CONCEPT CLEARANCE FOR RFA

PAGE (Population Architecture for Genomics and Epidemiology) Phase II

National Advisory Council for Human Genome Research

February, 2012

Purpose

The National Human Genome Research Institute (NHGRI) proposes an RFA to assemble, sequence and analyze a population resource of approximately 2,000 well-phenotyped, non-European ancestry (non-EA) individuals to expand understanding of ancestral differences in disease associations observed in PAGE Phase I. This four-year program will extend the initial experience of PAGE I to include whole genome sequencing (WGS) of non-EA participants. The goals of this program are to: 1) using WGS, identify disease-associated genomic regions where between-population differences may be attributable to differences in LD structure; 2) build a population resource of non-EA individuals whose comprehensive genotype, phenotype and association data would serve as a reference population for the scientific community; and 3) explore the associations of sequence variation with a broad range of PAGE phenotypes, including conditions with disproportionate disease burdens in non-EA persons.

Background

In July, 2008, NHGRI funded 5 grants in response to two RFA's titled "Epidemiologic Investigation of Putative Causal Genetic Variants" (RFA-HG-07-014, Study Investigators; RFA-HG-07-015, Coordinating Center). The four Study Investigators (SI) include a consortium of 5 multi-site cardiovascular cohorts, the Women's Health Initiative, the Multiethnic Cohort and one site which includes the National Health and Examination Survey and the BioVU DNA repository. In addition to providing logistical and administrative support, the PAGE Coordinating Center (CC) was originally funded to coordinate cross-study analyses and quality control (QC) of association-level data and has now evolved to QC, genotype calling and ancestry adjustment of individual-level data.

PAGE I analyses leverage the considerable breadth and depth of phenotypes from these cohorts to compare cross-population allele frequencies, relative risks, and distribution of phenotypes associated with particular variants and their environmental modifiers among populations of European and non-EA (African American, Hispanic/Latino, East Asian, Native Hawaiian, or American Indian) descent. PAGE evolved from genotyping approximately 75-200 SNPs per year to Metabochip genotyping of ~200K SNPs, including ancestry-specific fine mapping variants, for a range of metabolic and cardiovascular traits. Results from the Metabochip pilot study (discussed at September 2009 Council and funded by ARRA) in approximately 5,000 PAGE African Americans demonstrated a significant number of GWAS associations that do not generalize to non-EA populations, demonstrating the importance of dissecting GWAS signals in an ethnic-specific way. For the remaining project period, PAGE will analyze Metabochip data on approximately 69,000 participants of African American, Hispanic/Latino, East Asian, and Native Hawaiian descent.

At the beginning of the project period, PAGE organized an approach to analyzing large-scale phenotype data by forming phenotype harmonization working groups and facilitating cross-study analysis, allowing all cohorts to contribute to a wide range of analyses. Development of these plans required negotiating the complex internal structure and dynamics of several parent cohorts. In addition to conducting carefully harmonized and *a priori* analyses of specific traits, PAGE is analyzing and depositing association data across nearly all phenotypes, yielding data suitable for preliminary, highly exploratory phenome-wide association studies (PheWAS). To encourage broad dissemination of these data, the investigators have shortened the publication embargo on the PheWAS data from 12 to 6 months. After initial challenges were addressed, PAGE has been able to ramp up its productivity. When a renewal of the program was first considered last February, 3 PAGE papers had been published and 7 submitted. As of January, 2012, 18 papers have now been published, 11 submitted and 51 in preparation. At its November 2011 meeting, the PAGE External Scientific Panel noted the sharp increase in productivity, the large numbers of well-phenotyped non-EA participants, the breadth of available phenotypes, the availability of dense genotyping data, and a strong overall contribution of PAGE to several fields and across a range of journals in strongly endorsing a renewal of the program.

