FASEB
Federation of American Societies for Experimental Biology

Representing 125,000 Researchers

301.634.7000
www.faseb.org

9650 Rockville Pike
Bethesda, MD 20814

April 5, 2017

Office of Science Policy
National Institutes of Health
6705 Rockledge Drive, Suite 750
Bethesda, MD 20892

Dear NIH Science Policy Team:

The Federation of American Societies for Experimental Biology (FASEB) appreciates the opportunity to provide comments in response to the National Institutes of Health's (NIH's) "Request for Information on Processes for dbGaP Data Submission, Access, and Management" (NOT-OD-17-044). FASEB is composed of 30 scientific societies representing 125,000 biomedical and biological investigators. We commend NIH for seeking stakeholder input to enhance the Database of Genotypes and Phenotypes (dbGaP) and offer cautious support for expanding access to summary statistics. To balance potential risks and benefits, we recommend use of a registration model with an educational component for accessing most types of summary statistics. However, for studies involving sensitive phenotypes, vulnerable populations, children, or smaller sample sizes, we recommend that summary statistics only be available through a controlled-access system. Below we provide specific details in responses to three issues raised in the RFI.

**Alternate controlled-access models**

FASEB supports continued use of controlled access for whole genome sequences and any other datasets or data products that pose a risk for re-identification of research participants. When selecting a model for controlled-access, we encourage NIH to consider approaches that minimize administrative burden for researchers depositing and accessing data. Similarly, NIH also should ensure that the adopted model is uniformly implemented across all NIH genomic databases.

**Benefits and risks associated with the availability of genomic study summary statistics**

FASEB recognizes that the exchange of research findings increases the efficiency of scientific research and can accelerate discoveries that improve human health. We appreciate NIH's consideration of alternative access models for data products that present a much lower risk for re-identification than individual-level data. Below we describe two types of risk and suggest mitigating them through use of registered- and controlled-access models.

First, accurate interpretation of genomic data and associations are challenging due to their complexity and many potential complicating factors. FASEB recommends a registration model that incorporates a brief educational module for first-time users. The educational component could be tailored to the type of user,

with different versions for clinicians, researchers, and members of the public. For researchers, this training should emphasize rigorous and reproducible research practices for using these data products. The modules should also affirm the ethical and legal responsibilities of users to maintain the privacy of research participants and clearly state that users are prohibited from attempting to re-identify human subjects. Registration and training should be transferable across all NIH genomic databases.

Second, providing greater access to summary statistics increases risks for some research participants. In limited circumstances, summary statistics could be used to determine whether a particular individual participated in a particular study; the nature of the study could, in turn, reveal further health information about that individual. There is also concern that summary statistics from research on sensitive phenotypes and vulnerable populations could be misused to further marginalize these groups. Unless these risks were explicitly covered in the informed consent process, summary statistics should only be available through a controlled-access system for any studies involving sensitive phenotypes, vulnerable populations, children, or smaller sample sizes.

**Clinical Use of Genomic Research Data Maintained in Controlled-Access in dbGaP**

Allele frequencies and other summary statistics can help inform the interpretation of clinical test results. However, FASEB is concerned about the potential harm from misinterpretation and misapplication of such summary information – especially in the context of healthcare decisions. Information that could mislead clinicians and patients may inadvertently reduce patient safety. As described in the prior section, we advocate that NIH implement a short educational module as part of the registration process for accessing dbGaP summary statistics. The module should describe best practices for clinical use and introduce users to concepts such as correlation versus causation, effect size, probability of significance, and the risks of generalizing results to other populations.

FASEB appreciates NIH's engagement of the biomedical community as it explores ways to enhance dbGaP. We are supportive of expanded access to summary statistics provided that (1) there is an educational component on appropriate and rigorous use; and (2) summary statistics from more sensitive studies remain only available through a controlled-access system. Please do not hesitate to contact me if FASEB can provide further assistance.

Sincerely,

Hudson H. Freeze, PhD
FASEB President