# DATA USE CERTIFICATION AGREEMENT

***(December 15, 2023, version)***

This Data Use Certification Agreement outlines the terms of use for requested controlled-access datasets maintained in NIH-designated data repositories under the NIH Genomic Data Sharing Policy (e.g., the NIH database of Genotypes and Phenotypes (dbGaP)). The Addendum to this Agreement outlines additional terms and information which are specific to each requested dataset such as:

- Data Use Limitation(s)
- Sponsoring NIH Institute or Center
- Responsible Data Access Committee
- Study Description
- Suggested Acknowledgement Statement

## 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 maintained 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. NIH expects that, through Data Access Request (DAR) process, approved users of controlled-access datasets recognize any restrictions on data use established by the Submitting Institutions through the Institutional Certification, and as stated on the dbGaP study page.

Definitions of the underlined terminology in this document are found in section 14.

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 (the "Requester") as represented by the Institutional Signing Official designated through the eRA Commons system, and the NIH. The effective date of this Agreement shall be the DAR Approval Date, as specified in the notification of approval of the Data Access Committee (DAC).

## TERMS OF ACCESS

### 1. Research Use

The Requester agrees that if access is approved, (1) the PI named in the 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 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 proposed research (i.e., a Research Use Statement). Investigators interested in using Cloud Computing for data storage and analysis must request permission to use Cloud Computing in the DAR and identify the Cloud Service Provider (CSP) 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 as described in the Research Use Statement. If the Approved Users plan to collaborate with investigators outside the Requester, 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 Requester Collaborators from the Requester, adding datasets to an approved project). Access to the requested dataset(s) is granted for a period of **one (1) year**, with the option to renew access or close-out a project at the end of that year.

Submitting Investigator(s), or their collaborators, who provided the data or samples used to generate controlled-access datasets subject to the NIH GDS Policy and who have Institutional Review Board (IRB) approval and who meet any other study specific terms of access, 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 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 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 any Approved Users may use the dataset(s) only in accordance with the parameters described on the study page and in the Addendum to this Agreement for the appropriate research use, as well as any limitations on such use, of the dataset(s), as described in the DAR, and as required by law.

Through the submission of this DAR, the Requester and Approved Users acknowledge receiving and

[1] If contractor services are to be utilized, 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 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 Requester must submit their own DAR.

reviewing a copy of the Addendum which includes Data Use Limitation(s) for each dataset requested. The Requester and Approved Users agree to comply with the terms listed in the Addendum.

Through submission of the DAR, the PI and Requester agree to submit a Project Renewal or Project Close-out prior to the expiration date of the one (1) year data access period. The PI also agrees to submit an annual Progress Update prior to the one (1) year anniversary[2] of the project, as described under *Research Use Reporting* (Term 11) below.

By approving and submitting the attached DAR, the Institutional Signing Official provides assurance that relevant institutional policies and applicable local, state, tribal, and federal laws and regulations, as applicable, have been followed, including IRB approval, if required. Approved Users may be required to have IRB approval if they have access to personal identifying information for research participants in the original study at their institution, or through their collaborators. 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, technology transfer) have reviewed the relevant sections of the NIH GDS Policy and the associated procedures and are in agreement with the principles defined.

The Requester acknowledges that controlled-access datasets subject to the NIH GDS Policy may be updated to exclude or include 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**

The PI agrees that information about themselves and the approved research use will be posted publicly on the dbGaP website. The information includes the PI's name and Requester, 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. Approved Users also agree not to generate information (e.g., facial images or comparable representations) that could allow the identities of research participants to be readily ascertained. These provisions do 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 including any Approved User conducting "human subjects research" within the scope of 45 CFR 46 must comply with the requirements contained therein.

**5. Certificate of Confidentiality**

[2] The project anniversary date can be found in "My Projects" after logging in to the dbGaP authorized-access portal.

