MODEL DATA USE CERTIFICATION AGREEMENT

Sequence-based analysis of human breast tissues

(October 27, 2015, version)

PLEASE COMPLETE OR DELETE HIGHLIGHTED SECTIONS AS APPROPRIATE

INTRODUCTION AND STATEMENT OF POLICY

The National Institutes of Health (NIH) has established NIH-designated data repositories (e.g., database of Genotypes and Phenotypes (dbGaP), Sequence Read Archive (SRA), NIH Established Trusted Partnerships) for securely storing and sharing controlled-access human data submitted to NIH under the NIH Genomic Data Sharing (GDS) Policy. Because the volume of human genomic and phenotypic data contained in these repositories is substantial and, in some instances, potentially sensitive (e.g., data related to the presence or risk of developing particular diseases or conditions and information regarding family relationships or ancestry), data must be shared in a manner consistent with the research participants' informed consent, and the confidentiality of the data and the privacy of participants must be protected.

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 and the NCI that approved users of controlled-access datasets obtained through this DAR recognize any restrictions on data use established by the submitting institution through the Institutional Certification and stated on the dbGaP study page.

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 (an "Approved User"), the PI's 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 and any trainee, employee, or contractor[1] working on the proposed research project

[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

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. Investigators interested in using cloud computing[2] for data storage and analysis must indicate in their Data Access Request (DAR) that they are requesting permission to use cloud computing and identify the cloud service provider (CSP)[3] or providers and/or Private Cloud System (PCS) that they propose to use. They must also submit a Cloud Computing Use Statement as part of the DAR that describes the type of service and how it will be used to carry out the proposed research described in the Research Use Statement of the DAR. If investigators plan to collaborate with investigators outside their own institution, the investigators at each external site must submit an independent DAR using the same project title and Research Use Statement, and if using the cloud, Cloud Computing Use Statement. 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 controlled-access datasets subject to the GDS Policy 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.

NCI Specific Terms:
None

Data Use Limitations:
The eNCI DAC recognizes the Data Use Limitations of each consent group stated on the dbGaP study page (*to show the content, mouse over title of the consent group in the Authorized Access section*).

2. Requester and Approved User Responsibilities

The Requester agrees through the submission of the DAR that the PI named has reviewed and understands the principles for responsible research use and data management of the genomic datasets as defined in the NIH Security Best Practices for Controlled-Access Data Subject to the GDS Policy. The

---

agreement requirements to ensure that any of the contractor's employees who have access to the data adhere to the GDS Policy, this Data Use Certification Agreement, and the NIH Security Best Practices for Controlled-Access Data Subject to the GDS Policy. Note that any scientific collaborators, including contractors**,** who are not at the same institution as the PI must submit their own DAR.

[2] The National Institute for Standards and Technology defines cloud computing as a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. For more information see: http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf

[3] The National Institute for Standards and Technology defines a cloud service provider as a company that offers some component of cloud computing to other businesses or individual, typically Infrastructure as a Service (IaaS), Software as a Service (SaaS) or Platform as a Service (PaaS), as defined by the National Institute of Standards and Technology. For more information see: http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf

Requester and Approved Users further acknowledge that they are responsible for ensuring that all uses of the data are consistent with national, tribal, and state laws and regulations, as appropriate, as well as relevant institutional policies and procedures for managing sensitive genomic and phenotypic data. The Requester certifies that the PI 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). The Requester and all Approved Users may use the dataset(s) only in accordance with the parameters described on the dbGaP website for the appropriate research use, as well as any limitations on such use, of the dataset(s) and as described in the DAR and as required by law.

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 prior to the 1-year anniversary of the DAR, as described under *Research Use Reporting* below. Failure to submit a renewal or to complete the close-out process, including confirmation of data destruction by the Institutional Signing Official, may result in termination of all current data access and/or suspension of the PI and all associated 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 compliance measures affecting the Requester.

Approved Users who 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 national, tribal, and state laws and regulations, as applicable, 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 GDS Policy and the associated procedures and are in agreement with the principles defined.

In some cases, NIH anticipates that controlled-access datasets subject to the GDS Policy will be 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. In addition, and if applicable, this information may include the Cloud Computing Use Statement and name of the CSP or PCS. Citations of publications resulting from the use of controlled-access datasets obtained through this DAR may also be posted on the dbGaP website.

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 NIH controlled-access datasets obtained through the attached DAR and any derivatives[4] of controlled-access datasets and further agree not to distribute controlled-access datasets and derivatives of controlled-access datasets to any entity or individual not identified in the submitted DAR. If Approved Users are provided access to controlled-access datasets subject to the GDS Policy 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 the attached DAR may be securely transmitted within the collaborative group. Approved Users are expected to follow all data security practices and other terms of use defined in this agreement, the NIH Security Best Practices for Controlled-Access Data Subject to the GDS Policy and the Requester's IT security requirements and policies.

The Requester and Approved Users acknowledge responsibility for ensuring the review and agreement to the terms within this Data Use Certification Agreement and the appropriate research use of controlled-access data obtained through the attached DAR and any derivatives of controlled-access datasets by research staff associated with any approved project, subject to applicable laws and regulations. Controlled-access datasets obtained through the attached DAR and any derivatives of controlled-access datasets, in whole or in part, may not be sold to any individual at any point in time for any purpose.

PIs 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 Data Use Certification, in which the new Requester agrees to the GDS Policy, must be approved by the relevant NIH DAC(s) before controlled-access data may be re-accessed. As part of the close-out process, all copies and versions of the datasets retrieved from NIH-designated controlled-access databases as well as any derivatives of controlled-access datasets stored at the institution and/or CSP must be destroyed and destruction confirmed by the Signing Official, as described below.

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 controlled-access dataset(s) and any derivatives of controlled-access datasets 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

[4] Any data containing individual-level information that are generated or inferred from controlled-access datasets (e.g. imputed or annotated data) obtained from NIH-designated data repositories (e.g., dbGaP).

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 eNCI 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 eNCI DAC notification, the Requester, through the PI and the Institutional Signing Official, agree to submit to the eNCI 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:
eNCI Data Access Committee URGENT email: ncidac-urgent@mail.nih.gov

GDS mailbox: gds@mail.nih.gov

NCI, 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 eNCI 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 GDS Policy as summarized below:

> Achieving maximum public benefit is the ultimate goal of data distribution through the NIH-designated 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 Best Practices for the Licensing of Genomic Inventions and its Research Tools Policy.
>
> 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 Controlled-Access Datasets Subject to the GDS Policy

It is NIH's intent to promote the dissemination of research findings from controlled-access dataset(s) subject to the GDS Policy 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.

Approved Users agree to acknowledge the Contributing Investigator(s) who submitted data from the original study to dbGaP, the primary funding organization that supported the Contributing Investigators, and the NIH-designated data repository, in all oral and written presentations, disclosures, and publications resulting from any analyses of controlled-access data obtained through the attached DAR. Approved Users further agree that the acknowledgment shall include the dbGaP accession number to the specific version of the dataset(s) analyzed. A sample acknowledgment statement for the Sequence-based analysis of human breast tissues dataset(s) follows:

The data sets used for the analyses described in this manuscript were obtained from dbGaP at www.ncbi.nlm.nih.gov/gap through dbGaP accession number phs000676.

9. Research Use Reporting

To assure adherence to NIH policies and procedures for genomic data, PIs agree to provide annual progress updates as part of the annual project renewal or project close-out processes, prior to the expiration of the 1-year data access period. PIs who are seeking renewal or close-out of a project agree to complete the appropriate online forms and provide specific information such as how the data have been used, including publications or presentations that resulted from the use of the requested dataset(s), a summary of any plans for future research use (if the requester is seeking renewal), any violations of the terms of access described within this Data Use Certification Agreement and the implemented remediation, and information on any downstream intellectual property generated from the data. PIs 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. Information provided in the progress updates helps NIH evaluate program activities and may be considered by the NIH GDS governance committees as part of NIH's effort to provide ongoing oversight and management of data sharing activities subject to the GDS Policy.

Note that any inadvertent or inappropriate data release incidents should be reported to the eNCI 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 controlled-access data obtained through the attached DAR, the NIH, the eNCI, and Contributing Investigators do not and cannot warrant the results that may be obtained by using any data included therein. NIH, the eNCI, 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

Upon project close-out, all Approved Users agree to destroy all copies, versions, and derivations of the dataset(s) retrieved from NIH-designated controlled-access databases, on both local servers and hardware, and if cloud computing was used, delete the data and cloud images from cloud computing provider storage, virtual and physical machines, databases, and random access archives, except as required by publication practices, institutional policies, or law to retain them.

The Requester and PI acknowledge that the NIH or the eNCI may terminate this agreement and immediately revoke access to all controlled-access datasets subject to the GDS Policy 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 eNCI.

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 Principal Investigator (PI)'s approved Research Use Statement. The Information Technology (IT) Director indicated on the Data Access Request, as well as any staff members and trainees under the direct supervision of the PI 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 PI (e.g., not a member of the PI's laboratory) who assists with the PI's research project involving controlled-access data subject to the GDS Policy. 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 controlled-access data subject to the GDS Policy.

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

Cloud Computing: A model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

Cloud Service Provider (CSP): A company that offers some component of cloud computing to other businesses or individual, typically Infrastructure as a Service (IaaS), Software as a Service (SaaS) or Platform as a Service (PaaS).

Data Access Request (DAR): A request submitted to a Data Access Committee for a specific "consent group" specifying the data to which access is sought, the planned research use, and the names of collaborators and the IT Director. The DAR is signed by the PI 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 Use Certification Agreement (DUC): An agreement between the Approved Users, the Requester, and NIH regarding the terms associated with access of controlled-access datasets subject to the GDS Policy and the expectations for use of these datasets.

Approved User Code of Conduct: Key principles and practices agreed to by all research investigators requesting access to controlled-access data subject to the GDS Policy. 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/GDS_Code_of_Conduct.html)

Information Technology (IT) Director: Generally, a senior IT official with the necessary expertise and authority to affirm the IT capacities at an academic institution, company, or other research entity. The IT Director is expected to have the authority and capacity to ensure that the NIH Security Best

Practices for Controlled-Access Data Subject to the NIH GDS Policy and the institution's IT security requirements and policies are followed by the Approved Users.

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: The Signing Official has institutional authority to legally bind the institution in grants administration matters. The label, "Signing Official," is used in conjunction with the NIH eRA Commons. The individual fulfilling this role may have any number of titles in the grantee organization, but is typically located in its Office of Sponsored Research or equivalent. The Signing Official reviews Data Access Request applications submitted by Principal Investigators and on behalf of the institution, agrees to adhere to the terms described in the Data Use Certification Agreement if the application is submitted to NIH.

Private Cloud System (PCS): A cloud infrastructure provisioned for exclusive use by a single organization comprising multiple consumers (e.g., business units). It may be owned, managed, and operated by the organization, a third party, or some combination of them, and it may exist on or off premises.

Progress Update: Information included with the annual Data Access Request (DAR) renewal or close-out summarizing the analysis of controlled-access datasets obtained through the DAR and any publications and presentations derived from the work.

Project Close-out: Termination 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. The project close-out process is completed in the dbGaP Auhorized Access System.

Project Renewal: Renewal of a PI's access to controlled-access datasets for a prior-approved project.

Requester: The home institution or organization of the PI that applies to dbGaP for access to controlled-access data subject to the GDS Policy.

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 eNCI 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 eNCI DAC notification, the Requester, through the PI and the Institutional Signing Official, agree to submit to the eNCI 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:
eNCI Data Access Committee URGENT: ncidac-urgent@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 eNCI and NIH to assure that plans and procedures that are developed to address identified problems are mutually acceptable and consistent with applicable law.