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 Purn@biology.rutgers.edu

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

Title of Proposed Ms.: Analysis of Familial Combined Dyslipidemia predisposing factors

Name of Lead Author: Irene Predazzi

Name of Corresponding Author (if different): Will Bush
Email Address: William.s.bush@vanderbilt.edu Telephone Number: 615-343-8608

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

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charles Kooperberg</td>
<td>WHI</td>
<td><a href="mailto:clk@fhcrc.org">clk@fhcrc.org</a></td>
</tr>
<tr>
<td>Cara Carty</td>
<td>WHI</td>
<td><a href="mailto:ccarty@fhcrc.org">ccarty@fhcrc.org</a></td>
</tr>
<tr>
<td>Dana Crawford</td>
<td>EAGLE</td>
<td><a href="mailto:crawford@chgr.mc.vanderbilt.edu">crawford@chgr.mc.vanderbilt.edu</a></td>
</tr>
<tr>
<td>Will Bush</td>
<td>EAGLE</td>
<td></td>
</tr>
<tr>
<td>Robert A. Wild</td>
<td>WHI</td>
<td><a href="mailto:Robert-wild@ouhsc.edu">Robert-wild@ouhsc.edu</a></td>
</tr>
<tr>
<td>Lisa W. Martin</td>
<td>WHI</td>
<td><a href="mailto:lwmartin@mfa.gwu.edu">lwmartin@mfa.gwu.edu</a></td>
</tr>
<tr>
<td>Jennifer H. Lin</td>
<td>WHI</td>
<td><a href="mailto:jhlin@rics.bwh.harvard.edu">jhlin@rics.bwh.harvard.edu</a></td>
</tr>
<tr>
<td>Jian Gong</td>
<td>WHI</td>
<td><a href="mailto:jgong@fhcrc.org">jgong@fhcrc.org</a></td>
</tr>
<tr>
<td>Jennifer Robinson</td>
<td>WHI</td>
<td><a href="mailto:jennifer-g-robinson@uiowa.edu">jennifer-g-robinson@uiowa.edu</a></td>
</tr>
<tr>
<td>(TBA)</td>
<td>SOL</td>
<td></td>
</tr>
<tr>
<td>George Papanicolaou</td>
<td>SOL/NHLBI</td>
<td></td>
</tr>
</tbody>
</table>
Partner studies in PAGE not collaborating in this ms. proposal:

<table>
<thead>
<tr>
<th>Study</th>
<th>Contacted? Y/N</th>
<th>Declined? / Other?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Names, affiliations, email address of non-PAGE investigators proposed as co-authors:

Sergio Fazio, Vanderbilt University Medical Center, Atherosclerosis Research Unit, Division of Cardiovascular Medicine, 383 PRB-2220 Pierce Avenue, Nashville, TN 37232-6300, USA

Scott Williams, Department of Genetics, Geisel School of Medicine, Dartmouth College, 78 College ST, HB 6044, Hanover, NH 03755, USA

II. SCIENTIFIC RATIONALE (Please be specific and concise)

Familial combined Hyperlipidemia (FCHL) is the most prevalent genetic dyslipidemia in Western Society, with a prevalence of 1:100\(^1\) and is an established risk factor for premature coronary artery disease\(^2\). FCHL is defined by the combination of different lipid phenotypes, such as hypercholesterolemia, hypertriglyceridemia, and low HDL cholesterol levels. The phenotype is variable, with the affected subject switching from a high triglyceride (TG) to a high LDL syndrome depending on lifestyle and circumstances (weight gain/loss, alcohol use/abuse, drugs affecting lipids, etc.) and with family members showing different types of dyslipidemia (mostly high LDL in some, mostly high TG in others). However, HDL levels tend to be low in all cases.

