ARIC Manuscript Proposal # 1857

PC Reviewed: 10/10/11	**Status: A**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

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).

PAGE Ms. Number: _____ **Submission Date :** _____ **[Approval Date:** _____**]**

Title of Proposed Ms.: Phenome-Wide Investigation of Pleiotropy Using Pilot MetaboChip Data

I. INVESTIGATOR INFORMATION:

Name of Lead Author:
Sarah Pendergrass
Email Address: sarah.a.pendergrass@vanderbilt.edu
Telephone Number: 603-369-7709

Junior Investigator? Y

Names, affiliations and email address of PAGE Investigators proposed as co-authors, not in paper order:

N, N	Affiliation in PAGE	Email
Christy L. Avery	CALiCo	christy_avery@unc.edu
Gerardo Heiss	CALiCo	gerardo_heiss@unc.edu
Kristine R. Monroe	CALiCo	kmonroe@usc.edu
North, Kari	CALiCo	kari_north@unc.edu
Buyske, Steve	CC	buyske@stat.rutgers.edu
Tara C. Matise	CC	matise@Biology.Rutgers.Edu
Ewa Deelman	CC	deelman@isi.edu
Petra Buzkova	CHS	buzkova@u.washington.edu
Crawford, Dana	EAGLE	dana.crawford@chgr.mc.vanderbilt.edu
Pendergrass, Sarah	EAGLE	Sarah.a.pendergrass@vanderbilt.edu
Ritchie, Marylyn	EAGLE	ritchie@chgr.mc.vanderbilt.edu
Torstenson, Eric	EAGLE	torstenson@chgr.mc.vanderbilt.edu
Chu-Nan Hsu	ISI	chunnan@isi.edu
Jose-Luis Ambite	ISI	ambite@isi.edu
Chris Haiman	MEC	Christopher.Haiman@med.usc.edu
Loic Le Marchand	MEC	Loic@crch.hawaii.edu
Lynne R. Wilkens	MEC	Lynne@crch.hawaii.edu
Hindorff, Lucia	NHGRI	hindorffl@mail.nih.gov
Anastasia Wise	NHGRI	anastasia.wise@nih.gov
Authors TBD after	WHI	

proposal accepted		

Partner studies in PAGE not collaborating in this ms. proposal: NA

II. SCIENTIFIC RATIONALE (Please be specific and concise)

The field of phenomics has been investigating the network structure that exists among large arrays of phenotypes, and genome-wide association studies (GWAS) have been used to investigate the relationship between genetic variation and single diseases/outcomes. We have been utilizing a novel approach that combines both the exploration of phenotypic structure and genotypic variation, known as a phenome-wide association study (PheWAS) [1], where comprehensive associations are calculated between all available SNPs and phenotypes. In the first PheWAS study of PAGE, associations were calculated between year 1 SNPs and phenotypes from the PAGE consortium (Pendergrass et al. 2011, *manuscript in preparation*).

PAGE is currently pursuing studies with the MetaboChip custom genotyping array. The MetaboChip contains metabolism associated single nucleotide polymorphisms (SNPs) identified in genome-wide association studies up to 2009 (33.2% of the SNPs on the array) as well as fine mapping SNPs for select metabolic GWAS-identified variants drawn from multiple ethnic groups and the 1000 Genomes Project. There are also additional SNPs, described in further detail here: http://www.sph.umich.edu/csg/kang/MetaboChip. Thus far, three groups of the PAGE network, Atherosclerosis Risk in Communities (ARIC), Multiethnic Cohort (MEC), and the Women's Health Initiative (WHI) have genotyped a total of 6,359 African-American individuals with the MetaboChip. The numbers of phenotypes available for the pilot study vary: 98 in ARIC, 43 in MEC, and 121 in WHI. These phenotypes fall within phenotypic domains such as anthropometry, type 2 diabetes, hypertension, and inflammation.

We propose a PheWAS manuscript utilizing the pilot MetaboChip results of PAGE. A PheWAS utilizing MetaboChip data could elucidate a more comprehensive picture of the associations between phenotype and genotype through utilizing the unique combination of SNPs from previously published associations as well as fine mapping SNPs available with MetaboChip. An additional unique feature of these data is the focus on genotypic information only from African Americans. The results of this PheWAS have the potential to show novel relationships between SNPs and phenotypes; identify pleiotropy; provide novel mechanistic insights; and foster hypothesis generation.

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

Using our agnostic PheWAS approach we have comprehensively calculated the association between the variation of 161,098 SNPs of the MetaboChip that have passed QC, and the following numbers of phenotypes:

ARIC: 98
MEC: 43
WHI: 121

The number of phenotypes used for this study is smaller than the number of phenotypes used in other PAGE PheWAS research, however there is significant diversity across the types of phenotypes chosen for the pilot project. These phenotypes fit within the following phenotype classes:

Activity level	White blood cell count	HDL
Albumin	Body Mass Index	LDL
Alcohol consumption	Heart Failure	Apoliprotein
Height	MI	Triglycerides
Hematocrit	Stroke	Total Cholesterol
Hemoglobin	Artery Surgery	Smoking
Platelet count	Creatinine	Type 2 diabetes
Heart Rate	Cystatin	Pregnancy
C-reactive Protein	D-Dimer Test	Hysterectomy
Fibrinogen	Glucose	Menopause
Blood Pressure	Insulin	Menarche
Hormone Use		

The variant selection for MetaboChip was focused on specific traits, and novel associations may exist between genotype and phenotype that have been previously unreported, particularly for African Americans. Our primary focus will be to characterize what pleiotropy may exist within these data. Our definition of pleiotropy will be an association result passing our p-value cutoff for more than one phenotype class for the same SNP. Because not all phenotypes that have been investigated in GWAS are represented in this data, we will also consider a result pleiotropic if we see a novel phenotype class result for a SNP that has shown a significant result in the NHGRI GWAS catalog for a different phenotype. While these results will not prove the impact of genetic variation on the expression of more than one phenotype, our results will provide a view of the "landscape" of potential pleiotropy that may exist across the genomic regions represented on MetaboChip, and could be used in hypothesis generation for future investigations beyond the scope of this manuscript proposal. We are likely to see known associations represented in our results, significant associations for related phenotypes that reflect more the correlation between phenotypes than the impact of genetic variation on multiple phenotypes, as well as novel associations. We will be asking the following questions: *Does pleiotropy exist across the variants covered in the MetaboChip for African-Americans? Are there areas on the MetaboChip showing more pleiotropy than others? Are the phenotype classes of such regions closely related, or do some regions represent a wider diversity of phenotypes?*

To answer these questions, a set of association analyses have been performed on the MetaboChip pilot data and preliminary plots have been generated. Figure 1 shows regions where two or more phenotype classes had significant results for the same SNP. We counted the number of associations for two or more phenotypic classes for SNPs on MetaboChip, for $p < 5E\text{-}4$. The phenotype classes listed above were used, except for HDL, LDL, Triglycerides, Total Cholesterol, and Apoliprotein A& B, which were combined into a single phenotype class for this plot, as the lipid phenotypes are highly correlated. The diversity of the phenotypes represented for SNPs of MetaboChip is more apparent when not having each lipid phenotype class counted separately, which inflate our "counts" at each SNP greatly.

It will be of interest to be able to discern what phenotype classes are represented at each SNP where there is more than on significant phenotype-class association, so we intend to take these results and for the manuscript create a primary manuscript figure more like Figure 2, which is from the NHGRI Catalog of Published Genome-Wide Association Studies[2]. All significant results at a cutoff of 4E-5 will be plotted in this way, showing in a single figure the variety of significant

phenotypic class associations across all chromosomes, and locations of multiple phenotype classes at a single location, with colored circles indicating different phenotype classes.

We have information from the University of Michigan regarding the selection of various regions of MetaboChip (http://www.sph.umich.edu/csg/kang/MetaboChip/) that we will use for interpretation and potentially visualization of association results. For instance, some regions of MetaboChip have previous association with HDL cholesterol levels, and this information can be used along with the results of this study to determine whether or not the associations found here diverge from previous associations for these SNPs.

In order to account for SNPs that have previously been examined by PAGE, we will link the results of the MetaboChip study to those of the PAGE PheWAS in the database. This will allow for the potential filtering out of results that may already exist within PAGE. We have many more SNPs available with the MetaboChip array, and the MetaboChip contains has diversity in the types of SNPs represented, limiting the number of possible overlaps between previously genotyped PAGE SNPs and MetaboChip.

For the table of the manuscript we intend to show the results for a single phenotype, likely the results for the D-dimer phenotype. This phenotype, to our knowledge, is the least likely phenotype to overlap with other PAGE manuscripts under preparation. D-dimer levels are an interesting phenotype, as they are correlated with thrombotic events in older adults. In addition, this phentoype has only been characterized in one GWAS thus far, in European Americans. We are currently determining if the association results for this phenotype will prove of interest for our manuscript.

The possibility of false positives due to multiple hypothesis testing will be an issue with the calculation of this many associations. A p-value cutoff of 4E-5 may be not stringent enough. However, a bonferroni correction is likely too stringent for this kind of exploratory research. Our power here is also limited in many cases, as we are limited in sample size. Thus, we will use permutation of the data to evaluate our false positive rate, as all of the genotypic/phenotypic data results exist in one location, the PAGE CC. Because there is correlation structure between phenotypes for each individual, we will permute all of the phenotypes as a group for each individual. In this way the phenotypic correlation structure will be maintained. We are currently testing to see how long permutation testing will take, which will define how many permutations we will be able to accomplish.

We are considering all SNPs of this study to have the same weight in terms of the calculation of the associations. However, in exploring the results, we will consider some results more important than others. There are fine mapping SNPs on MetaboChip, and LD structure as well that differs in African Americans compared to the European Decent populations, that may impact our results. Unlike in the previous PAGE PheWAS analysis where the SNPs were candidate SNPs distributed over large distances, in the case of this study, plots can be made of all results together, exposing trends across genotyped regions. Also, there is annotation indicating what kind of SNP each variant represented on the array is. Taking into account both of these pieces of information can help inform how novel or interesting any single result of this PheWAS study is.

Because this study is only in African-Americans, we can only characterize what we find for one population. Many of these SNPs were chosen from research focused on European Descent populations. We also have a limited sample size. Even with these limitations, the opportunity here to explore the results of MetaboChip across a wide diversity of phenotypes within a population underrepresented in GWAS, is worthwhile and novel.

It is important to clarify that, even with the overlap between specific phenotypes and SNPs between our PheWAS work and other PAGE MetaboChip manuscript groups, our work has a different focus from the other manuscript groups and should not interfere (in terms of publication of associations or in any other way) with the work of the other manuscript groups. First, many groups

are waiting for the full MetaboChip data, rather than the pilot data, before producing a manuscript. For those groups using the pilot data, the focus of those manuscripts is different from our PheWAS proposal. The other groups are each focused on a single, or very related group of, well harmonized phenotype(s). The other groups will be presenting results for SNPs with p-values. We will be using all of the phenotypes and not harmonizing phenotypes. We will be showing in plots our pleiotropic findings passing a significance cutoff, and other than the D-Dimer results we will not be sharing specific p-values and the actual SNP IDs. As with our previous PheWAS work we will consider any findings of known SNP-Phenotype associations an indication for our group that the high throughput PheWAS approach is functional. In addition, our manuscript(s) from the MetaboPheWAS analysis will be circulated with the rest of the MetaboChip writing groups, so there will be the opportunity for other groups to identify results that might repeat manuscript specific results across writing groups.

Comment [SP1]: A comment from one of the reviewers to note: Let me start by saying that I applaud your efforts to avoid duplicating results from other PAGE manuscripts. This is a very well thought out approach. I would hate for this to end up limiting the manuscript; however, as this is a potentially high impact paper. I would therefore argue that this manuscript should be written as the best paper possible, taking these issues into consideration, rather than the other way around to maximize impact.

a. **Study Questions/Hypotheses**.

This is an exploratory analysis to uncover the spectrum of phenotypes associated with various SNPs on the MetaboChip. We hypothesize that we will replicate some known results, find a series of results where associated phenotypes are not independent of the phenotypes of previously reported associations, as well as identify novel associations. Of great interest will be the identification of potential pleiotropy.

