| | |
|---|---|
| **SOP#: RPS-23** | **Registering a Clinical Trial in dbGaP** |
| **Version #: 2.0** | **Next Review Date: 8/2019** |
| **Approved Date: 8/2017** | **Review Interval Period: Biennial** |
| **NCI Clinical Director Signature:** | William L. Dahut -S<br>Digitally signed by William L. Dahut -S<br>DN: c=US, o=U.S. Government, ou=HHS, ou=NIH, ou=People, 0.9.2342.19200300.100.1.1=0010106299, cn=William L. Dahut -S<br>Date: 2017.09.01 10:40:52 -04'00' |

**POLICY**

The Genomic Data Sharing (GDS) Policy applies to all NIH intramural research that generates large-scale human or non-human genomic data as well as the use of these data for subsequent research. Large-scale data include genome-wide association studies (GWAS), single nucleotide polymorphisms (SNP) arrays, and genome sequence, transcriptomic, metagenomic, epigenomic, and gene expression data.

Investigators should work with their designated Genomic Program Administrator (GPA) to register all studies with human genomic data that fall within the scope of the GDS Policy, in the database of Genotypes and Phenotypes (dbGaP portal).

**For clinical trials, the study registration process will start once data for the project have been generated, processed and checked.

**PURPOSE**

The purpose of this standard operating procedure is to provide instructions on registering a Clinical Trial in dbGaP.

**RESOURCES**

- NIH Genomic Data Sharing Policy
- NIH Genomic Data Sharing Website
- CCR Genomic Data Sharing Website

**Genomic Program Administrators (GPA)**
Kathleen Calzone, PhD, RN, AGN-BC, FAAN
240-760-6178, calzonek@mail.nih.gov

**GPA Assistant**
Anjan Purkayastha, PhD
240-760-7408, anjan.purkayastha@nih.gov

## PROCEDURES

### STEP 1: Study Registration Process

- Once data for your project have been generated, processed and checked, contact GPA via email informing them that you would like to register your study with dbGaP. Attach an electronic copy of the signed Institutional Certification (IC) Memo.
  - **NOTE:** You should register your study even if you have applied for and have been granted a data sharing exception.
- GPA will send you:
  - dbGAP study configuration template and
  - dbGAP Basic Study Information Sheet. Example templates of both are included in Appendices A and B.
- Complete the required information and email the following to the GPA:
  - Completed dbGaP study configuration information and
  - Completed dbGaP Basic Study Information Sheet
- GPA will:
  - Ensure that forms are completely and accurately filled out
  - Register the study in the dbGaP portal
  - Invite PI to review the registered study and determine who will be the data submitter – PI or designee as assigned by the PI

### STEP 2: Confirmation of Accuracy

- After the registration process is complete, an email invitation from dbGaP will be sent to the PI to review the study information.
- PI reviews the information for completeness and accuracy. For any questions or concerns on any of aspect of the registered study, or if there are any errors in the registration data, please email your GPA admin for help.

# Study Configuration Template

See example: http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000001.v1.p1

| **Entrez Study Name (character limit is 75 with spaces): a short study name that will appear in Entrez. The short Study Name should be the same between study versions. http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap. Example: NEI Age-Related Eye Disease Study (AREDS)** |
|---|
| |

| **Study Report Name (no character limit): a comprehensive study webpage name that will appear on the upper left hand corner of the study webpage. This name length can be longer than the Entrez Study Name. This name can change between different study versions, since each study version will have a different webpage. http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000001.v1.p1 Example: National Eye Institute (NEI) Age-Related Eye Disease Study (AREDS)** |
|---|
| |

| **Study Type: the study type(s) (Longitudinal, Case-Control, Case Set, Control Set, Parent-Offspring Trios, Cohort, etc).** |
|---|
| |