Multiple defects have been associated with the etiology of FCHL:
- Adipose tissue dysfunction
- Hepatic overproduction of VLDL
- Multi-organ insulin resistance and hepatic fat accumulation
- Unbalanced metabolism of lipoprotein particles in plasma
- Impaired clearance of Apo-B containing particles\(^2\)

Several genetic studies have investigated the etiology of FCHL, using basically three different approaches:
- Candidate gene approach (considering the multiple defects previously listed)
- Linkage analysis
- Genome wide association studies\(^2\)

However, most of these focused on families of European descent, with some focus on Mexican and Asian populations\(^2\). African-descent populations have not been examined. None of these studies have successfully identified robust genetic factors for FCHL.

Furthermore, given the multiple lipid traits involved in the phenotype definition, it is much likely that genetic contribution to FCHL manifests in multiple pathways rather than as the exclusive manifestation of a single gene defect. To date, no studies have assessed how multiple genes and epistasis affect disease susceptibility.

III. OBJECTIVES AND PLAN (Please be specific and concise)
a. **Study Questions/Hypotheses.**

Under the general hypothesis that metabolic related variants confer FCHL we will address three specific but correlated questions that drive each one of the aims. The first is simply whether certain single variants are consistently associated to FCHL in multiple populations (Aim 1); the second is whether gene-gene interactions rather than gene variants alone are responsible for the insurgence of the disorder (Aim 2); and the third questions is whether different populations have different population-specific susceptibility loci to the FCHL, as opposed to the first question (Aim 3).

We propose to perform the following aims:

Aim 1 will determine whether MetaboChip variants are consistently associated to FCHL in multiethnic the PAGE studies.

Aim 2 will determine whether interactions between MetaboChip variants (modifier effects or epistasis) are associated to FCHL in multiethnic the PAGE studies.

Aim 3 will perform admixture mapping in a sample containing multiple ethnic groups using the variants in the MetaboChip.

The information regarding race/ethnicity will be used to stratify for ancestry in the analyses of both aim 1 and aim 2.

b. **Study populations, study design for each**

The proposed analyses will include all populations participating in the Population Architecture using Genomics and Epidemiology (PAGE) Study. The PAGE Study consists of four sites: CALiCo Consortium, Epidemiologic Architecture of Genes Linked to Environment (EAGLE), Multiethnic Cohort (MEC), and Women’s Health Initiative (WHI).

The CALiCo Consortium for this proposal includes ARIC and CHS with a total of approximately 58,000 men and women from diverse race/ethnic groups in the US, ranging in age from childhood to advanced age. MEC contains five major ethnic groups of older men and women in Hawaii and California. WHI contains over 68,000 postmenopausal women in the clinical trial and over 93,000 women in the observational study. These three studies represent traditional prospective epidemiological cohorts – a study design generally thought to be the gold standard for assessing the effects of environmental exposures and genetic factors.

The EAGLE study will exclusively use samples from BioVU, the Vanderbilt DNA biobank. BioVU is a repository of clinical blood samples linked to de-identified electronic medical records collected at Vanderbilt Medical Center in Nashville, TN. Race/ethnicity is determined through third-party report in the EMR and is confirmed using principal components analysis.

c. **Variant/SNPs (Specify)**

Given the paucity of knowledge regarding the phenotype of interest, all polymorphisms available on the Illumina MetaboChip$^3$ will be tested. For some analyses, we may restrict SNP selection to either common (MAF > 0.05) or rare (MAF < 0.05) alleles.

d. **Phenotype(s) (Specify)**

The phenotype of interest is designated Familial Combined Hyperlipidemia, as it has been shown to cluster in families. Because most PAGE studies are population-based, we will focus more broadly
on Combined Dyslipidemia, or Combined Hyperlipidemia, as a population-level analog of FCHL defined as a combination of the following parameters:
- Total cholesterol >= 200 mg/dl
- TG >= 150 mg/dl (fasting)
- HDL<= 40 mg/dl in men (<= 50 mg/dl in women)

Individuals having the combination of these parameters will be considered as cases, all the others will be considered as controls.

A parallel analysis will consider two continuous traits defined in the following way:
- HDL/TC (proportion)
- TC-HDL (difference)

For the clinic-based population (EAGLE/BioVU), we will consider FHCL cases for patients with ICD9 codes 272.2 (disorders of lipid metabolism).

e. Covariates (Specify)

In addition to genetic factors, the following covariates will be added to the model if available on the majority of participants/patients:
- Age
- Gender
- BMI
- Smoking status (ever/never)
- Treatments (ever/never for cholesterol-lowering therapy)
- Menopause
- Diabetes (y/n)
- Blood Pressure
- Principal component identified ethnicity
- Self-identified ethnicity
- Alcohol consumption

f. Main statistical analysis methods

For all analyses, we expect access to individual-level data. SOL-approved analysts will access individual-level SOL/CALiCo data along with data from MEC, WHI, and EAGLE/BioVU through data use agreements.

- Linear regression will be used for continuous traits: HDL/TC and TC-HDL as predicted variables (analysis performed on all samples)
- Logistic regression for binary outcomes: FCHL (yes/no) as predicted variables.

All models will be stratified by race/ethnicity (self-reported/administrator reported and genetically-determined) will be unadjusted and adjusted for the covariates listed above if available. We anticipate that the largest groups available for analysis will be African Americans, Hispanics/Mexican Americans, and Asians. For these single-marker analyses in Aim 1, the number of statistical tests will be based on the entire Metabochip dataset (~200,000 variants), making the Bonferroni conservative significance threshold (<2.5x10^{-7}) approximate genome-wide significance.

For burden analysis, we will examine both quantitative and binary outcomes using collapsing tests over the Metabochip gene regions. A variety of tools have been developed to facilitate these analyses, such as the sequence kernel association test (SKAT)⁴. We will rely on our PAGE investigators and non-PAGE collaborators with extensive experience and expertise in this area,
particularly those involved in the Exome Sequencing Project. And, in the event that meta-analysis is required, we will rely on our PAGE and non-PAGE collaborators with expertise in this evolving area of research (such as those involved in meta-analysis of rare variants in the CHARGE consortium).

We will use Biofilter software\(^4\) to test for epistasis within biologically relevant pathways represented by the Metabochip. Biofilter will conduct linear and logistic regression modeling as above, for pairs of SNPs. Likelihood ratio tests will be used to examine the statistical significance of including an interaction term between pairs of SNPs. Due to the high number of correlated tests, we will conduct false-discovery rate correction using the method of Storey et al, implemented in the R ‘qvalue’ package. FDR thresholds will be set < 20%.

We will use Structure software, developed by Pritchard and co-workers\(^5\), to perform ancestry mapping analyses.

g. Ancestry information used? No ___ Yes X How is it used in the analyses?
Ancestry information as generated by each study for each sample will be used to confirm ethnicity for stratified analyses. We will also adjust for AIMs-based principal components. Because interaction models may be more susceptible to population stratification, we will adjust for 10-20 PCs for Biofilter analyses. We will work with the PAGE Coordinating Center to estimate PCs and to determine the appropriate number for adjustment by population.

h. Anticipated date of draft manuscript to P&P: __________

i. What manuscript proposals listed on www.pagestudy.org/index.php/manuscripts/ are most related to the work proposed here? Approved PAGE ms. numbers: 031
   – If any: Have the lead authors of these proposals been contacted for comments and/or collaboration? Yes X No ___

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

We request final called genotypes for all Metabochip samples with the needed lipid trait measures.

Aggregate/summary data to be generated by investigators of the study(ies) mentioned:

[ x ] EAGLE; [ x ] CALiCO; [ x ] MEC; [ x ] WHI; [ x ] CC; [ ] Other:________________________
If CALiCo, specify [ x ] ARIC; [ ] CARDIA; [ ] CHS; [ ] SHS-Fam; [ ] SHS-Cohort; [ ] SOL

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

V. REFERENCES