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

The MetaboChip currently is genotypic data only from those of African-American ancestry. Thus far, three groups of the PAGE network, ARIC, MEC, and WHI have genotyped a total of 6,359 African-American individuals with the MetaboChip platform. This will be used for our analysis.

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

A total of 161,098 SNPs on the MetaboChip have passed QC.

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

This is the current number of phenotypes.

ARIC: 98
MEC: 43
WHI: 121

The phenotypes cover several phenotypic domains, such as anthropometry, type 2 diabetes, hypertension, and inflammation and will be reported in this way for the manuscript.

e. **Covariates (Specify)**

We will be adjusting models for sex, and the first two principle components of ancestry.

f. **Main statistical analysis methods**

Linear or logistic regression depending on whether the phenotype is continuous or categorical. For continuous phenotypes, associations will be calculated both with the phenotype variables

untransformed, but also in the form ln(1+y) form, where y is the phenotypic variable. For variables with multiple categories, binning will be used to create new variables of the form "A versus not A".

g. **Ancestry information used? No __ Yes_X__ How is it used in the analyses?**

We will be adjusting for the first two principle components of ancestry

h. **Anticipated date of draft manuscript to P&P: __September 2011_____**

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

Phenotype-Wide Association Study for Exploration of Novel SNP and Phenotype Relationships within PAGE - manuscript 19

Phenotype-Wide Association Study for Exploration of Novel SNP and Phenotype Relationships within EAGLE/NHANES - manuscript 18

- **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): **Check all that apply:**

All associations will be calculated by Steve Buyske of the CC.

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

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

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

1. Pendergrass, S.A., et al., *The use of phenome-wide association studies (PheWAS) for exploration of novel genotype-phenotype relationships and pleiotropy discovery.* Genetic epidemiology, 2011. **35**(5): p. 410-22.
2. Hindorff, L.A., et al., *Potential etiologic and functional implications of genome-wide association loci for human diseases and traits.* Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(23): p. 9362-7.

