CONCEPT CLEARANCE

Program Announcement (PA) for an X01-based Solicitation to Provide Access to Large-scale Sequencing Capacity for Association Studies

National Advisory Council for Human Genome Research
September 10, 2007

Background

Whole-genome association studies, in which hundreds of thousands of variants are genotyped across the genome in hundreds to thousands of samples, have been highly successful over the last year in identifying gene regions that are associated with common diseases. After replication of the initial findings, the next step generally is to sequence the implicated regions to determine their complete patterns of variation. The goal can potentially be threefold: to identify all common variants in a region that might be driving the original association; to identify secondary variants or haplotypes in the same region that also affect disease risk; and to search for rare variants of large effect in the same locus. The sequence data produced will include a comprehensive set of the variants in these regions (SNPs; copy-number variants; small and large insertions, deletions, and inversions), as well as their frequencies and LD relationships. These data can then be used by investigators with case and control data to narrow the genotype association peaks and prioritize sets of candidate causal variants for subsequent studies, such as functional analyses, that will be pursued to understand the genetic basis of disease.

Scope

As part of its Medical Sequencing program, the NHGRI plans to provide the capacity for sequencing a number of candidate regions identified in genome-wide association studies. Such studies will have been supported by NIH (by the Genes, Environment and Health Initiative (GEI) or ICs) or by other funding agencies.

Sequencing technology is approaching the point that it is feasible to sequence regions from hundreds to thousands of individuals. The amount of sequencing that will be provided for each study will depend on the number of regions that are strongly associated with disease and the amount of LD in those regions.

The NHGRI-funded sequencing centers will perform the sequencing and provide the data to NCBI. Fragmentary data (such as individual Sanger sequencing reads) and a minimal set of phenotype data will be in a public-access area, and assembled data and the complete set of phenotype data will be in a controlled-access area. The NHGRI and GWAS policies will apply to these data.

Mechanism of support

The X01 is a new mechanism that was developed to provide access to NIH resources, but no funds; it allows the use of the automated NIH application tracking system and of the peer review system to evaluate the scientific merit of the requests. NHGRI proposes to issue a program announcement for X01 requests for access to the sequencing capacity at the NHGRI-funded large-scale sequencing centers.

Review and decision processes

The request for access to NHGRI-funded sequencing capacity will be in the form of an X01 application, which can be relatively short (5-10 pages) and will consist primarily of the scientific justification for the sequence data, the plan for sequencing (samples, regions), and a description of how the data will be used. A peer review committee organized by NHGRI will evaluate the scientific merit of the proposed sequencing and the experimental design for each project.

Review considerations will include:

1. Significance of the proposed study: How significant is the disease and how important would the proposed data be for providing insight into its biology? Would sequencing be expected to yield useful information narrowing association peaks and providing a comprehensive look at variation in the associated regions?
2. Strength of evidence for associations in the proposed regions: Is the evidence for the associations replicated and convincing?
3. The amount of sequencing needed: How large are the genomic regions to be sequenced, how many and what types of samples would be used (e.g., samples that provided strong support for the association peaks, samples with particular haplotypes), what quality of sequence data would be needed, and would special features of particular sequencing platforms be required?
4. Strength of plans for subsequent studies.
5. Capability of the disease community: Is it able to use the data?
6. Informed consent: Were the consent processes used to obtain the samples consistent with NHGRI policies, such as those for data release?

The decision to provide access will be based on:

1. Scientific merit
2. Balance of the NHGRI's portfolio for disease-based sequencing
3. Minority participation as study subjects

NHGRI will also try to expand this program by encouraging other ICs to provide additional funding to the sequencing centers for particular studies, as long as the other NHGRI-required criteria are met (consent, scientific merit).

Number of projects

NHGRI anticipates granting at least 15 access requests a year. This number is expected to increase with the expected increase in NHGRI-supported sequencing capacity that should be afforded by the new sequencing technologies.

Additional funding

NHGRI will not provide any funds for the analysis of the sequence data generated through this program, for re-consent, for genotyping, or for clone distribution. ICs and GEI will be responsible for supporting those activities if they choose to do so.

Timeline

A notice about this PA will be put in the NIH Guide in September, 2007, and the PA will be released in December.