



Standard Operating Procedure (SOP) for Deposition of Genomic Data from Non-Commercial Biospecimen Sources into Open Access Databases

Last Revised: August 30th 2023

SOP OVERVIEW

Regarding deposition of data generated from non-commercial sample sources into an open-access repository, the Broad Institute voluntarily adheres to the National Institutes of Health's Genomic Data Sharing Policy of 2015, regardless of the funding source used in data generation. For samples collected after the effective date of the policy (January 25, 2015), consent forms associated with sample collection must explicitly reference future data deposition in an open access repository, and deposition must be approved by an Institutional Review Board (IRB). For samples collected before the effective date of the policy, open access deposition is permitted at the discretion of the reviewing IRB.

PURPOSE AND SCOPE

This SOP describes the type of genomic research data generated from non-commercially sourced human biospecimens and cell lines that may be deposited into an open access repository, and provides detailed instructions on the steps necessary to deposit such genomic research data into such an open access repository. The SOP applies to individual level whole genome and whole exome data generated from samples that are not commercially available (e.g. those that are provided by an academic collaborator).

The following types of data do NOT fall within the scope of this SOP:

- gene expression calls
- somatic variant calls
- non-genetic read counts

Other types of data (e.g. SNP array and fingerprinting data) may also be subject to this SOP at the discretion of Broad scientific leadership and/or the NIH.

This SOP applies to the data described above regardless of the regulatory determination applied to the project. This means that sequencing projects that were determined to be “not human subjects research” or those in which the Broad was determined to be “not engaged in human subjects research,” in addition to those that were reviewed by an IRB, are subject to this SOP.

This SOP does not apply to data that has already been released into the public domain (e.g. the 1000 Genomes Project).

DEFINITIONS AND ACRONYMS

Code of Federal Regulations (45 CFR 46): HHS regulations that govern human subject research in the United States (also referred to as the “Common Rule,” as it was adopted by 15 federal agencies).

Commercially available biospecimens or cell lines: biospecimens or cell lines that are available for purchase from a vendor for research.

Consent Form: A form signed by a research participant indicating they are voluntary participants in the research protocol.

Controlled-access data repository: a data repository that requires review by a Data Access Committee (DAC) to approve deposition of and access to data.

Data Access Committee (DAC): A committee charged with reviewing data access requests to ensure their congruence with any applicable restrictions on data access or use.

De-Identified Data: Data that has been stripped of the 18 HIPAA identifiers as listed in 45CFR164.514.

Database of Genotypes and Phenotypes (dbGaP): a controlled access data repository maintained by the National Institutes of Health to archive, curate and distribute information produced by studies investigating the interaction of genotype and phenotype.

Genomic Research Data: Genomic data that include sequence information from the genome or the transcriptome.

Gene Expression Omnibus (GEO): an open-access repository maintained by the National Institutes of Health to serve genomic research data from diverse technologies, including microarrays and RNA sequencing,

Gene Expression Omnibus (GEO) Deposition Requirements: the requirements for deposition into GEO include explicit language in the consent that states the data will be available in an open-access repository, if the biospecimens were collected after January 25, 2015.

Genomic Data Sharing (GDS) Policy (2015): a policy issued by the National Institutes of Health, effective January 25, 2015, that sets forth expectations designed to ensure the broad and responsible sharing of genomic research data.

Health Insurance Portability and Accountability Act (HIPAA): a US law that defines privacy standards to protect patients' medical records and other health information obtained by health plans, doctors, hospitals and other health care providers.

Institutional Review Board (IRB): an administrative body established to protect the rights and welfare of human research subjects.

Memo of Determination (MOD): A memo written by the Director of the Office of Research Subject Protection after review of the consent form and in template format (or substantially similar format) provided in Attachment A.

Not Engaged in Human Subjects Research: A designation that applies to research in which human subjects research is taking place at a collaborating institution (typically where samples are collected) but Broad is not engaged in that research according to the 2008 OHRP “Guidance on Engagement of Institutions in Human Subjects Research”). There are several possible subcategories of this designation (e.g. fee-for-service), which is issued by Broad’s Office of Research Subject Protection.

Not Human Subject Research (NHSR): a designation that applies to research that is conducted in samples derived from human materials but does not involve a human subject as defined by HHS. This designation is made by the Broad's Office of Research Subject Protection based on the information provided by the study team regarding the research project.

Exempt Research: Research involving human subjects that meet one of eight criteria outlined in the Code of Federal Regulations, and which are therefore exempt from some federal regulations that apply to non-exempt research. Research exemptions for Broad- based projects must be made by the Broad's Office of Research Subject Protection or the project's IRB-of-record.