In planning for a potential renewal of PAGE, NHGRI staff presented to the National Advisory Council for Human Genome Research (NACHGR) in February 2011 a plan to extend PAGE for one year while encouraging greater productivity and assessing the direction a renewal should take, particularly with respect to sequencing. We believe that the time is opportune to pursue a WGS-based renewal in FY2013 for the following reasons. First, the costs of next generation sequencing are decreasing such that a $1000 WGS will likely be within reach in the next 4 years. Second, although sequencing efforts, mostly of whole exomes, are underway, inclusion of non-EA participants is limited. A WGS-based PAGE renewal would facilitate discovery of ancestry group-specific variants across a range of common traits and across the allelic frequency spectrum. These data would be also useful in developing a customized multi-ethnic SNP chip informed by ancestry-specific association data on a wide range of phenotypes for use by the scientific community. Third, ongoing efforts are also piecemeal with regards to disease. Developing a framework for identifying notable genotype-phenotype associations in high dimensionality genomic and phenotype data will be a key product of PAGE II. That the vast majority of GWAS signals are located in intronic and intergenic regions suggests the benefits of sequence-based disease risk characterization will not be fully realized until a comprehensive effort at WGS in population-based cohorts is established.

Research scope and objectives

For PAGE II, two RFAs would fund 4 SI's and 1 CC, respectively, to identify appropriate participants, sequence whole genomes, and analyze a cohort of approximately 2,000 non-EA study participants. Although the exact numbers of study participants to be included depend heavily on sequencing costs, we propose that African descent and Hispanic/Latino cohorts of at least 1000 participants each would be broadly useful as reference populations. Each applicant would assemble a multidisciplinary team of individuals with expertise in genomics, specific diseases, and computational and statistical methods to address the program goals: 1) using WGS, identify disease-associated genomic regions where between-population differences may be attributable to differences in LD structure; 2) build a population resource

of non-EA individuals whose comprehensive genotype, phenotype and association data would serve as a reference population for the scientific community; and 3) explore the associations of sequence variation with the full range of PAGE phenotypes, including diseases with disproportionate disease burdens in non-EA persons.

To accomplish goal 1, awardees would work cooperatively to coordinate selection of study participants and create a cross-study analysis pipeline. Investigators would propose a set of scientific aims to analyze WGS data in relation to one or at most a few conditions, recognizing that cross-study analyses, prioritized by the PAGE II Steering Committee, would be done where possible. Applicants would also propose a technical plan to conduct WGS in the proposed samples. Given rapidly decreasing sequencing costs, the RFA will allow NHGRI the flexibility to consider WGS costs at the time of funding to determine whether WGS would be performed at each site or subcontracted centrally through the CC.

To accomplish goal 2, grantees would agree to contribute individual-level sequence and phenotype data to be made publicly available through dbGaP or a similar resource. In addition to phenotypes, study participants would also be selected on the basis of appropriate consent for data deposition and return of relevant research results. Limited funds would be provided for reconsent of study participants only with strong justification; for example, if the study participants were critical to the success of the program, or were from unique populations. Given the modest scale of this program, applicants will be asked to describe the full set of study participants and their associated phenotypes available for sequencing and analysis, should additional funding or sequence capacity be made available during the project period.

To accomplish goal 3, grantees will propose methodological approaches to leverage the considerable phenotype data available to PAGE II. These efforts may include expanding on the PheWAS design, development of new methods to simultaneously incorporate high dimensionality genotype and phenotype data, and/or dissemination of association data in a user-friendly manner. Applicants will describe a comprehensive set of phenotypes available for analysis and data dissemination.

With the complexity of sequence data come the potential to identify variants of secondary interest and the question of whether research results should be returned to participants. Applicants will be asked to describe their plans for deciding whether and how to return research results, incorporating sequencing and validation in CLIA-certified labs where appropriate. Applicants will also be expected to address ethical, legal and social concerns that arise when consenting individuals for this program and for returning research results, and to include appropriate expertise to address these concerns.

Mechanism of support

As for PAGE I, these RFAs will use the U01 Cooperative Agreement mechanism.

Funds available

NHGRI intends to commit $5M in FY2013, to be split among 4 SI awards and 1 CC award. Funding a total of approximately $20M over 4 years is anticipated.