

University of Kansas Medical Center

Human Research Protection Program

POLICY on Genome Wide Association Studies (GWAS) at KUMC

The NIH GWAS policy is a data sharing policy implemented in 2008 for NIH-funded Genome-Wide Association studies, or GWAS. The NIH policy facilitates the sharing of large datasets containing coded, de-identified genotypic and phenotypic data obtained from NIH supported or conducted research. The NIH policy applies to data obtained prospectively as well as retrospectively from existing specimens. A key element of the NIH policy is the expectation that data from NIH-supported GWAS will be deposited into the NIH GWAS data repository, currently designated as the database of Genotypes and Phenotypes (dbGaP), at the National Center for Biotechnology Information (NCBI), within the National Library of Medicine (NLM).

The NIH Database of Genotypes and Phenotypes (dbGAP) has two components:

1. An open access portion that will be freely available to the public and will include:
 - a. The protocol;
 - b. Questionnaires;
 - c. Variables measured; and,
 - d. Other supporting documentation
2. A controlled access portion that will only be available to researchers who have been approved by an NIH Data Access Committee (DAC). This portion will include coded phenotype, exposure, genotype and pedigree data, and summary statistics.

The data submission portion of the KUMC policy applies to investigators who receive any NIH funding for genome-wide analysis of specimens. The GWAS policy has no threshold; therefore, any NIH funding for GWAS triggers requirements under the policy. The KUMC-generated data submitted for inclusion in the NIH GWAS data repository will be coded and de-identified by the submitting investigator from KUMC, but the investigator may retain the key to the code that would link to specific individuals. KUMC investigators may request submission of data and / or access to the NIH GWAS data repository through the IRB submission process.

In summary, a project is covered by the KUMC - GWAS policy, if:

1. The PI will obtain NIH funds to conduct GWAS studies through a new application or continuing funding application submitted after 1/25/2008; or
2. The PI has agreed to voluntarily submit genotype or phenotype data to the NIH GWAS registry; or
3. The PI is applying to NIH to request access to the controlled portion of the dbGaP.

Key Terms

Genome Wide Association Study is defined in the NIH policy as “any study of genetic variation across the entire human genome that is designed to identify genetic associations with observable traits or the presence of a disease or condition.”

Coded means (1) identifying information (such as name or social security number) that would enable the investigator to readily ascertain the identity of the individual to whom the private information or specimens pertain has been replaced with a number, letter, symbol, or combination thereof (i.e., the code); and (2) a key to decipher the code exists, enabling linkage of the identifying information to the private information or specimens. http://grants.nih.gov/grants/policy/hs/coded_synopsis.htm

De-identified means that the identities of data subjects cannot be readily ascertained or otherwise associated with the data by the repository staff or secondary data users (45 CFR 46.102(f)), the 18 identifiers enumerated at section 164.514(b)(2) of the HIPAA Privacy Rule are removed and the submitting institution has no actual knowledge that the remaining information could be used alone or in combination with other information to identify the subject of the data.

Certificates of Confidentiality are issued by the National Institutes of Health (NIH) and other HHS agencies to protect identifiable research information from forced or compelled disclosure. They allow the investigator and others who have access to research records to refuse to disclose identifying information on research participants in civil, criminal, administrative, legislative, or other proceedings, whether federal, state, or local. Certificates of Confidentiality may be granted for studies collecting information that, if disclosed, could have adverse consequences for subjects, such as damage to their financial standing, employability, insurability, or reputation. By protecting researchers and institutions from being compelled to disclose information that would identify research subjects, Certificates of Confidentiality help to minimize risks to subjects by adding an additional level of protection for maintaining confidentiality of private information.

Certificates of Confidentiality protect subjects from compelled disclosure of identifying information but do not prevent the voluntary disclosure of identifying characteristics of research subjects. Researchers, therefore, are not prevented from voluntarily disclosing certain information about research subjects, such as evidence of child abuse or a subject's threatened violence to self or others.