Office of Research Subject Protection (ORSP): The Broad Institute office responsible for compliance with all local, state, and federal laws, regulations and guidance related to the use of humans and human materials in research. ORSP reports to the Broad's Chief Compliance Officer.

Open-Access Repository: A data repository that is accessible to anyone regardless of institutional affiliation or professional status and that operates without review by a Data Access Committee.

ORSP Portal: an online platform where Broad staff members can upload documents for ORSP review (either as stand-alone submissions, or in preparation for an IRB protocol applications), and store consent forms and regulatory approvals.

DESCRIPTION

The process for making a determination whether genomic research data is eligible for deposition into an open access database consists of the following steps:

1. The Broad PI obtains a blank copy of the consent form that was used when the biospecimens were collected. If the samples were collected under an IRB-approved waiver of consent, documentation of this determination is required.
2. To be eligible to initiate a request the relevant study must have undergone appropriate regulatory review (IRB review of human subjects research, or ORSP review of "not human subjects research" or research in which the Broad is "not engaged.")
3. The regulatory determination is entered into the ORSP portal by ORSP staff.
4. For data generated from samples collected after January 25, 2015, deposition in an open-access repository, all of the following are required:
 - i) the consent form explicitly describes open access sharing

- ii) the consent form describes future research use of the sample (either broad general research use or genomic research);
 - iii) the consent form does not include language that implies limited access to the sample or data;
5. For data generated from samples collected before January 25, 2015, the consent form must not explicitly preclude broad sharing through an open-access repository. NOTE: If sample collection pre-dates January 25, 2015, and no consent was obtained, or the consent form is not available, data may still be eligible for open access deposition at the discretion of the IRB that originally approved sample collection.
 6. Provided that the criteria in either #4 or #5 are met, the request to share data in an open access repository must then be submitted to and approved by an IRB. Typically this approval must be provided by the IRB that originally approved the biospecimen collection consent form, although in some cases a Broad IRB-of-record may be substituted. Written approval is required; this may be either in the form of an amendment to the collection protocol, or in a letter or email from the IRB at their discretion.
 7. If the deposition is approved by the IRB, The Director of the Office of Research Subject protection completes a Memo of Determination (MOD) in a format similar to that provided in Appendix A.
 8. The MOD is reviewed by the Director of ORSP. If the Director of ORSP rejects the MOD, then the genomic research data cannot be deposited in an open-access repository.
 9. If the Director of ORSP accepts and signs the MOD, the Broad Investigator receives an approval letter from the IRB stating that the data may be deposited in an open access database.
 10. The MOD and any related documentation is stored in the ORSP portal.

REFERENCES

- Code of Federal Regulations 45CFR46
- Code of Federal Regulations 45CFR164.514

CONTACTS AND SUBJECT MATTER EXPERTS

- Office of Research Subject Protection
- Director, Office of Research Subject Protection

KEYWORDS

Human subjects data, data deposition, GEO, open access databases



APPENDIX A: MEMO OF DETERMINATION: OPEN ACCESS DISTRIBUTION OF DATA
DERIVED FROM BIOSPECIMENS THAT ARE NOT COMMERCIALY AVAILABLE

Specimen Source and Data Destination Information

Source from which Biospecimens were acquired:
Title on Informed Consent Form:
Repository in which Data will be Deposited:
Date range of sample collection:

Prerequisites

___ If samples were collected after January 25, 2015, ORSP has reviewed the consent form(s) and confirmed the following:

- ___ the consent includes a description of potential future use in any type of research (general research use), or specifically references genomic research;
- ___ the consent explicitly describes open-access data sharing
- ___ the consent does not include language that implies limited access to the data

___ If samples were collected before January 25, 2015 and an consent form is available, ORSP has reviewed the consent and confirmed that it does not explicitly prohibit data sharing in an open-access repository.

___ If samples were collected before January 25, 2015 and a consent form is not available, the IRB of the collecting institution has reviewed and approved data deposition in an open-access repository.

___ The consent and project description have been entered into the ORSP portal as a part of a request for a “not human subjects research” determination, or has been reviewed by a Broad IRB-of-record.
ORSP ID # _____

___ An IRB has approved data deposition in an open-access repository

Determination

___ Data derived from this commercial biospecimen source, from samples collected under the consent form cited above, MAY ___/ MAY NOT ___ be deposited into an open access repository

Signature

Director, Office of Research Subject Protection/Date: