DATA USE CERTIFICATION AGREEMENT

The Cancer Genome Atlas (TCGA)
(August 20, 2014 version)

INTRODUCTION AND STATEMENT OF POLICY

The National Institutes of Health (NIH) has established two central data repositories for The Cancer Genome Atlas (TCGA): the Cancer Genomics Hub (CGHub) stores lower level sequence data and the TCGA Data Portal stores all other data types and higher-level analyses of the sequence data. Systems implemented for the database of Genotypes and Phenotypes (dbGaP) also manage the application, approval and authentication processes for investigators wishing to access Closed Access TCGA data. The terms outlined in this Data Use Certification do not apply to Open Access data. Users should review dbGaP policies for securely storing and sharing human data submitted to NIH under the *Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS)*, but users shall note that TCGA data are not GWAS data and have slightly different policies associated with them as outlined in this Data Use Certification. Implicit in the establishment of these data management systems is that scientific progress in genomic research will be greatly enhanced if the data are readily available to all scientific investigators and shared in a manner consistent with the human subjects protocols under which participant's data and samples were donated to TCGA.

Access to human genomic data will be provided to research investigators who, along with their institutions, have certified their agreement with the expectations and terms of access detailed below. It is the intent of NIH, National Cancer Institute (NCI-NIH), and the National Human Genome Institute (NHGRI-NIH) that approved users of TCGA datasets recognize restrictions on data use established by the submitting institution through the Institutional Certification and stated on TCGA's Publication Guidelines page (http://cancergenome.nih.gov/publications/publicationguidelines).

Definitions of terminology used in this document are found in the Appendix.

The parties to this agreement include: the Principal Investigator (PI) requesting access to the genomic study dataset (the "Approved User"), his/her home institution as represented by the Institutional Signing Official designated through the eRA Commons system (the "Requester"), and the relevant NIH Institute or Center (IC). The effective date of this agreement shall be the Project Approval Date, as specified on the Data Access Committee (DAC) approval notification.

TERMS OF ACCESS

1. Research Use

The Requester agrees that if access is approved, (1) the PI named in the Data Access Request (DAR) and (2) those named in the "Senior/Key Person Profile" section of the DAR, including the Information Technology Director or his/her designee, and any trainee, employee, or contractor[1] working on the proposed research

[1] If contractor services are to be utilized, the principal investigator (PI) requesting the data must provide a brief description of the services that the contractor will perform for the PI (e.g., data cleaning services) in the research use statement of the DAR. Additionally, the Key Personnel section of the DAR must include the name of the contractor's employee(s) who will conduct the work. These requirements apply whether the contractor carries out the work at the PI's facility or at the contractor's facility. In addition, the PI is expected to include in any contract agreement requirements to ensure that any of the contractor's employees who have access to the data adhere to the *NIH Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies*, the data use certification agreement, and the dbGaP data.

project under the direct oversight of these individuals, shall become Approved Users of the requested dataset(s). Research use will occur solely in connection with the approved research project described in the DAR, which includes a 1-2 paragraph description of the research objectives and design. New uses of these data outside those described in the DAR will require submission of a new DAR; modifications to the research project will require submission of an amendment to this application (e.g., adding or deleting collaborators from the same institution, adding datasets to an approved project). Access to the requested dataset(s) is granted for a period of 1 year as defined below.

Contributing Investigators, or their direct collaborators, who provided the data or samples used to generate an NIH genomic dataset and who have Institutional Review Board (IRB) approval, if applicable, for broad use of the data are exempt from the limitation on the scope of the research use as defined in the DAR.

2. Requester and Approved User Responsibilities

The Requester agrees through the submission of the DAR that the PI named in the DAR has reviewed and understands the principles for responsible research use and data handling of the genomic datasets as defined in the NIH GWAS Data Sharing Policy and as detailed in this Data Use Certification (DUC) agreement. The Requester and Approved Users further acknowledge that they are responsible for ensuring that all uses of the data are consistent with federal, state, and local laws and regulations and any relevant institutional policies. The Requester certifies that the Approved User is in good standing (i.e., no known sanctions) with the institution, relevant funding agencies, and regulatory agencies and is eligible to conduct independent research (i.e., is not a postdoctoral fellow, student, or trainee).

Through submission of the DAR, the PI agrees to submit either a project renewal or close-out request prior to the expiration date of the 1-year data access period. The PI also agrees to submit an annual progress update or a final progress report at the 1-year anniversary of the DAR, as described under *Research Use Reporting* below. Failure to submit a renewal or complete the close-out process, including confirmation of data destruction by the Signing Official, may result in termination of all current data access and/or suspension of the Approved User and all associated key personnel and collaborators from submitting new DARs for a period to be determined by NIH. Repeated violations or unresponsiveness to NIH requests may result in further measures affecting the Requester.

Approved Users who may have access to personal identifying information for research participants in the original study at their institution or through their collaborators may be required to have IRB approval. By approving and submitting the attached DAR, the Institutional Signing Official provides assurance that relevant institutional policies and applicable federal, state, and local laws and regulations (if any) have been followed, including IRB approval if required. The Institutional Signing Official also assures through the approval of the DAR that other institutional departments with relevant authorities (e.g., those overseeing human subjects research, information technology, or technology transfer) have reviewed the relevant sections of the NIH GWAS Data Sharing Policy and the associated procedures and are in agreement with the principles defined.

NIH anticipates that TCGA datasets will be regularly updated with additional information. Unless otherwise indicated, all statements herein are presumed to be true and applicable to the access and use of all versions of these datasets.

3. Public Posting of Approved Users' Research Use Statement

PIs agree that if they become Approved Users, information about themselves and their approved research use will be posted publicly on the dbGaP website. The information includes the Approved User's name and institution, project name, research use statement, and a non-technical summary of the research use statement.

Security Best Practices Requirements. Note that any scientific collaborators, including contactors, who are not at the same institution as the PI must submit their own DAR.

In addition, citations of publications resulting from the use of NIH genomic datasets may be posted on NIH data repository websites.

4. Non-Identification

Approved Users agree not to use the requested datasets, either alone or in concert with any other information, to identify or contact individual participants from whom data and/or samples were collected. This provision does not apply to research investigators operating with specific IRB approval, pursuant to 45 CFR 46, to contact individuals within datasets or to obtain and use identifying information under an IRB approved research protocol. All investigators conducting "human subjects research" within the scope of 45 CFR 46 must comply with the requirements contained therein.

5. Non-Transferability

The Requester and Approved Users agree to retain control of the data and further agree not to distribute data obtained through the DAR to any entity or individual not covered in the submitted DAR. If Approved Users are provided access to NIH genomic datasets for inter-institutional collaborative research described in the research use statement of the DAR, and all members of the collaboration are also Approved Users through their home institution(s), data obtained through this DAR may be securely transmitted within the collaborative group.

The Requester and Approved Users acknowledge responsibility for ensuring the review and agreement to the terms within this DUC and the appropriate research use of NIH genomic data by research staff associated with any approved project, subject to applicable laws and regulations. NIH genomic datasets obtained through this DAR, in whole or in part, may not be sold to any individual at any point in time for any purpose.

Approved Users agree that if they change institutions during the access period they will complete the DAR close-out process before moving to their new institution. A new DAR and DUC, in which the new institution agrees to the NIH GWAS Data Sharing Policy, must be approved by the relevant NIH DAC(s) before data access resumes. As part of the close-out process, any versions of the data stored at the prior institution should be destroyed and destruction confirmed in writing by the Signing Official, as described below. However, with advance written notice and approval by the NCI/NHGRI TCGA DAC to transfer responsibility for the approved research project to another Approved User from the PI's prior institution, the data do not need to be destroyed.

6. Data Security and Data Release Reporting

The Requester and Approved Users, including the institutional Information Technology Director or his/her designee, acknowledge NIH's expectation that they have reviewed and agree to handle the requested dataset(s) according to the current *dbGaP Security Best Practices*, including its detailed description of requirements for security and encryption. These include, but are not limited to:

- All Approved Users have completed all required computer security training required by their institution, for example, the http://irtsectraining.nih.gov/, or the equivalent.
- The data will always be physically secured (e.g., through camera surveillance, locks on doors/computers, security guard).
- Servers must not be accessible directly from the internet, (e.g., they must be behind a firewall or not connected to a larger network) and unnecessary services should be disabled.
- Use of portable media (e.g., CD, flash drive or laptop) is discouraged, but if necessary then they should be encrypted consistent with applicable law.
- Updated anti-virus/anti-spyware software is used.
- Security auditing/intrusion detection software that regularly scans and detects potential data intrusions should be in place.
- Strong password policies for file access are used.

- All copies of the dataset are destroyed, as permitted by law and local institutional policies, whenever any of the following occurs:
  - the DUC expires and renewal is not sought;
  - access renewal is not granted;
  - DAC requests destruction of the dataset; and
  - continued use of the data would no longer be consistent with the DUC.

In addition, the Requester and Approved Users agree to keep the data secure and confidential at all times and to adhere to information technology practices in all aspects of data management to assure that only authorized individuals can gain access to NIH genomic datasets. This agreement includes the maintenance of appropriate controls over any copies or derivatives of the data obtained through this Data Access Request.

Requesters and Approved Users agree to notify the NCI/NHGRI TCGA DAC of any unauthorized data sharing, breaches of data security, violations in the presentation and publication embargo period, or inadvertent data releases that may compromise data confidentiality within 24 hours of when the incident is identified. As permitted by law, notifications should include any known information regarding the incident and a general description of the activities or process in place to define and remediate the situation fully. Within 3 business days of the NCI/NHGRI TCGA DAC notification, the Requester, through the Approved User and the Institutional Signing Official, agree to submit to the NCI-NHGRI TCGA Data Access Committee a more detailed written report including the date and nature of the event, actions taken or to be taken to remediate the issue(s), and plans or processes developed to prevent further problems, including specific information on timelines anticipated for action.

All notifications and written reports of data security incidents should be sent to:

tcgadac@mail.nih.gov
NCI/NHGRI TCGA Data Access Committee
Marked: URGENT

The NCI-NIH, NHGRI-NIH, and the NIH, or another entity designated by NIH may, as permitted by law, also investigate any data security incident. Approved Users and their associates agree to support such investigations and provide information, within the limits of applicable local, state, and federal laws and regulations. In addition, Requesters and Approved Users agree to work with the NCI/NHGRI TCGA DAC and NIH to assure that plans and procedures that are developed to address identified problems are mutually acceptable and consistent with applicable law.

7. Intellectual Property

By requesting access to genomic dataset(s), the Requester and Approved Users acknowledge the intent of the NIH that anyone authorized for research access through the attached DAR follow the intellectual property (IP) principles in the NIH GWAS Data Sharing Policy as summarized below:

> Achieving maximum public benefit is the ultimate goal of data distribution through the NIH genomic data repositories. The NIH encourages broad use of NIH-supported genotype-phenotype data that is consistent with a responsible approach to management of intellectual property derived from downstream discoveries, as outlined in the NIH's Best Practices for the Licensing of Genomic Inventions and its Research Tools Policy (see http://www.ott.nih.gov/).
>
> The NIH considers these data as pre-competitive and urges Approved Users to avoid making IP claims derived directly from the genomic dataset(s). It is expected that these NIH-provided data, and conclusions derived therefrom, will remain freely available, without requirement for licensing. However, the NIH also recognizes the importance of the subsequent development of IP on downstream

discoveries, especially in therapeutics, which will be necessary to support full investment in products to benefit the public.

8. Research Dissemination and Acknowledgement of NIH Genomic Study Datasets

It is NIH's intent to promote the dissemination of research findings from NIH genomic dataset(s) as widely as possible through scientific publication or other appropriate public dissemination mechanisms. Approved Users are strongly encouraged to publish their results in peer-reviewed journals and to present research findings at scientific meetings and are asked to adhere to the TCGA Publication Guidelines and Moratoria (http://cancergenome.nih.gov/publications/publicationguidelines).

Approved Users agree to acknowledge the TCGA Research Network in all oral and written presentations, disclosures, and publications resulting from any analyses of the data. A sample acknowledgment statement for The Cancer Genome Atlas dataset(s) follows:

> "The results published here are in whole or part based upon data generated by The Cancer Genome Atlas managed by the NCI and NHGRI. Information about TCGA can be found at http://cancergenome.nih.gov."

9. Research Use Reporting

To assure adherence to NIH policies and procedures for genomic data, Approved Users agree to provide annual progress updates on how these data have been used, including presentations, publications, and the generation of intellectual property. This information helps NIH evaluate program activities and may be considered by the NIH governance committees as part of NIH's effort to provide ongoing oversight and management of NIH genomic data sharing activities.

Progress updates are provided as part of the annual project renewal or project close-out processes, prior to the expiration of the 1-year data access period. Approved Users who are seeking renewal or close-out of a project agree to complete the appropriate online forms and provide specific information such as publications or presentations that resulted from the use of the requested dataset(s), a summary of any plans for future research use, any violations of the terms of access described within this DUC and the implemented remediation, and information on any downstream intellectual property generated from the data. Approved Users also may include general comments regarding topics such as the effectiveness of the data access process (e.g., ease of access and use), appropriateness of data format, challenges in following the policies, and suggestions for improving data access or the program in general.