Effective June 11, 2017 the Certificate of Confidentiality (Certificate) issued for the database of Genotypes and Phenotypes (dbGaP) is subject to the requirements of section 301(d) of the Public Health Service Act (42 U.S.C. 241(d)). Moreover, as of October 1, 2017 dbGaP is required to adhere to the *NIH Policy for Issuing Certificates of Confidentiality* (NOT-OD-17-109). Therefore, Approved Users of dbGaP, whether or not funded by the NIH, who access a copy of information protected by a Certificate held by dbGaP, are also subject to the requirements of the Certificate of Confidentiality and subsection 301(d) of the Public Health Service Act.

Under Section 301(d) of the Public Health Service Act and the *NIH Policy for Issuing Certificates of Confidentiality*, recipients of a Certificate of Confidentiality shall not:

- Disclose or provide, in any Federal, State, or local civil, criminal, administrative, legislative, or other proceeding, the name of such individual or any such information, document, or biospecimen that contains identifiable, sensitive information about the individual and that was created or compiled for purposes of the research, unless such disclosure or use is made with the consent of the individual whom the information, document, or biospecimen pertains; or
- Disclose or provide to any other person not connected with the research the name of such an individual or any information, document, or biospecimen that contains identifiable, sensitive information about such an individual and that was created or compiled for purposes of the research.

Disclosure is permitted only when:

- Required by Federal, State, or local laws (e.g., as required by the Federal Food, Drug, and Cosmetic Act, or state laws requiring the reporting of communicable diseases to State and local health departments), excluding instances of disclosure in any Federal, State, or local civil, criminal, administrative, legislative, or other proceeding;
- Necessary for the medical treatment of the individual to whom the information, document, or biospecimen pertains and made with the consent of such individual;
- Made with the consent of the individual to whom the information, document, or biospecimen pertains; or
- Made for the purposes of other scientific research that is in compliance with applicable Federal regulations governing the protection of human subjects in research.

**6. Non-Transferability**

The Requester and Approved Users agree to retain control of NIH controlled-access datasets obtained through the attached DAR, and any Data Derivatives of controlled-access datasets, and further agree not to distribute controlled-access datasets and Data Derivatives of controlled-access datasets to any entity or individual not identified in the submitted DAR. If the Approved Users are provided access to controlled-access datasets subject to the NIH 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. Each Approved User will 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 Agreement and the appropriate research use of controlled-access data obtained through the attached DAR and any Data Derivatives of controlled-access datasets by research staff associated with any approved project, subject to applicable laws and regulations. Requester and Approved Users agree that controlled-access datasets obtained through the attached DAR and any Data 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.

The PI agrees that if they change institutions during the access period they will complete the Project Close-out process (See Term 13 for more details) before moving to their new institution. A new DAR, in which the new Requester agrees to the Data Use Certification Agreement and the Genomic Data User Code of Conduct, must be approved by the relevant NIH DAC(s) before controlled-access data may be re-accessed.

**7. Data Security and Unauthorized Data Release**

The Requester and Approved Users, including the Requester's IT Director, acknowledge NIH's expectation that they have reviewed and agree to manage the requested controlled-access dataset(s) and any Data Derivatives of controlled-access datasets according to NIH's expectations set forth in the current NIH Security Best Practices for Controlled-Access Data Subject to the GDS Policy and the Requester's IT security requirements and policies. The Requester, including the Requester's IT Director, agree that the Requester'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 NIH GDS Policy.

The Requester and PI agree to notify the appropriate DAC(s) 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 DAC notification, the Requester agrees to submit to the DAC(s) 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. The Requester agrees to provide documentation verifying that the remediation plans have been implemented. Repeated violations or unresponsiveness to NIH requests may result in further compliance measures affecting the Requester.