However, if a researcher intends to make such voluntary disclosures, the consent form should clearly indicate this. Furthermore, Certificates of Confidentiality do not prevent other types of intentional or unintentional breaches of confidentiality. As a result, investigators and IRBs must use other appropriate mechanisms and procedures to protect the confidentiality of identifiable private information to be obtained in the proposed research. <http://www.hhs.gov/ohrp/policy/certconf.html>

PROCEDURES

I) Initial proposals for submitting data to the GWAS database of Genotypes and Phenotypes (dbGaP)

- a) **PI Actions:** The PI must complete and submit the following documents for HSC -IRB review and certification:
 - i) An application for a new study in eIRB which indicates a request for GWAS submission by uploading a Project Description in the eIRB Supporting Documents section;
 - ii) A Data Sharing Plan / Form uploaded into the eIRB application in the Supporting Documents section which details the following:
 - (1) A description of all data fields (genotype and phenotype) being submitted to the GWAS repository;
 - (2) A description of the method(s) to be used to code data for transmission to the NIH; investigators should outline plans to de-identify the data according to the following criteria:
 - (a) A random, unique code should be assigned to the data;
 - (b) The identifiers outlined in the HIPAA Privacy Rule should be removed prior to the submission; and
 - (c) The identities of subjects should not be readily ascertained or otherwise associated with the data by the repository staff or secondary data users.
 - (3) A description of how the code key(s) will be maintained by the PI;
 - (4) A written confirmation by the PI that the code keys will never be shared with the NIH;
 - (5) A description of any limitations on use of data or samples (e.g., subjects signed consent forms stipulating use of their data or specimens only in particular fields of research);
 - (6) A statement as to whether whole genome or whole exome sequencing will be performed on any samples being submitted; and,
 - (7) The name of the applicable Project Officer at NIH.
 - iii) After IRB approval, the PI verifies that the Institutional Certification letter is obtained from the Office of the Vice Chancellor for Administration and included in the application to NIH.
- b) **Additional Informed Consent Requirements for GWAS studies**
 - i) *Purpose of the research project* should include succinct explanation of the proposed study, who is sponsoring the research, the nature of the genetic research or analysis and the plans to submit genetic data to a national repository
 - ii) *Description of the research procedures* should include topics such as:
 - (1) The process for the collection of samples (blood or other tissue) and health information;
 - (2) A description of the types of demographic, genetic and clinical information that will be submitted to the repository;
 - (3) How samples and health information will be coded and stored;
 - (4) Whether there will be access by local investigators to the research participant's medical records and, if so, the process for accessing them (e.g., one-time vs. ongoing collection of information from medical records);
 - (5) The duration of storage;
 - (6) Whether and how samples and health information will be shared with qualified external investigators for appropriate research use both during the study period and after the study ends;
 - (7) A general description of the types of external researchers who will have access to samples and data (e.g., academic, industry, government); and
 - (8) Whether and how future contact (i.e. re-contact) is planned.

- iii) *Financial compensation, costs and commercialization* must be clearly communicated to potential research participants, including:
 - (1) Whether there will be any financial compensation for taking part in the study;
 - (2) If the research could lead to the development of new medical breakthroughs, there is no plan for research participants to receive any part of the profits generated from these breakthroughs / products; and
 - (3) A statement explaining who owns the samples.
- iv) *Potential risks of participating in GWAS studies* need to be clearly articulated. Possible risks vary depending upon the study protocol, but as with most genomics research, the potential risks are centered on psychological and social risks for the research participant and, possibly, their family. Additional risks specific to GWAS may include:
 - (1) Risks related to broad sharing of phenotype and genomic data (*e.g.* genotype, DNA sequence, expression profiles, etc.);
 - (2) Risks of the data sharing model for the study (*e.g.* the possibility that the coded ¹ data may be released to members of the public, insurers, employers, and law enforcement agencies);
 - (3) Risks of receiving information that is unwanted by the participant;
 - (4) Risks of computer security breaches or other unanticipated distributions arising from maintaining data in an electronic format;
 - (5) Risks to relatives or identifiable populations or groups;
 - (6) The uncertainty of findings related to genetic risk for a given disease or trait; and
 - (7) Privacy risks, both those known and those unforeseen at this time.
- v) *The potential scientific benefits from GWAS research*: Potential benefits to the research participant and to others should be described in the consent form. It is important to include potential benefits for society, but investigators should distinguish between potential benefits to the individual research participant versus society.
- vi) *Risks to confidentiality* section of the consent document should describe the level of confidentiality of the research data and the measures planned to maintain confidentiality.
 - (1) Participants should know whether their samples will be anonymous/non-identifiable (*i.e.* personal identifiers will not be kept with their sample and the sample will not have a code number that can be used to identify the participant).
 - (2) Samples are coded and considered de-identified (*i.e.* any identifying information such as name or SS# will be replaced with a code and only a few authorized people will have access to this code to link samples and data back to personal identifiers).
 - (3) The use of Certificates of Confidentiality should be addressed here if applicable.
- vii) *Returning results to the research participants*: Subjects must be informed about whether or not they will receive results of testing on their genetic samples. The decision on whether to return research results to participants (either individual research results or general research findings through newsletters, study Web site, etc) should be made by the study investigator in consultation with the IRB. If return of results is planned, the format and process should be clearly communicated in the consent document.
- viii) *The mechanism for requesting withdrawal of information from the repository and limits to the ability to withdraw*:
 - (1) Participants have the right to withdraw from the study at any time and the implications and consequences of withdrawal should be discussed in this section of the form and as part of the overall consent process.

- (2) For certain genomic studies, complete withdrawal and information may not be possible once information has been posted for broad data sharing. In such circumstances, a full explanation of the inability to withdraw all information should be provided.
- (3) For studies where individual-level genomic, demographic, and health data will be deposited in a public or controlled-access data repository for broad sharing with the research community, the consent form should align with the data repository policy on whether or not information may be withdrawn.
- c) IRB Review and Determination For any submission of phenotype data and genotype data to an NIH GWAS repository, the IRB must determine the following during a review conducted at a convened meeting:
 - i) All the information to be submitted meets the HIPAA standard for de-identification;
 - (1) the identities of data subjects cannot be readily ascertained or otherwise associated with the data by the repository staff or secondary data users (45 CFR § 46.102(f)); and
 - (2) the identifiers enumerated at section 164.514(b) (2) of the HIPAA Privacy Rule are removed;
 - ii) The proposed data sharing plan is appropriate and meets federal requirements;
 - (1) To reduce the risk of re-identification, the IRB may limit the types or fields of phenotype data that the PI may submit to the GWAS repository.
 - iii) IRB should consider the extent to which the genotype and other phenotype information associated with the participants could be used to identify an individual or his or her family members by matching the genotype/phenotype datasets to other sources of information;
 - iv) GWAS studies of these data do not pose any risks (e.g., stigmatization) to particular populations or groups;
 - v) Specimens and data were or will be collected in compliance with 45 CFR 46 (federal requirements governing human subjects research)and state law;
 - vi) The proposed consent form meets the criteria above;
 - (1) If the consent forms impose use restrictions that are inconsistent with broad data sharing, the IRB may impose use restrictions upon the data submitted to the NIH, (e.g., a restriction to specify that the data may be disclosed by NIH only for the study of a particular disease or only for non-commercial research).
 - vii) A Certificate of Confidentiality is or is not required for a particular study.
- d) Certificates of Confidentiality- Investigators should determine whether a Certificate of Confidentiality has been obtained for their research or, if one has not been obtained, to recommend to the IRB if it would be appropriate to do so. Ultimately, the IRB will determine if a Certificate of Confidentiality is required for a particular study.

II) Previously approved research that did not anticipate GWAS data submission

- a) PI Actions: The PI must complete and submit the following documents for IRB review and certification:
 - i) An amendment / modification to the protocol of an existing active study in eIRB, indicating a request for GWAS submission in the revised Project description to be uploaded in the Supporting Documents section;
 - ii) A Data Sharing Plan / Form uploaded into the eIRB application in the Supporting Documents section which details the following:
 - (1) A description of all data fields (genotype and phenotype) being submitted to the GWAS repository;

- (2) A description of the method(s) to be used to code data for transmission to the NIH; (investigators should outline plans to de-identify the data according to the following criteria:
 - (a) A random, unique code should be assigned to the data;
 - (b) The identifiers outlined in the HIPAA Privacy Rule should be removed prior to the submission; and
 - (c) The identities of subjects should not be readily ascertained or otherwise associated with the data by the repository staff or secondary data users;
 - (3) A description of how the code key(s) will be maintained by the PI;
 - (4) A written confirmation by the PI that the code keys will never be shared with the NIH;
 - (5) A description of any limitations on use of data or samples (e.g., subjects signed consent forms stipulating use of their data or specimens only in particular fields of research);
 - (6) A statement as to whether whole genome or whole exome sequencing will be performed on any samples being submitted; and,
 - (7) The name of the applicable Project Officer at NIH.
- iii) If the submission to the IRB is an amendment / modification to an existing study and will involve consent forms previously approved and signed by KUMC subjects without GWAS-specific language, the PI must provide the IRB with a copy of each version of the consent form. The form(s) must be uploaded in the eIRB application. To facilitate review and approval, the PI may wish to provide an opinion to the IRB concerning whether the language in any previous version of the consent form(s) would prohibit a submission of subject data to NIH.
 - iv) After IRB approval, the PI verifies that the Institutional Certification letter is obtained from the Office of the Vice Chancellor for Administration and included in the application to NIH.
- b) IRB Review and Determination IRB reviews the data sharing plan and the consent form previously signed by the subjects and determines whether or not to approve the GWAS submission. The determination is based on the following factors:
- i) All the information to be submitted meets the HIPAA standard for de-identification;
 - (1) the identities of data subjects cannot be readily ascertained or otherwise associated with the data by the repository staff or secondary data users (45 CFR § 46.102(f)); and
 - (2) the identifiers enumerated at section 164.514(b) (2) of the HIPAA Privacy Rule are removed;
 - ii) The proposed data sharing plan is appropriate and meets federal requirements;
 - (1) To reduce the risk of re-identification, the IRB may limit the types or fields of phenotype data that the PI may submit to the GWAS repository.
 - iii) IRB should consider the extent to which the genotype and other phenotype information associated with the participants could be used to identify an individual or his or her family members by matching the genotype/phenotype datasets to other sources of information;
 - iv) GWAS studies of these data do not pose any risks (e.g., stigmatization) to particular populations or groups;
 - v) Specimens and data were collected in compliance with 45 CFR 46 (federal requirements governing human subjects research) and state law;
 - vi) The previous consent form adequately covers the proposal to send data to NIH (i.e., possibility of sharing the data outside KUMC, use for future research);
 - (1) The IRB may determine that the original consent forms are not consistent with the requirements for sharing, that no research consent forms were signed, or the risk of re-identification seems too high;

- (2) If the consent forms impose use restrictions that are inconsistent with broad data sharing, the IRB may impose use restrictions upon the data submitted to the NIH, (e.g., a restriction to specify that the data may be disclosed by NIH only for the study of a particular disease or only for non-commercial research);
- vii) Whether past or current subjects should be re-consented:
 - (1) If the research consent was not consistent with GWAS sharing requirements or there are no research consent forms (e.g., samples were obtained under a surgical consent or waiver of consent) subjects must be contacted and consented before the PI may submit their data to the NIH GWAS repository. If appropriate, the IRB may approve an amendment to allow the PI to re-contact subjects to seek their re-consent, or, in limited circumstances and when consistent with regulations, the IRB may waive consent to allow submission of data;
 - (2) For the submission of data or specimens from studies previously approved by the HSC – IRB (in addition to the items in number 1, above), the IRB must determine that existing or past versions of the consent forms are not inconsistent with sharing genotype or phenotype information with the NIH GWAS repository. In reviewing this requirement, the IRB should consider the following questions:
 - (a) Do any of the consent forms contain statements such as “your data/specimens will not be shared” or “will only be seen by the research team?”
 - (b) Were any of the specimen donors minors?
 - (c) Did the original consent forms limit future use to specific projects or disease states, or to non-commercial research?
 - (d) Were any of the data (e.g., medical record information) collected under a waiver of consent?
 - viii) If the consent form should be revised for future enrollees.
 - ix) A Certificate of Confidentiality is or is not required for a particular study.
- c) Certificates of Confidentiality- Investigators should determine whether a Certificate of Confidentiality has been obtained for their research or, if one has not been obtained, to recommend to the IRB if it would be appropriate to do so. Ultimately, the IRB will determine if a Certificate of Confidentiality is required for a particular study.

