Genome-Wide Association Studies (GWAS)

Guidance for Developing Data-sharing Plans for GWAS

Under the NIH GWAS Policy, NIH expects grantees to share data from GWAS with the NIH GWAS data repository, or provide an appropriate explanation for not sharing if meeting the expectations of the policy regarding data sharing is not possible. Plans for sharing data, or an explanation for not sharing, should be included as part of the Research Plan of the application. In the PHS 398 application, place this information in Section K. Resource Sharing; in the SF424 (R&R) it should be addressed in the PDF attachment of the PHS 398 Research Plan Component, Item 17. If the data-sharing plan is accepted by the Institute/Center, it will become a term of the Notice of Award. The ability to share, and the richness of the data for maximizing usefulness of future research, may be considered by the IC as part of award decisions.

In developing a data-sharing plan, the applicant should consider the following:

- Will the data be shared in accordance with the NIH GWAS Policy?
  - State whether your IRB or Privacy Board has approved the data-sharing aspects of your project, or the timeline for such approval.
  - If your application or proposal is being considered for funding and involves human subjects research, IRB approval typically is required prior to grant award (just-in-time) or finalization of a contract.

- When data are shared through the NIH GWAS data repository:
  - Which data will be shared?
    - Describe the data elements, study populations, and study documents that will be included in the data-sharing. The minimum expectation is that data generated and used for funded analyses and documentation sufficient for interpretation of data (e.g., study protocol and manuals, data collection instruments) will be shared.
  - Whose data will be shared?
    - Will data from all study participants that are included in the analyses be shared?
    - If only a subset will be shared, provide the rationale (e.g., tiered consent that addresses/addressed sharing).
  - When will the data be shared?
    - Provide a timeline for data-sharing. The expectation is that data will be shared once the data have been cleaned.
  - Will there be restrictions on use limitations for the shared data?
    - This should be determined in consultation with your institution. Your institution will document appropriate uses of the data and data use limitations as part of the institutional certification. Verification of the appropriate uses of the data and data use limitations stated in your data-sharing plan may be requested from your institution prior to award.
    - E.g., Use may be limited to studies of specific conditions or traits, or certain types of uses or users (e.g., non-commercial use only).

- Is additional consent needed for sharing data as described in the NIH GWAS Policy? If a new consent process is necessary, the application may include plans and budget for a new consent process.

- If the data will not be shared through the NIH GWAS data repository:
  - What is the justification for not sharing?
    - E.g., For use of existing data/specimens, your institution may determine that consent is not adequate for data-sharing as described in the NIH GWAS Policy, and it may not be possible to obtain additional consent (e.g., due to the age of the data/specimens).
    - Other issues may preclude sharing such as local laws and limitations, concerns about harms to individuals or groups, or other cases where the expectations of the data submission cannot be met.
  - Specify an alternative data-sharing plan, if another data-sharing mechanism is acceptable.

<u>Sample data-sharing plans</u>

1. Data will be shared through the NIH GWAS repository – existing specimens and data

The University of Blue Skies IRB has advised us that the genome-wide association data produced through this award may be shared through the NIH GWAS data repository, consistent with data-sharing under the NIH data-sharing policy for genome-wide association studies. Institutional certification is underway and will be provided prior to data submission. The IRB has reviewed and verified [see Appendix 3] that:

- The submission of data to the NIH GWAS data repository and subsequent sharing for research purposes are consistent with the informed consent of study participants from whom the data were obtained;
- The investigator's plan for de-identifying datasets is consistent with the standards outlined in the NIH GWAS Policy;
- It has considered the risks to individuals, their families, and groups or populations associated with data submitted to the NIH GWAS data repository; and
- The genotype and phenotype data to be submitted were collected in a manner consistent with 45 C.F.R. Part 46.

We will share study documents, individual-level genotype and phenotype data, and other study information described in the Policy, for the SUNNY study population. We will share the individual-level genome-wide genotyping data produced as part of Specific Aim 1, and the individual-level phenotypic data included in the analyses under Specific Aim 2. Genotyping data will include the genotypes as well as the intensity files used to call the genotypes. Phenotype data will include the outcomes and traits described in Table 2, and the additional analysis variables described in Table 3. A number of these are derived variables, and we will provide the underlying variables as well. The SUNNY study also has collected a number of additional variables which will be shared (Appendix 2). Data will be shared for all of the study participants included in these analyses; participants who did not give consent for sharing data will be

excluded from the study ("Study population" section). For study documentation, the SUNNY study protocol and manual of operations, as well as questionnaire and data abstraction forms, will be provided. We will share the genotype and phenotype data once the genotyping data have been cleaned; we expect the cleaning to be complete no more than two months after genotyping is finished. This is month 12 in the proposed study timeline. The consent form states that the data and specimens may be used for any genetic studies of mood and factors that affect mood. Data use limitations have been verified, and appropriate uses provided in more detail, by the University of Blue Skies IRB (Appendix 3).

We also plan to release the data to qualified researchers who wish to collaborate with the SUNNY study investigators. The availability of data for collaborators will be advertised on the SUNNY study web site. The data will be available through a secure FTP site maintained by the University's IT department, or sent on encrypted physical media via trackable mail.

We acknowledge the intellectual property and scientific publication elements of the NIH GWAS Policy.

2. *Data will be shared through the NIH GWAS data repository – specimens/data not yet collected*

The University of Blue Skies IRB has advised us that the genome-wide association data produced through this award may be shared through the NIH GWAS data repository, consistent with data-sharing under the NIH data-sharing policy for genome-wide association studies. The IRB will review the data-sharing aspect of this project at their March 2008 meeting. We have asked the IRB to verify that:

- The submission of data to the NIH GWAS data repository and subsequent sharing for research purposes will be consistent with the informed consent requested of study participants from whom the data and specimens will be obtained;
- The investigator's plan for de-identifying datasets is consistent with the standards outlined in the NIH GWAS Policy;
- It has considered the risks to individuals, their families, and groups or populations associated with data to be submitted to the NIH GWAS data repository; and
- The genotype and phenotype data to be submitted will be collected in a manner consistent with 45 C.F.R. Part 46.

We will share study documents, individual-level genotype and phenotype data, and other study information described in the Policy, for the SUNNY study population. We will share the individual-level genome-wide genotyping data produced as part of Specific Aim 1, and the individual-level phenotypic data included in the analyses under Specific Aim 2. Genotyping data will include the genotypes as well as the intensity files used to call the genotypes. Phenotype data will include the outcomes and traits described in Table 2, and the additional analysis variables described in Table 3. A number of these are derived variables, and we will provide the underlying variables as well. The SUNNY study will also collected a number of additional variables which will be shared (Appendix 2). Data will be shared for all of the study participants included in these analyses; participants who do not give consent for sharing data will be

excluded from the study ("Study population" section). For study documentation, the SUNNY study protocol and manual of operations, as well as questionnaire and data abstraction forms, will be provided. We will share the genotype and phenotype data once the genotyping data have been cleaned; we expect the cleaning to be complete no more than two months after genotyping is finished. This is month 24 in the proposed study timeline. The draft consent form provides for tiered consent, including data being used for any studies of genetics and health, or data being used for any genetic studies of mood and factors that affect mood. Data use limitations will be verified, and appropriate uses provided in more detail, by the University of Blue Skies IRB by April 2008.

We also plan to release the data to qualified researchers who wish to collaborate with the SUNNY study investigators. The availability of data for collaborators will be advertised on the SUNNY study web site. The data will be available through a secure FTP site maintained by the University's IT department, or sent on encrypted physical media via trackable mail.

We acknowledge the intellectual property and scientific publication elements of the NIH GWAS Policy.

3. *Data will be shared through the NIH GWAS data repository, after an additional consent process*

The University of Blue Skies IRB has advised us that the genome-wide association data produced through this award may be shared through the NIH GWAS data repository, only after participants have given additional consent for data-sharing. Thus, we have included an additional consent process as part of our study design. Institutional certification will be provided prior to data submission. The IRB will review the additional consent process, materials and data-sharing plan at their March 2008 meeting. We have asked the IRB to verify that:

- The submission of data to the NIH GWAS data repository and subsequent sharing for research purposes will be consistent with the additional informed consent requested of study participants from whom the data and specimens were obtained;
- The investigator's plan for de-identifying datasets is consistent with the standards outlined in the NIH GWAS Policy;
- It has considered the risks to individuals, their families, and groups or populations associated with data to be submitted to the NIH GWAS data repository; and
- The genotype and phenotype data to be submitted were collected in a manner consistent with 45 C.F.R. Part 46.

We will share study documents, individual-level genotype and phenotype data, and other study information described in the Policy, for the SUNNY study population. We will share the individual-level genome-wide genotyping data produced as part of Specific Aim 1, and the individual-level phenotypic data included in the analyses under Specific Aim 2. Genotyping data will include the genotypes as well as the intensity files used to call the genotypes. Phenotype data will include the outcomes and traits described in Table 2, and the additional analysis variables described in Table 3. A number of these are derived variables, and we will provide the underlying variables as well. The SUNNY study also has collected a number of additional

variables which will be shared (Appendix 2). Data will be shared for all of the study participants included in these analyses; participants who did not give consent for sharing data will be excluded from the study ("Study population" section). For study documentation, the SUNNY study protocol and manual of operations, as well as questionnaire and data abstraction forms, will be provided. We will share the genotype and phenotype data once the genotyping data have been cleaned; we expect the cleaning to be complete no more than two months after genotyping is finished. This is month 24 in the proposed study timeline. The draft consent form provides for tiered consent, including data being used for any studies of genetics and health, or data being used for any genetic studies of mood and factors that affect mood. Data use limitations will be verified, and appropriate uses provided in more detail, by the University of Blue Skies by April 2008.

We also plan to release the data to qualified researchers who wish to collaborate with the SUNNY study investigators. The availability of data for collaborators will be advertised on the SUNNY study web site. The data will be available through a secure FTP site maintained by the University's IT department, or sent on encrypted physical media via trackable mail.

We acknowledge the intellectual property and scientific publication elements of the NIH data-sharing policy for genome-wide association studies.