BROAD INSTITUTE

Standard Operating Procedure (SOP) for Deposition of Genomic Data from Commercially Available Biospecimens into Open Access Databases

Last Revised: August 30th 2023

SOP OVERVIEW

Use of commercially available human biospecimens and cell lines is not typically considered to be research involving human subjects. In particular, it does not meet the federal definition of "work involving a living individual about whom an investigator conducting research: (i) obtains information or biospecimens through intervention or interaction with the individual, and uses, studies, or analyzes the information or biospecimens; or (ii) obtains, uses, studies, analyzes, or generates identifiable private information or identifiable biospecimens." Most commercially available biospecimens are fully de-identified, are available for any research purpose, and can often be publicly purchased through online vendors. As a result, most projects that use these lines and specimens do not undergo IRB review but are instead determined to be "not human subjects research" (NHSR) after review by the Broad's Office of Research Subject Protection (ORSP). Because the biospecimens are commercially available, the data resulting from their use in research, including genomic research data, is also eligible for public distribution, provided that the requirements set in this SOP are met and the steps outlined are followed.

PURPOSE AND SCOPE

This SOP describes the type of genomic research data generated from commercially available 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 biospecimens that are commercially available for purchase by any individual, regardless of institutional or professional affiliation.

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 (SNP array and fingerprinting data), depending on size, may also be subject to this SOP at the discretion of Broad scientific leadership and/or the NIH.

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 Office of Research Subject Protection staff after review of the consent form and in template format (or substantially similar format) provided in Attachment A.

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.

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. To be eligible to initiate a request the relevant study must meet the 45CFR46 federal regulation of "not human subjects research" and the specimens must be commercially available biospecimens or cell lines.

2. The Broad PI obtains from the source of the commercially available samples a blank copy of the consent form that was used when the biospecimens were collected. NOTE: If

consent was not obtained, or the consent form is not available to Broad for review, AND the samples were collected before January 25, 2015 (the effective date of the NIH genomic data sharing policy), data may still be eligible for open-access deposition at the discretion of the IRB. If samples were collected after January 25, 2015, and the consent form is not available to Broad for review, open-access deposition of the data is at the discretion of the commercial vendor and the IRB, in consultation with the NIH. Samples collected after January 25, 2015, without any consent, are not eligible for open-access deposition.

3. The consent form and project description are entered into the ORSP portal as part of an official request for a "not human subjects research" determination by ORSP.

4. The consent form is reviewed by ORSP. For genomic research data to be eligible for deposition in an open-access repository all of the following requirements must be met:

    i) the consent form includes a description of potential use in any type of research;

    ii) the consent form does not explicitly prohibit open-access sharing or, if deposition into GEO, all GEO deposition requirements are met (i.e. explicit description of open-access sharing);

    iii) the consent form describes future research use of the sample;

    iv) the consent form does not include language that implies limited access to the sample or data;

5. Provided that the criteria above are met, the request to share data in an open access repository must then be submitted to and approved by one of the Broad's IRBs-of-record, including, but not limited to the IRB that originally approved the biospecimen collection consent form.

6. If the deposition is approved by the IRB, the Office of Research Subject protection staff will complete a Memo of Determination (MOD) in a format similar to that provided in Attachment A.

7. The MOD is reviewed by the Director of ORSP.

8. 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

BROAD INSTITUTE

APPENDIX A: MEMO OF DETERMINATION:
DISTRIBUTION OF DATA DERIVED FROM COMMERCIALLY AVAILABLE BIOSPECIMENS

Specimen Source and Data Destination Information

Commercial Vendor from which Biospecimens were Purchased:
Title on Informed Consent Form (if available):
Repository in which Data will be Deposited:
Date range of sample collection:

Prerequisites

____Proof of commercially availability of biospecimens has been documented to the Broad's Office of Research Subject Protection (ORSP) (e.g. vendor's internet URL, purchase receipt, etc)

____If samples were collected after January 25, 2015:

____A copy of the consent form used in the original sample collection has been obtained and reviewed by ORSP

___ORSP review of the sample collection consent form confirms the following:

___the consent includes a description of potential future use in any type of research (general research use);
___ the consent explicitly describes open-access sharing
___the consent does not include language that implies limited access to the samples or data (i.e. broad sharing is described).

____If samples were collected before January 25, 2015 and a 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, or a Broad IRB of record, has reviewed and approved data deposition in an open-access repository.

___ The consent (if available) 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.

___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 (if available), MAY ___/ MAY NOT ___be deposited into an open-access repository

Signature

Director, Office of Research Subject Protection/Date: