cc: EJ

DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service

National Institutes of Health
National Human Genome
Research Institute
31 Center Drive MSC 2152
Building 31, Room 4B09
Bethesda, MD 20892-2152
(301) 496-0844
FAX (301) 402-0837

CONFIDENTIAL

May 30, 2000

TO: Ruth L. Kirschstein, M.D.
Acting Director, NIH

FROM: Director, NHGRI

SUBJECT: Collaboration with The SNP Consortium

Formed in the spring of 1999, The SNPs Consortium (TSC) is a group of pharmaceutical and electronics companies that are collaborating to produce a set of publicly available SNPs, in the belief that generating such a common set of pre-competitive data will benefit both the public and the private sector. NIH participated in the early discussions about the utility of a large, publicly available set of SNPs. Harold Varmus, Maria Freire and Francis Collins all attended the initial discussion. The final plan agreed to among the TSC members was specifically designed to be synergistic with NIH SNP discovery grants and discussions have continued on a regular basis. The TSC is using DNA samples from the set developed by NIH for this purpose and is depositing data in the NCBI database, dbSNP.

Over the course of 1999, there were a number of conversations between NIH and the SNPs Consortium about potential collaborations. Arthur Holden of TSC met with Harold Varmus and also addressed the IC Directors. In December 1999, a specific project for collaboration was discussed with the IC Directors (see attached background document prepared for the IC discussion). The proposal was to evaluate emerging technology for high throughput scoring of SNPs. It was decided that it would be advantageous for NIH to collaborate with TSC on this project and Harold agreed to contribute $700,000 from the DDF (see attached memo). After this decision was made, however, the TSC Board decided on a more limited evaluation which it would finance on its own. Therefore the $700,000 NIH contribution has not been spent.

Communication between NIH and TSC has continued, and a new opportunity for cooperation has developed over the last few months. In part, this opportunity has arisen because the TSC decided to change its strategy for finding SNPs to a more efficient process that will yield two to three times as many SNPs and takes advantage of the now almost complete draft sequence. The new strategy calls for generation of a set of sequence reads from both ends of the inserts in a whole genome plasmid library. This is

of great interest to the public sequencing effort. In addition to being useful for generating SNPs, the paired end sequences can also be used to improve the working draft version of the human sequence, as the data will allow order and orientation to be assigned to the sequence contigs in the draft sequence.

Specifically, the objective of the proposed collaborative effort will be to generate at least two million pairs of sequence reads, and preferably as many as three million. This is twice as much data of this sort as either the TSC or the public sequencing effort could generate on its own in a reasonable period of time. From such a large data set, an additional 500,000 to 1,000,000 SNPs could be generated (currently, the combined targets for the TSC and NIH efforts are about 500,000 to 600,000 SNPs). Similarly, the average length of sequence contigs in the working draft would be increased from about 10,000 to 15,000 base pairs to about 35,000 to 40,000 base pairs, significantly increasing the chance that investigators will be able to identify complete genes, including controlling elements, in the draft sequence.

Thus, collaboration will address the objectives of both the TSC and NIH. The overall database of SNP discovery will be significantly improved, as will the quality and utility of the draft sequence. Through combined support from TSC and NIH, these important objectives could be achieved very quickly - a sufficient number of end sequence pairs could be generated by the end of September 2000. This would benefit the community immensely by making the draft sequence much more useful for many types of experiments.

I therefore request that you approve redirection of the $700,000 from the DDF, already intended to enhance NIH/TSC collaboration, to this new highly synergistic initiative. NHGRI is planning to contribute about $10 million to this collaborative effort; TSC and The Wellcome Trust will fund the rest. The funds will go to the groups that are already sequencing for TSC, namely the Washington University Genome Sequencing Center, the Whitehead Institute Center for Genomic Research and the Sanger Centre. The National Advisory Council for Human Genome Research approved requests for supplements to the Whitehead and Washington University human sequencing grants for this purpose at its meeting on May 22. However, $10.9M is needed to fully support the Washington University and Whitehead efforts; thus, the DDF funds are critically needed to ensure that we take full advantage of this important opportunity.

Francis S. Collins, M.D., Ph.D.

Attachments