Note that any inadvertent or inappropriate data release incidents should be reported to the NCI/NHGRI TCGA DAC according to the agreements and instructions under Term 6.

10. Non-Endorsement, Indemnification

The Requester and Approved Users acknowledge that although all reasonable efforts have been taken to ensure the accuracy and reliability of TCGA data, the NIH, the NCI/NHGRI TCGA DAC, and Contributing Investigators do not and cannot warrant the results that may be obtained by using any data included therein. NIH, the NCI/NHGRI TCGA DAC, and all contributors to these datasets disclaim all warranties as to performance or fitness of the data for any particular purpose.

No indemnification for any loss, claim, damage, or liability is intended or provided by any party under this agreement. Each party shall be liable for any loss, claim, damage, or liability that said party incurs as a result of

its activities under this agreement, except that NIH, as an agency of the United States, may be liable only to the extent provided under the Federal Tort Claims Act, 28 USC 2671 et seq.

11. Termination and Violations

This DUC will be in effect for a period of 1 year from the date the dataset(s) are made accessible to the Approved User ("Approved Access Date"). At the end of the access period, Approved Users agree to report progress, and renew access or close-out the project. Upon project closure-out, Approved Users agree to destroy all copies of the requested dataset(s), except as required by publication practices or law to retain them. Copies of NIH genomic dataset(s) may not need to be destroyed if, with advance notice and approval by the NCI/NHGRI TCGA DAC, the project has been transferred to another Approved User at the same institution. In this case, documentation must be provided that other Approved Users are using the dataset(s) under a DAC-approved DAR.

The Requester and Approved User acknowledge that the NIH or the NCI/NHGRI TCGA may terminate this agreement and immediately revoke access to all TCGA genomic datasets at any time if the Requester is found to be no longer in agreement with the policies, principles and procedures of the NIH and the NCI/NHGRI TCGA DAC.

****

By submission of the attached DAR,

- ☐ The Requester through the Institutional Signing Official attests to the Approved Users' qualifications for access to and use of NIH genomic dataset(s) and agrees to the NIH principles, policies, and procedures for the use of the requested datasets as articulated in this document, including the potential termination of access should any of these terms be violated.

- ☐ Requesters and the PI further acknowledge that they have shared this document, and the NIH GWAS Data Sharing Policy, and procedures for access and use of genomic datasets with any Approved Users, appropriate research staff, and all other Key Personnel and collaborators identified in the DAR.

- ☐ Institutional Signing Officials acknowledge that they have considered the relevant NIH GWAS policies and procedures, that they have shared this document and the relevant policies and procedures with appropriate institutional departments, and have assured compliance with local institutional policies related to technology transfer, information technology, privacy, and human subjects research.

- ☐ The Requestor and PI acknowledge that they reviewed and understand the TCGA-specific publication policies as outlined in Section 8.

- ☐ Institutional Signing Officials acknowledge that their institute is solely responsible for the conduct of all individuals who have access to the data under the DAR, including investigators, contractor staff (both on and off-site) and trainees.

APPENDIX

DEFINITIONS

Approved User: A user approved by the relevant Data Access Committee(s) to access one or more datasets for a specified period of time and only for the purposes outlined in the investigator's approved research use statement. Staff members and trainees under the direct supervision of the Approved User are also Approved Users and must abide by the terms laid out in the Data Use Certificate agreement.

Collaborator: An individual who is not under the direct supervision of the principal investigator (PI) (e.g., not a member of the PI's laboratory) who assists with the PI's dbGaP research project. Internal collaborators are employees of the Requester and work at the same location/campus as the PI. External collaborators are not employees of the Requester and/or do not work at the same location as the PI, and consequently must be independently approved to access dbGaP data.

Contributing Investigator: An investigator who submitted a genomic dataset to an NIH-designated data repository (e.g., dbGaP).

Data Access Request (DAR): A request submitted to a Data Access Committee for a specific project specifying the data to which access is sought, the planned research use, and the names of collaborators and the Information Technology Director. The DAR is signed by the investigator requesting the data and her/his Institutional Signing Official. Collaborators and project team members on a request must be from the same institution or organization.

Data Derivative: any data including individual-level data or aggregate genomic data that stems from the original dataset deposited (e.g. imputed or annotated data) in NIH-designated data repositories (e.g., dbGaP). Summary information that is expected to be shared through community publication practices in not included in this term.

Data Use Agreement (DUC): An agreement between the Approved User, the Requestor, and NIH regarding the terms associated with dbGaP data access and the expectations for use of dbGaP datasets.

dbGaP Approved User Code of Conduct: Key principles and practices agreed to by all research investigators requesting access to NIH controlled-access genomic data. The elements within the Code of Conduct reflect the terms of access in the Data Use Certification agreement. Failure to abide by the Code of Conduct may result in revocation of an investigator's access to any and all approved datasets. (See https://dbgap.ncbi.nlm.nih.gov/aa/GWAS_Code_of_Conduct.html)

Information Technology (IT) Director: Individual with the necessary expertise and authority to vouch for the IT capacities at an academic institution, company, or other research entity and the ability of that institution to comply with NIH data security expectations. The IT Director is to be included as key personnel in the Data Access Request.

Institutional Certification: Certification by the Institution that delineates, among other items, the appropriate research uses of the data and the uses that are specifically excluded by the relevant informed consent documents. (See http://grants.nih.gov/grants/guide/notice-files/NOT-OD-07-088.html)

Institutional Signing Official: Generally, a senior official at an institution with the authority to sign on behalf of the submitting investigator or an investigator who has submitted a Data Access Request or Project Request to NIH, authorized to enter their institution into a legally binding contract, and who is credentialed through the eRA Commons system.

Progress Update: Information included with the annual Data Access Request (DAR) renewal or close-out

summarizing the analysis of dbGaP datasets obtained through the DAR and any publications and presentations derived from the work.

Project Closeout: Closeout of a research project that used controlled-access data from an NIH-designated data repository (e.g., dbGaP) and confirmation of data destruction when the research is completed and/or discontinued. (See http://www.ncbi.nlm.nih.gov/books/NBK63627/)

Project Renewal: Renewal of a principal investigator's access to controlled-access datasets for a prior-approved project before the expiration date of a Data Access Request or Project Request. (See http://www.ncbi.nlm.nih.gov/books/NBK63627/#DArequest.the_research_related_to_my_dbg)

Requester: The home institution or organization of the principal investigator that applies to dbGaP for access to NIH genomic data.

Senior/Key Persons: Collaborators at the home institution of the data submitter or requester, such as the Information Technology Director.

Addendum to the Data Use Certification Agreement
Modification of Data Security Terms and Best Practices

Effective for all dbGaP Data Access Requests submitted on or after March 23, 2015,
Section 6 of the Data Use Certification Agreement is replaced in its entirety by the following:

6. Data Security and Data Release Reporting

The Requester and Approved Users, including the institutional IT Director, acknowledge NIH's expectation that they have reviewed and agree to manage the requested dataset(s) according to the current NIH Security Best Practices for Controlled-Access Data Subject to the GDS Policy and the institutional IT security requirements and policies, and that the institution's IT security requirements and policies are sufficient to protect the confidentiality and integrity of the NIH controlled-access data entrusted to the Requester.

If approved by NIH to use cloud computing for the proposed research project, as outlined in the Research and Cloud Computing Use Statements of the Data Access Request, the Requester acknowledges that the IT Director has reviewed and understands the cloud computing guidelines in the NIH Security Best Practices for Controlled-Access Data Subject to the GDS Policy.

Requesters and PIs agree to notify the TCGA DAC of any unauthorized data sharing, breaches of data security, or inadvertent data releases that may compromise data confidentiality within 24 hours of when the incident is identified. As permitted by law, notifications should include any known information regarding the incident and a general description of the activities or process in place to define and remediate the situation fully. Within 3 business days of the TCGA DAC notification, the Requester, through the PI and the Institutional Signing Official, agree to submit to the TCGA Data Access Committee a detailed written report including the date and nature of the event, actions taken or to be taken to remediate the issue(s), and plans or processes developed to prevent further problems, including specific information on timelines anticipated for action.

All notifications and written reports of data security incidents should be sent to:
TCGA Data Access Committee URGENT: tcgadac@mail.nih.gov
GDS mailbox: gds@mail.nih.gov

NIH, or another entity designated by NIH may, as permitted by law, also investigate any data security incident. Approved Users and their associates agree to support such investigations and provide information, within the limits of applicable local, state, and federal laws and regulations. In addition, Requesters and Approved Users agree to work with the TCGA and NIH to assure that plans and procedures that are developed to address identified problems are mutually acceptable and consistent with applicable law.