# Genetics of cardiovascular risk factors in a large founder population birth cohort-The Northern Finland Birth Cohort 1966 study (NFBC1966)
# Data Use Certification

## Introduction and Statement of Policy

The NIH GWAS Data Repository was developed to archive and distribute the results of studies provided by Contributing Investigators examining the relationship between genotype and phenotype. Such studies include genome-wide association studies, medical sequencing, and molecular diagnostic assays. Implicit in dbGaP is the view that scientific progress in this area will be greatly enhanced if the data produced by these studies are readily available to all investigators in the research community.

Dataset access will only 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 the NIH and NHLBI that Approved Users of GWAS datasets recognize any restrictions on data use delineated within the original informed consent agreements of contributing studies, as identified by the submitting institutions and stated on dbGaP.

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

The parties between whom this agreement is made include: the Principal Investigator (PI) requesting access to the GWAS dataset, his/her home institution as represented by the Institutional Signing Official designated through the eRA Commons system, and the NHLBI, National Institutes of Health. The effective date of this agreement shall be the Project Approval Date, as specified on the Data Access Committee approval notification.

## Study Background

The Northern Finland Birth Cohorts program (NFBC) was initiated in the 1960s in the two northernmost provinces of Finland to study risk factors involved in pre-term birth and intrauterine growth retardation, and the consequences of these early adverse events on subsequent morbidity and mortality. To our knowledge, the NBFCs are the only population cohorts in which data were obtained from early fetal life (including maternal health during pregnancy) to adulthood. The NFBC1966 includes 12,058 live births to mothers in the two northern-most provinces of Finland. Two decades later, a second cohort of 9432 births was obtained (NFBC1986). In NFBC1966 pregnancies were followed prospectively from the first antenatal contact (10-16th week). After birth, the offspring were examined and then again underwent clinical evaluation at ages 1y, 7y, 14-16y and 31y. At each visit, a wide range of phenotypic, lifestyle and demographic data were gathered by questionnaires and clinical examinations. For the most part, NFBC1986 has undergone similar evaluations to NFBC1966. Linkage to national registries includes hospitalization, deaths, education, medication, pensions, and provides up-to-date demographic and clinical information for members of both cohorts. DNA samples were obtained from 5,923 subjects from NFBC1966 and 6688 subjects from NFBC1986. Data coverage, 96% of all births in 1966 and 99% in 1986, is highly representative for the whole population. The NFBC program comprises more than 20 different projects coordinated by the Center of Lifecourse Disease studies in Northern Finland (COLD) at Oulu University. The prospective data collected from the NFBCs form a unique resource, allowing the study of disease emergence, and of the importance of genetic, biological, social and behavioral risk factors.

The genome-wide association (GWA) study sponsored through the STAMPEED program of NHLBI employed genomic DNA samples previously collected by the NFBC1966 study and

stored in the DNA repository of the National Institute for Health and Welfare, Finland. This NHLBI sponsored RO1 project aimed to identify genetic variants contributing to metabolic and cardiovascular diseases (CVD). In addition to de-identified genome wide genotypic data, a selected list of phenotypic data related to CVD including weight, height, BMI, HDL, LDL, total cholesterol, triglyceride, glucose, insulin and fasting status, are also available in dbGaP. A summary of the GWAS for the NFBC1966 cardiovascular risk traits can be found in Sabatti et al., Nature Genetics 41: 35-46, 2009

## Terms of Access

### 1. Research Use

The Requester agrees that if access is approved, the Principal Investigator named in the Data Access Request (DAR) submitted to the NIH, those named in the "Senior/Key Person Profile" portion of the DAR, which should include the Information Technology Director or his/her designee, and any trainee or employee working on the proposed research project under the direct supervision of these individuals, shall become Approved Users of the requested dataset(s). Research use will occur solely in connection with the 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., the addition of new aims related to the approved project, adding or deleting collaborators from the same institution, and the potential addition of new NIH GWAS datasets to an approved project). The Requester and all Approved Users further agree to use the dataset(s) only in accordance with the parameters described on the dbGaP Web site for the appropriate research use, and any limitations on such use, of the dataset(s) and as required by law.

Research access to the requested dataset(s) is granted for a period of one (1) year as defined below.