**All notifications and written reports of data security incidents and policy compliance violations should be sent to the DAC(s) indicated in the Addendum to this Agreement.**

NIH, or another entity designated by NIH may, as permitted by law, also investigate any data security incident or policy violation. Approved Users and their associates agree to support such investigations

and provide information, within the limits of applicable local, state, tribal, and federal laws and regulations. In addition, Requester and Approved Users agree to work with the NIH to assure that plans and procedures that are developed to address identified problems are mutually acceptable and consistent with applicable law.

**8. Policy Compliance Violations**

The Requester and Approved Users acknowledge that the NIH may terminate the DAR, including this Agreement and immediately revoke or suspend access to all controlled-access datasets subject to the NIH GDS Policy at any time if the Requester is found to be no longer in agreement with the principles outlined in the NIH GDS Policy, the terms described in this Agreement, or the Genomic Data User Code of Conduct. The Requester and PI agree to notify the NIH of any violations of the NIH GDS Policy, this Agreement, or the Genomic Data User Code of Conduct data within 24 hours of when the incident is identified. Repeated violations or unresponsiveness to NIH requests may result in further compliance measures affecting the Requester.

The Requester and PI agree to notify the appropriate DAC(s) 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 DAC notification(s), the Requester agrees to submit to the DAC(s) 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. The Requester agrees to provide documentation verifying that the remediation plans have been implemented. Repeated violations or unresponsiveness to NIH requests may result in further compliance measures affecting the Requester.

**All notifications and written reports of data management incidents should be sent to the DAC(s) indicated in the Addendum to this Agreement.**

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

**9. 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.

**10. Dissemination of Research Findings and Acknowledgement of Controlled-Access Datasets Subject to the NIH GDS Policy**

It is NIH's intent to promote the dissemination of research findings from use of controlled-access dataset(s) subject to the NIH 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 Submitting Investigator(s) who submitted data from the original study to an NIH-designated data repository, the primary funding organization that supported the Submitting Investigator(s), 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 is provided for each dataset in the Addendum to this Agreement.

**11. Research Use Reporting**

To assure adherence to NIH GDS Policy, the PI agrees to provide annual Progress Updates as part of the annual Project Renewal or Project Close-out processes, prior to the expiration of the one (1) year data access period. The PI who is 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 PI is seeking renewal), any violations of the terms of access described within this Agreement and the implemented remediation, and information on any downstream intellectual property generated from the data. The PI also may include general comments regarding suggestions for improving the data access process 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 stewardship of data sharing activities subject to the NIH GDS Policy.

**12. 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 and Submitting Investigator(s) do not and cannot warrant the results that may be obtained by using any data included therein. NIH 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.

**13. Termination and Data Destruction**

Upon Project Close-out, the Requester and Approved Users agree to destroy all copies, versions, and Data Derivatives 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, in accord with the NIH Security Best Practices for Controlled-Access Data Subject to the NIH Genomic Data Sharing (GDS) Policy. However, the Requester may retain these data as necessary to comply with any institutional policies (e.g., scientific data retention policy), law, and scientific transparency expectations for disseminated research results, and/or journal policies. A Requester who retains data for any of these purposes continues to be a steward of the data and is responsible for the management of the retained data in accordance with the NIH Security Best Practices for Controlled-Access Data Subject to the NIH Genomic Data Sharing (GDS) Policy, and any institutional policies. Any retained data may only be used by the PI and Requester to support the findings (e.g., validation) resulting from the research described in the DAR that was submitted by the Requester and approved by NIH. The data may not be used to answer any additional research questions, even if they are within the scope of the approved Data Access Request, unless the Requester submits a new DAR and is approved by NIH to conduct the additional research. If a Requester retains data for any of these purposes, the relevant portions of Terms 4, 5, 6, 7, 8, and 13 remain in effect after termination of this Data Use Certification Agreement. These terms remain in effect until the data is destroyed.

**14. 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 Certification 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 NIH 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 NIH GDS Policy.

**Cloud Computing:** 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 NIST Special Publication 800-145.

**Cloud Service Provider (CSP):** A company or institution 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 NIST Special Publication 800-145.

**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. Requester Collaborators and project team members on a request must be from the same organization.

**Data Derivative:** Data derived from controlled-access datasets obtained from NIH-designated data repositories. Examples of derived data include imputed datasets and single nucleotide polymorphisms.

**Data Use Certification (DUC) Agreement:** An agreement between the Approved User, the Requester, and NIH regarding the terms associated with access of controlled-access datasets subject to the NIH GDS Policy and the expectations for use of these datasets.

**Genomic Data User Code of Conduct:** Key principles and practices agreed to by all research investigators requesting access to controlled-access data subject to the NIH GDS Policy. The elements within the Genomic Data User Code of Conduct reflect the terms of access in the Data Use Certification Agreement. Failure to abide by the Genomic Code of Conduct may result in revocation of an investigator's access to any and all approved datasets.

**Information Technology (IT) Director:** An Approved User who is generally a senior IT official of the Requester with the necessary expertise and authority to affirm the IT capacities at the Requester. 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 Requester's IT security requirements and policies are followed by all of the Requester's Approved Users.

**Institutional Certification:** Certification by the Submitting 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. Further information may be found here.

**Institutional Signing Official:** The label, "Signing Official," is used in conjunction with the NIH eRA Commons and refers to the individual that has institutional authority to legally bind the institution in grants administration matters. The individual fulfilling this role may have any number of titles in the institution, but is typically located in its Office of Sponsored Research or equivalent. The Signing Official for the Requester reviews Data Access Request, Project Renewal, and Project Close-out applications submitted by Principal Investigators and legally binds the Requester to agree to adhere to the terms described in this Agreement if the application is submitted to NIH. The Institutional Signing Official for the Submitting Institution enters into the Institutional Certification and signs on behalf of the Submitting Investigator(s) who has submitted data.

**Principal Investigator (PI):** The investigator who prepares Data Access Requests (DARs), Project Renewals, and Project close-outs. The Principal Investigator plays a lead role in ensuring that management and use of controlled-access data remains consistent with the terms in the Data Use

Certification Agreement. To be able to submit a DAR, a Principal Investigator must be designated as such by their institution in eRA Commons *and* be a permanent employee of their institution at a level equivalent to a tenure-track professor or senior scientist with responsibilities that most likely include laboratory administration and oversight.

**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 Requester, 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 Authorized Access System.

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

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

**Submitting Institution:** An organization who submitted a genomic dataset to an NIH-designated data repository (e.g., dbGaP).

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

## Study specific DUC addendum

### phs000393 : DeCODE Genetics of Nicotine Dependence

Public Posting of Genomic Summary Results - Allowed.

NIH Data Access Committee (DAC) : Joint Addiction, Aging, and Mental Health DAC

Important Contacts : JAAMH-DAC@list.nih.gov ; URGENTJAAMHDAC@mail.nih.gov; GDS@mail.nih.gov

In the event of a data management incident, within 24 hours, please contact emails above.

IC Specific Access Term :

Three Methods for Accessing deCODE data for Genetics of Nicotine Dependence:
There are three methods for gaining access:
I) Predefined summary data is distributed to dbGAP. These summary data can be subsequently combined with other summary data.
II) Ad-hoc capability for scientists to define new summary statistics through the deCODE secure remote Disease Miner interface. These data can be combined with dbGAP primary data by uploading dbGAP data to the deCODE server to generate new summary statistics.
III) Full access to the project data within the data enclave in Iceland, with the support of deCODE staff. Under this circumstance project data can be analyzed together with dbGAP data by uploading dbGAP data to the deCODE server.
Data Access.