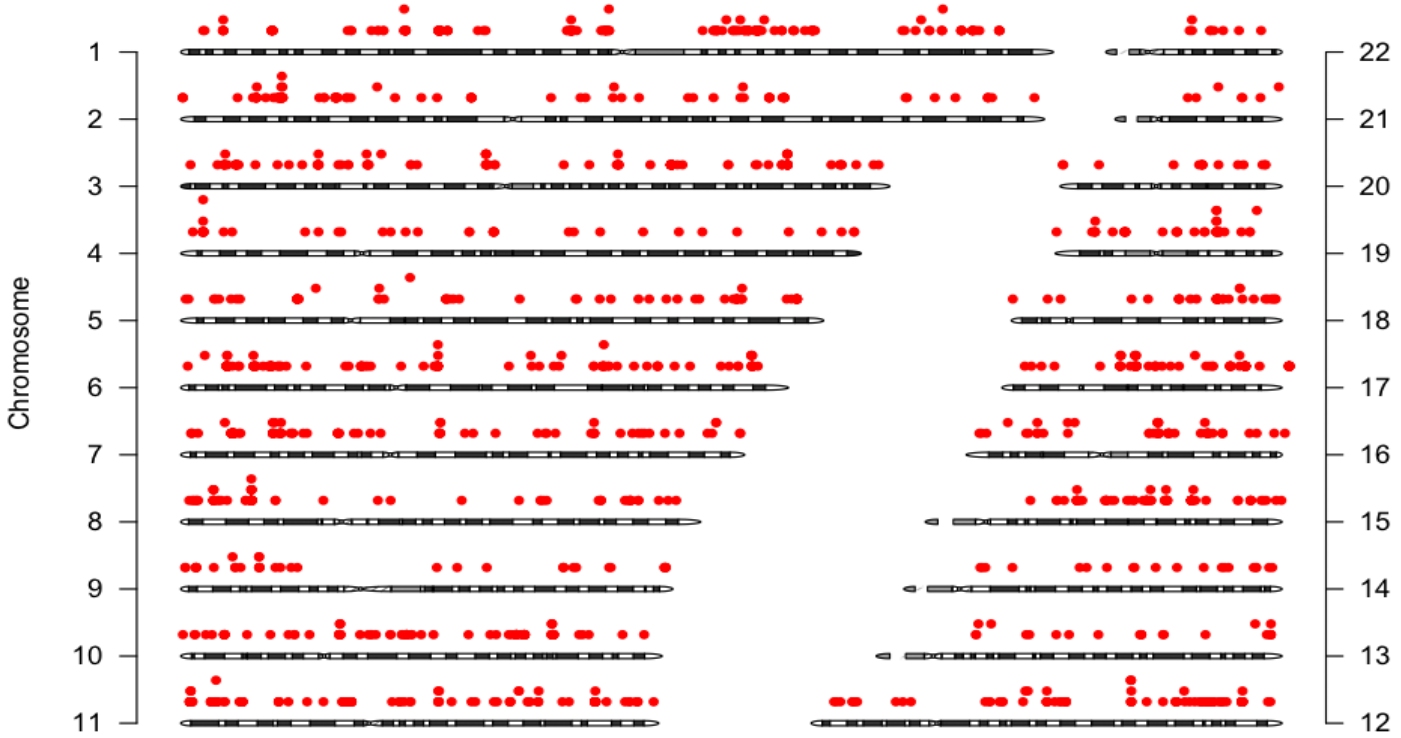

Figure 1 - Plot of pleiotropy counts across chromosomes, where p < 5E-4. This plot shows SNPs of Metabochip with two or more association results for distinct phenotype classes. The number of unique associations is plotted, varying from two to five distinct associations. All of the phenotype classes listed in this manuscript proposal were used, the lipids phenotypes (HDL, LDL, Triglycerides, Total Cholesterol, and Apolipoprotein) were combined together in a single group to reduce the number of counts found among these highly correlated phenotypes.

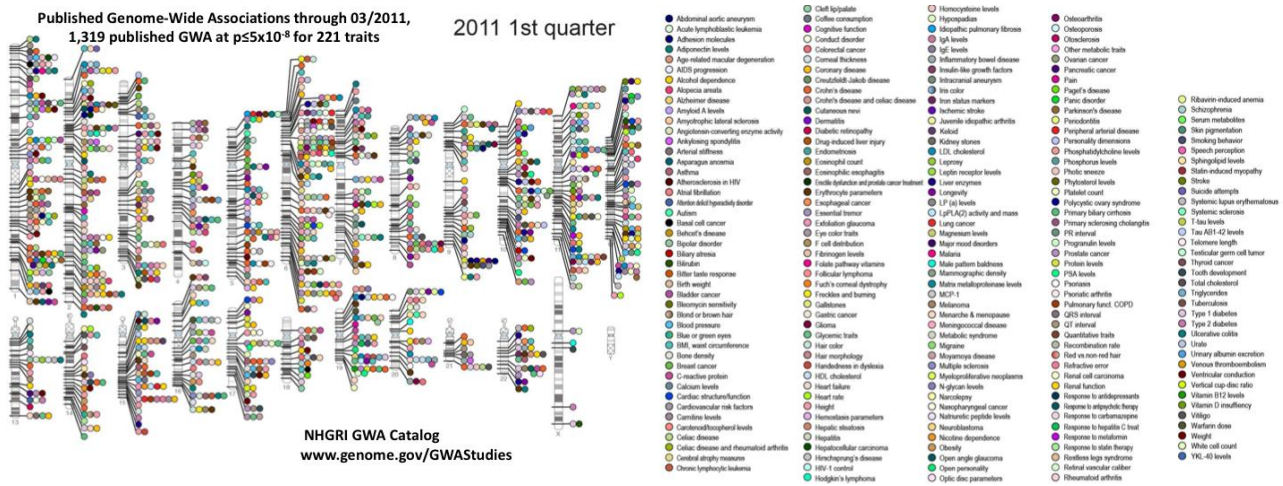

Figure 2 – NHGRI GWA Catalog Figure. We intend to create a figure similar to this for the MetaboPheWAS results. MetaboPheWAS results for each phenotype class with p-values < 5E-4 will be presented, with each phenotype class represented by a different colored circle. In this way significant results for multiple phenotypic classes at a single location can be represented, showing regions of pleiotropy in African-Americans with the variants of MetaboChip.