Contributing GWAS Investigators, or their direct collaborators, who provided the data or samples used to generate an NIH GWAS dataset and who have appropriate IRB approval, if applicable, for broader use of the data are exempt from the limitation on the scope of the research use as defined in the DAR.

**DATA USE LIMITATIONS FOR NFBC1966: There are no restrictions in the usage of the genomic results outside the study for which they were originally consented.**

### 2. Institutional and Approved User Responsibilities

The Requester agrees through the submission of the Data Access Request that the PI has reviewed and understands the principles for responsible research use and data handling of the GWAS datasets as defined in the NIH GWAS Data Sharing Policy (http://grants.nih.gov/grants/guide/notice-files/NOT-OD-07-088.html) and detailed in this Data Use Certification 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. Through submission of the DAR, the Principal Investigator also agrees to submit annual reports to the appropriate NIH Data Access Committee (DAC) describing the research use of the Approved Users as described under *"Research Use Reporting"* below.

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 for their research protocol. By approving and submitting the attached Data Access Request, the Institutional Signing Official provides assurance that relevant institutional policies and applicable federal, state, or 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 Data Access Request that other organizations within the institution with relevant authorities (e.g., the Office of Human Subjects Research, the Office of Information Technology, the Office of Technology Transfer, etc.) have reviewed the relevant sections of the NIH GWAS Data Sharing Policy and the associated procedures and are in agreement with the principles defined.

It is anticipated that, at least in some cases, these datasets 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 User's Research Use Statement

The Principal Investigator agrees that, if he or she becomes an Approved User, information about the proposed research use can be posted on a public web site that describes the GWAS research projects. The information may include the Approved User's name and institution, project name, Research Use Statement, and a Non-technical Summary of the Research Use Statement. In addition, citations resulting from the use of NIH GWAS datasets will be posted on the NIH GWAS Data Repository 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 phenotype data and DNA samples were collected. This agreement does not apply to research investigators operating with specific IRB approval to contact individuals within datasets under an approved IRB research protocol.

### 5. Non-Transferability

The Requester and Approved Users agree to retain control over the data and further agree not to distribute individual-level data in any form to any entity or individual not covered in the submitted Data Access Request under "Senior/Key Person Profile". If Approved Users are provided access to NIH GWAS datasets for inter-institutional collaborative research described in the research use statement of the Data Access Request, and all members of the collaboration are also Approved Users, derivatives of the dataset including individual-level data may be securely transmitted within the collaborative group for the purpose of quality control or other research requirements. All data security practices and other terms of use defined in this agreement and the *dbGaP Security Best Practices* (http://www.ncbi.nlm.nih.gov/projects/gap/pdf/dbgap_2b_security_procedures.pdf) for the original data are expected to be followed for the derived data, including any transmission of the data.

The Requester and Approved Users acknowledge responsibility for ensuring the review and agreement to the terms within this Data Use Certification and the appropriate research use of NIH GWAS data by research staff associated with any approved GWAS project, subject to applicable laws and regulations. NIH GWAS Datasets containing individual-level information, 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 submit a new Data Access Request and Data Use Certification in which the new institution will agree to the NIH GWAS data use policy before data use resumes. Any versions of data stored at the prior institution will be destroyed and documented through a final Data Use Report as described below. However, if advance written notice and approval by the NHLBI Data Access Committee is obtained to transfer responsibility for the approved research project to another Approved User within the same institution the data may not need to be destroyed.

## 6. Data Security and Data Release Reporting

The Requester and Approved Users, including the Information Technology Director or his/her designee acknowledge that they have reviewed and agree to handle the requested dataset(s) according to the current *dbGaP Security Best Practices* (http://www.ncbi.nlm.nih.gov/projects/gap/pdf/dbgap_2b_security_procedures.pdf), including its detailed description of requirements for security and encryption. This includes, but is not limited to:

- all approved users have completed all required institutional computer security training, e.g. 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, (i.e. must be behind a firewall or not connected to a larger network) and unnecessary services disabled;
- use of portable media, e.g., on a CD, flash drive or laptop, is discouraged, if necessary then they must be encrypted;
- use of updated anti-virus/anti-spyware software;
- security auditing/intrusion detection software, detection and regular scans of potential data intrusions;
- use of strong password policies for file access.
- all copies of the dataset should be destroyed whenever any of the following occurs:
  - the DUC expires
  - the NHLBI requests destruction of the dataset
  - the 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 ensure that only authorized individuals can gain access to NIH GWAS datasets. This agreement includes the maintenance of appropriate controls over any copies or derivatives of the data that include individual-level data.

Requesters and Approved Users agree to notify the NHLBI Data Access Committee 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. Notifications should include the known information regarding the incident and a general description of the activities or process in place to fully define and remediate the situation. Within 3 business days of the NIH notification, the Requester, through the Approved User and the Institutional Signing Official, agree to submit to the NHLBI DAC 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:**

NHLBI Data Access Committee Chair
Email: nhlbigeneticdata@mail.nih.gov

The NHLBI, the NIH, or another entity designated by the NIH may 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 NHLBI and the NIH to assure that plans and procedures developed to address identified problems are mutually acceptable.

## 7. Intellectual Property

By requesting access to GWAS dataset(s), the Requester and Approved Users acknowledge the intent of the NIH to see that anyone authorized for research access through the attached

Data Access Request follow the intellectual property principles within the NIH GWAS Policy for Data Sharing (http://grants.nih.gov/grants/guide/notice-files/NOT-OD-07-088.html#intellectual) as summarized below:

> Achieving maximum public benefit is the ultimate goal of data distribution through the NIH GWAS Data Repository. The NIH believes that these data should be considered as pre-competitive, and urges Approved Users to avoid making IP claims derived directly from the GWAS dataset(s). 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.
>
> It is expected that these NIH-provided data, and conclusions derived therefrom, will remain freely available, without requirement for licensing. The NIH encourages broad use of GWAS datasets coupled with a responsible approach to management of intellectual property derived from downstream discoveries in a manner consistent with the NIH's Best Practices for the Licensing of Genomic Inventions [hotlink www.ott.nih.gov/policy/genomic_invention.html] and the NIH Research Tools Policy [ grants.nih.gov/grants/intell-property_64FR72090.pdf].

## 8. Research Dissemination and Acknowledgement of NIH GWAS Datasets

It is the intent of the NIH to promote the dissemination of research findings from NIH GWAS 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, etc. In accord with the NIH GWAS Policy for Data Sharing (http://grants.nih.gov/grants/guide/notice-files/NOT-OD-07-088.html#scientific), and as expressed through the submission of the DAR, Approved Users agree not to submit findings using the NFBC1966 dataset(s), or updated versions thereof, for publication or presentation for a period of exclusivity for Contributing Investigators concluding with the Embargo Date identified on the NIH GWAS Data Repository.

Approved Users agree to acknowledge the NIH GWAS Data Repository, the Contributing Investigator(s) who contributed the phenotype data and DNA samples from his/her original study, and the primary funding organization that supported the contributing study in all oral and written presentations, disclosures, and publications resulting from any analyses of the data. Approved Users further agree that the acknowledgment shall include the dbGaP accession number to the specific version of the dataset(s) analyzed.

A sample statement for the acknowledgment of the NFBC1966 dataset(s) follows:

The Approved Users agree to acknowledge the contribution of NHLBI and NFBC1966 Study Investigators in any and all oral and written presentations, disclosures, and publications resulting from any and all analyses of such genetic analysis data.

If the Research Project involves collaboration with NFBC1966 Study Investigators, then the Approved Users will acknowledge NFBC1966 Study as co-authors, as appropriate, on any publication.

If the Research Project does not involved collaboration with NFBC1966 Study Investigators, then the manuscript shall include the acknowledgement below.

"The NFBC1966 Study is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with the Broad Institute, UCLA, University of Oulu, and the National Institute for Health and Welfare in Finland. This manuscript was not prepared in collaboration with investigators of the NFBC1966 Study and does not necessarily reflect the opinions or views of the NFBC1966 Study Investigators, Broad Institute, UCLA, University of Oulu, National Institute for Health and Welfare in Finland and the NHLBI.