Investigators approved by the dbGAP Joint Addiction, Aging, and Mental Health Data Access committee (JAAMH-DAC) will have immediate access the descriptive and summary data for GWAS results in dbGAP from the previously funded by 1R01DA17932-01A1. Remote access to individual genotype and phenotype data generated by 1R01DA17932-01A1 through a secure password protected web interface (through a secure VPN connection) through the internet will be made accessible as soon as the site is ready but no later than June 30, 2012. Researchers approved by the JAAMH-DAC may also gain immediate access to genotype and phenotype data generated by 1R01DA17932-01A1 through the data enclave model, where interested researchers will be able to analyze data on the deCODE site in Reykjavik, Iceland, starting July 1, 2011.

Process for Data Access

An investigator wishing access to deCODE data shall request access to the deCODE data through NIH dbGAP Joint Addiction, Aging, and Mental Health Data Access committee (JAAMH-DAC). Once granted access by the JAAMH-DAC, investigators may retrieve the summary data directly from dbGAP, or they could go to a stand-alone firewalled deCODE server and request an analysis of the data to create an unidentifiable summary statistic data set to be downloaded from the deCODE server. No individual level data from Icelandic research participants shall be deposited outside deCODE as specified by Icelandic law and all handling of personal data and use of samples shall adhere to Iceland laws. Researchers approved by the JAAMH-DAC will have access to data on the data described above through a secure password protected web interface (through a secure VPN connection) through the internet on servers located at deCODE headquarters in Iceland. The following informatics and statistical algorithms will be made available based on deCODE's Disease Miner Professional package without licensing fees.

• Clinical data – phenotypes: Import, manage, and query large datasets (including the phenotypeSDL query algorithm for defining phenotype datasets).
• Family data: Import data, cluster subfamilies and analyse pedigrees for extended families
• Genetic Analysis: Genome-wide association (Fishers-Exact, TDT, S-TDT, CNV, LD, Clark's inference algorithm
• Genome browser: High speed visualization and browsing with multiple analyses results in one view
• Marker and gene data base: Database of over 30,000 known genes and 9 million SNPs.

This interface will be flexible and allow inclusion of other algorithms upon request that may implemented by deCODE's statistics and informatics departments. The remote Disease Miner web interface does however not provide unrestricted access to primary data during the analysis stage and prevents data inference by applying a dynamic k-anonymization approach. The main principle is to defer modification of the personal primary data until statistical results are delivered to the end-user.

The dynamic k-anonymization approach differs from regular k-anonymization approaches that are used to protect privacy, by preventing data inference, typically bin data values into buckets or perturb then in some other way. This is because these approaches work on the entire data set that is to be released at any given time. Under the dynamic k-anonymization approach scientists can define patient sets (or control sets or inputs for QTL analysis) based on all the variables that are captured for each subject in the Disease Miner database. Behind the scenes (on deCODE's servers), these patient sets will be evaluated based on the actual primary data. However, if a scientist wants for instance to see the distribution of BMI in his patient set, the system ensures that no BMI value (range) is returned unless it applies to at least k-individuals. (The size and cardinality of the data set will be taken into consideration when determining the size of k, but it will be no larger than 10,) Similarly, when the deCODE servers calculate summary statistics such as odds-ratios (per marker or haplotype) or perform regression in QTL, deCODE servers use the actual primary data, while the actual data table for the primary data (or a subset of it) is never returned. The main benefit of this approach is that it applies to any data sets, regardless of their cardinality,

while the conventional approaches would limit the number of data attributes one can capture or work with per subject.

Thus, investigators given remote access will not be able to modify the algorithms used to analyze the data. Nor will the user be able to see any of the typical primary data tables. The investigator will however have an easy way of exploring the entire data set, e.g. define patient sets and analyze the distribution of one or more variables within these sets. This should provide scientists with an easy way, and full capacity, to define the statistical analysis remotely without any support or intervention from deCODE staff, which results files they will be able to download or view in the deCODE built-in genome browser.

Full access to primary data (with all its flexibility to modify algorithms) is only possible within the data enclave in Iceland.

Researchers approved by the JAAMH-DAC may also access data through the data enclave model, where interested researchers will be able to analyze data on site in Iceland. This Data Enclave would be a secure database associated with power processors but with no connections to the outside, therefore requiring access locally rather than remotely. Office space and efficient informatics assistance would be available to researchers visiting Iceland. Certain types of analysis may not be possible through remote access in a timely fashion, or would in some case require modification of software that need to be approved by the Data Protection Authority and the Scientific Ethics Board of Iceland.

Research access to the requested deCODE dataset is granted for a period of one (1) year.

An Investigator wishing to combine dbGAP data with summary deCODE data shall request access to the deCODE data through dbGAP. Once granted access by JAAMH-DAC, investigators may retrieve the summary data from dbGAP directly, or go to the deCODE server directly (see above) and request an analysis of the data to create an unidentifiable summary statistics dataset to be downloaded from the deCODE server to their own servers and combine with dbGAP data and any other data (as is typical of most meta-analyses through large consortia); or

An investigator wishing to combine primary sequence or genotype data from dbGAP with primary data from deCODE shall apply for access to download dbGAP data. At the same time, deCODE will apply for access to the same requested datasets as the investigator so that those data sets can be downloaded directly to the deCODE server without going through the 3rd party (investigator). The analysis will be done on the deCODE server. The dbGAP data will only temporarily reside on deCODE's server, for the limited time to conduct analyses approved by the access committee, and this period will be typically no longer than 3 months unless specifically approved by the access committee. While on DeCODE's data servers, dbGAP data shall be safeguarded to the same extent as the data being shared by deCODE. Summary data will be returned to the investigator.

An Investigator wishing to work and view directly deCODE primary sequence or genotype data generated with the support of R01DA17932 and combine with dbGAP data shall request access to dbGAP. In this case the investigator must travel to deCODE. Once granted access by the JAAMH-DAC, the investigator may travel to deCODE to work inside the deCODE data enclave. At the same time, deCODE will also need to apply for access to the same requested datasets as the investigator so those data sets can be download directly to the deCODE server without going through the 3rd party (investigator). The analysis will be done on the deCODE server. The dbGAP data will only temporarily reside on deCODE servers, for the limited time required to conduct the analyses approved by the access committee, and this period will typically be no longer than 3 months unless specifically approved by the access committee. While on deCODE's servers, dbGAP data would be safeguarded to the same extent as the data being shared by deCODE.

Acknowledgement Statement : "The authors acknowledge the contribution of data from Genes Contributing to Nicotine Dependence in Humans supported by R01DA17932 and accessed through dbGAP to the analysis presented in this communication, i.e. publication, presentation, or grant application."

Name : General Research Use

Consent Group # : 1

Abbreviation : GRU

The DeCODE Genetics of Nicotine Dependence study summary data accessed from dbGAP or summary data generated through remote access to the data from DeCODE Genetics of Nicotine Dependence study on the deCODE server may be used for general research purposes. The consents do not limit the types of studies that can be performed.

Data Use Limitation : In addition, Investigators granted on site access to the deCODE enclave in Iceland may also use the data for general research purposes, i.e. the consent do not limit use, with the following caveats: No individual level data from Icelandic research participants shall be deposited outside deCODE as specified by Icelandic law AND All handling of personal data and use of personal data shall adhere to Icelandic laws, including but not limited to the The Act on Patients rights No. 74/1997 and Government Regulation No. 286/2008 on Medical Research, The Data Protection Act No. 77/2000 based on the European Data Protective Directive, and various rules and government regulations promulgated there under the Minister of Justice and the Icelandic Data Protection authority. These include "Rules on the Obligation to Notify and Processing which Requires a Permit" No. 698/2004, and the The Act on biobanks, No. 110/2000 and Government Regulation No. 134/2001. For English translations contact the NIH dbGAP Joint Addiction, Aging, and Mental Health Data Access committee.