unknown MAF threshold exists for causal variants in a given region. A statistic will be computed for each possible MAF threshold and the threshold for that region that provides the strongest association signal will be selected as the evidence of association in that region.

**g. Ancestry information used? No __ Yes _X_ How is it used in the analyses?** To control for population stratification

**h. Anticipated date of draft manuscript to P&P:** ___Fall/Winter 2016-17________

**i. What manuscript proposals listed on www.pagenetstudy.org/index.php/manuscripts/ are most related to the work proposed here?** Approved PAGE ms. numbers: 47 ____ ____ ____

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

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

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 [ X ] ARIC; [ X ] CARDIA; [ ] SHS-Fam; [ ] SHS-Cohort; [X ] SOL
I, CLW, affirm that this proposal has been reviewed and approved by all listed investigators.

V. REFERENCES


**VI. IF USING SOL DATA** (Please provide the information below)

a. **Keywords:** kidney, genotype, race/ethnicity, creatinine, estimated glomerular filtration rate, albumin-creatinine ratio

b. **Using biomarker data?** Yes ___ No ___ X __

c. **Where will the SOL data analyses be performed?** Centrally, in cloud computing platform the PAGE CC will set up