## 9. Research Use Reporting

To ensure that NIH policies and procedures for participant protection and other elements of GWAS data use are adhered to, Approved Users agree to provide annual feedback on how these data have been used and any results that have been generated as a result of access to the data, including patents and publications. This information will be used by the NHLBI DAC staff as required for program oversight activities and by the NIH GWAS Governance committees as part of the NIH effort to provide ongoing oversight and management of all NIH GWAS activities.

Approved Users thus agree to provide a brief Annual Data Use Report on the research specified within the DAR submitted with this DUC. Approved Users who are seeking renewal agree to provide specific information in a renewal DAR. Those not seeking renewal agree to provide specific information to the Data Access Committee via the contact information below. Annual Data Use Reports will provide information regarding potentially significant findings and 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 Data Use Certification and the implemented remediation, and information on any downstream intellectual property generated as a result of the data. Approved Users also may be asked to provide general comments regarding topics such as the effectiveness of the NIH GWAS 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.

Annual Data Use Reports will be submitted on the first day of the month following the anniversary of the Approved Access Date assigned by the DAC and specified within the manifest file provided to Approved Users by the NIH GWAS Data Repository at the time that that data access is provided. The final Data Use Report within a given approval period will be submitted within the context of a renewal Data Access Request, or via a letter signed by the Institutional Signing Official and the Approved User.

**Annual Data Use Reports should be submitted to:**

NHLBI Data Access Committee Chair
Email: nhlbigeneticdata@mail.nih.gov

**Note that any inadvertent or inappropriate data release incidents should be reported to the NHLBI 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 NIH GWAS data, the NIH, the NHLBI, and Contributing GWAS Investigators do not and cannot warrant the results that may be obtained by using any data included therein. The NIH, the NHLBI 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 the NIH, as an agency of the United States, assumes liability only to the extent provided under the Federal Tort Claims Act, 28 U.S.C. 2671 et seq.

## 11. Termination and Violations

This Data Use Certification will be in effect for a period of 3 years from the date the data were made accessible to the Approved User ("Approved Access Date") on the condition of receiving the annual reports. At the end of the access period, Approved Users agree to destroy all

copies of the requested dataset(s), and all derivatives that contain individual-level information.

Consideration will be given to a renewal of this agreement upon submission of a new DAR. Copies of NIH GWAS dataset(s) may not need to be destroyed if, with advance notice and approval by the NHLBI Data Access Committee, documentation is provided that other Approved Users at the same institution are using the dataset(s) under an active DAC approved research project.

The NIH or the NHLBI may terminate this agreement and immediately revoke access to all NIH GWAS 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 NHLBI.

**********

By submission of the attached Data Access Request, the Requester through the Institutional Signing Official attest to the Approved Users' qualifications for access to and use of NIH GWAS dataset(s) and certify their agreement to the NIH principles, policies and procedures for the use of the requested datasets as articulated in this document.

Requesters and the Principal Investigator further acknowledge that they have shared this document and the NIH GWAS policies and procedures for access and use of GWAS datasets with any Approved Users, appropriate research staff, and all other Key Personnel 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 organizations, and have assured compliance with local institutional policies related to technology transfer, information technology, privacy, and human subjects research.

# Appendix

## Terms

**Requester:** The home institution/organization for the PI that will use the requested data
**Approved User:** Post-DAC approval will include the PI, Collaborators at the home institution, trainees to these investigators, and the IT Director or designee. All individuals named in the Senior/Key Persons portion of the DAR
**Institutional Signing Official:** Someone with the authority to sign on behalf of the Requester and credentialed through the eRA system as such
**Senior/Key Persons:** Collaborators at the home institution, trainees to these investigators, and the IT Director or designee.
**Contributing GWAS Investigator:** PI that submitted the GWAS dataset to dbGaP
**Data Access Request:** SF 424 (R&R) cover pages and requested attachments
**GWAS Dataset Derivatives:** any dataset including individual-level data that stems from the original dataset obtained through dbGaP.
**Annual Data Use Report:** submitted to the DAC on anniversary of access approval
**Final Data Use Report:** submitted to the DAC at the end of the approved access period when no additional access is being sought
**Information Technology Director:** someone with the authority to vouch for the IT capacities at an institution, or higher-level division of an institution (e.g., the School of Medicine)

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