III) Institutional Certification Letter

- a) The IRB is responsible for reviewing and verifying for the Institutional Official, that:
 - i) 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;
 - ii) The investigator’s plan for de-identifying datasets is consistent with the standards outlined in the policy;
 - iii) It has considered the risks to individuals, their families, and groups or populations associated with data submitted to the NIH GWAS data repository; and
 - iv) The genotype and phenotype data to be submitted were collected in a manner consistent with 45 C.F.R. Part 46.
- b) When the IRB has completed its review and made all necessary determinations to allow for sharing through submission or access to the data repository, KUMC will issue a certification letter signed by the Vice Chancellor of Administration as requested.
- c) The KUMC Institutional Official, Vice Chancellor for Administration, is responsible for certifying that data submission plans meet the following expectations defined in the GWAS policy:

- i) The data submission is consistent with all applicable laws and regulations as well as KUMC policies;
- ii) The appropriate research uses of the data and the uses that are specifically excluded by the informed consent documents are delineated; and
- iii) The identities of research participants will not be disclosed to the NIH GWAS data repository.
- d) The Institutional Certification letter will be sent to the PI, who must then submit it to NIH;
- e) For multi-site studies in which KUMC serves as the lead or coordinating site in a multi-institutional effort, the KUMC IRBs and the Institution will not make certifications for data or specimens to be submitted to an NIH GWAS repository from cooperating sites. It is incumbent upon KUMC investigators who serve as lead or coordinating PIs to maintain documentation from each individual site's respective IRB that a certification has been made. If KUMC IRB serves as the IRB of record for multi-institutional efforts, then KUMC certification applies to data and specimens submitted for the cooperating institutions by KUMC. The investigator should verify with the General Counsel's Office that the appropriate indemnity language is contained in the sub-site agreements.

IV) Investigators requesting access to the GWAS database

- a) PI Actions (from Points to Consider, NIH) Investigators from KUMC seeking data from the NIH GWAS data repository will submit to the NIH a Data Access Request along with a Data Use Certification that will stipulate a number of protections for research participants. Both the Data Access Request and the Data Use Certification must be co-signed by the investigator and by the appropriate designated Institutional Official, the Vice Chancellor for Administration, to document their joint agreement to follow NIH policy for the use of GWAS data obtained from the NIH GWAS data repository. The Data Use Certification will stipulate that, subject to applicable law, the investigator and KUMC will:
 - i) Use the data only for the approved research;
 - ii) Protect data confidentiality;
 - iii) Follow appropriate data security protections;
 - iv) Follow all applicable laws, regulations and local institutional policies and procedures for handling GWAS data;
 - v) Not attempt to identify individual participants from whom data within a dataset were obtained;
 - vi) Not sell any of the data elements from datasets obtained from the NIH GWAS data repository;
 - vii) Not share with individuals other than those listed in the request any of the data elements from datasets obtained from the NIH GWAS data repository;
 - viii) Agree to the listing of a summary of approved research uses within the NIH GWAS data repository along with his or her name and organizational affiliation;
 - ix) Agree to report violations of the GWAS policy to the appropriate Data Access Committee;
 - x) Acknowledge the GWAS policy with regard to publication and intellectual property; and
 - xi) Provide annual progress reports on research using the GWAS dataset.
- b) The recipient investigator will be expected to protect the data by following best practices for data security posted on the NIH GWAS data repository website at http://www.ncbi.nlm.nih.gov/projects/gap/pdf/dbgap_2b_security_procedures.pdf, or other dataset-specific recommendations as detailed for a given GWAS within the repository. Investigator must store data only on servers that are managed by KUMC Information Resources.

In addition, progress reports will be reviewed by the relevant DAC to verify continued appropriate use of the data.

- c) IRB Responsibilities Generally, access to GWAS data does not require IRB review and approval. However, there are some data sets that do require review and approval by the IRB. If IRB approval is required to access data the Investigator submits the appropriate IRB application for review and approval by the IRB. The NIH Data Access Committees (DACs), that must approve access to data in the Repository, may in some uncommon situations require local IRB review. If IRB review and approval is required to access data the investigator should contact the HSC-IRB for further instructions
- d) Institutional Certification: Once the investigator prepares the data access request and uploads the necessary documents to the NIH GWAS web site, IRB will be notified. The Vice Chancellor for Administration, who is the institutional official for human research, reviews the documentation and completes the submission by certifying the request at the GWAS web site.

APPLICABLE REGULATIONS AND GUIDANCE

1. 45 CFR 46.111(a)(7)
2. 21 CFR 56.111(a)(7)
3. HIPAA Privacy Rule
4. NIH Points to Consider for IRBs and Institutions in their Review of Data Submission Plans for Institutional Certifications - Under NIH's Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS) Version 11/12/07