**Molecular Data: the type, name and version of genotype platform, sequencing platform, arrays and molecular phenotype tools used in this study.**
**Column1 Type**
**Column2 Name and Version (include Version number where applicable)**
**Column3 Vendor**
**Column4 dbSNP Batch ID (if applicable)**
**Column5 Comments**

| Type | Name and Version | Vendor | dbSNP Batch ID | Comments |
|---|---|---|---|---|
| | | | | |
| | | | | |
| | | | | |

| **Study URL: the study URL(s) if applicable.** |
|---|
| |

| **Study Logo: the URL(s) to link the logo(s) (required if logo(s) are submitted).** |
|---|
| |

| **Description: an original summary description of the study. If the description is taken** |
|---|

| **verbatim from a published or soon to be published article, please submit copyright permission from the Journal. Summaries with copyrighted material must include the following within the description: "Reprinted from [Article Citation], with permission from [Publisher]."** |
| --- |
| |

| **Inclusion/Exclusion Criteria: the inclusion and exclusion criteria for cases, controls, trios, and study participants as applicable.** |
| --- |
| |

| **History: the study history as applicable.** |
| --- |

| |
|---|
| |

| |
|---|
| **Phenotype term(s)/Disease name(s): any number of phenotype term(s) and/or disease name(s) associated with this study. The phenotype term and disease name must be a MeSH term. To check, type query in the search box at http://www.ncbi.nlm.nih.gov/mesh/. Disease names will be ordered as submitted below. Please mark one MeSH term as the primary phenotype term with a star (*).** |
| |

| |
|---|
| **Gene(s): any number of gene(s) associated with this study. For gene symbols, please type query in the search box at http://www.ncbi.nlm.nih.gov/gene. Gene symbols will be ordered as submitted below.** |
| |

**Relevant Publications: use Pubmed IDs (http://www.ncbi.nlm.nih.gov/PubMed/). References will appear in the order submitted.**

***Article Type*: References provided should be assigned 1=Study article (GWAS), 2=Follow-up (GWAS, linkage, candidate gene, etc), 3=Non-GWAS, Phenotype, 4=Background, 5=Sequencing (tumor, whole-genome, short-read, somatic, germline, exome, etc.), 6=Gene expression profiling, 10=Other. An article can be one or more types.**

***Articles with PMID or No PMID***
**1. If the article has a Pubmed ID (PMID), please provide the PMID# only.**

**2. If the article, abstract, or book does not have a PMID, please provide the reference in the following formats below.**
**For Journals:**
**Line1 Authors**
**Line2 Title of Article or Abstract**
**Line3 Journal Name**
**Line4 Year;Volume:Start page-End page.**

| **For Books:**<br>**Line1 Authors**<br>**Line2 Chapter or Section Name**<br>**Line3 Book Name**<br>**Line4 Editors. Publication City: Publisher, Year.** | |
|---|---|
| **Article Type** | **Articles with PMID or No PMID** |
| | |
| | |
| | |

| **Study Attribution: will appear as submitted.**<br>**Column1 Header (i.e., Principal Investigator, Co-Investigator, Institute, Funding Source)**<br>**Column2 Name of the person or Name of Grant**<br>**Column3 Affiliation (include City, State, Country)**<br>**Example: http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000071.v2.p1** | | |
|---|---|---|
| **Header** | **Name** | **Affiliation** |
| | | |
| | | |

| **Grouping dbGaP studies "By Disease": All dbGaP studies will be grouped under one or more of the headings below. Please check all that apply. (http://www.nlm.nih.gov/mesh/2008/MeSHtree.C.html)** | |
|---|---|
| Bacterial Infections and Mycoses | |
| Virus Diseases | |
| Parasitic Diseases | |
| Neoplasms | |
| Musculoskeletal Diseases | |
| Digestive System Diseases | |
| Stomatognathic Diseases | |
| Respiratory Tract Diseases | |
| Otorhinolaryngologic Diseases | |
| Nervous System Diseases | |
| Eye Diseases | |
| Male Urogenital Diseases | |
| Female Urogenital Diseases and Pregnancy Complications | |
| Cardiovascular Diseases | |
| Hemic and Lymphatic Diseases | |

| | |
|---|---|
| Congenital, Hereditary, and Neonatal Diseases and Abnormalities | |
| Skin and Connective Tissue Diseases | |
| Nutritional and Metabolic Diseases | |
| Endocrine System Diseases | |
| Immune System Diseases | |
| Disorders of Environmental Origin | |
| Pathological Conditions, Signs and Symptoms | |
| Behavior and Behavior Mechanisms | |
| Psychological Phenomena and Processes | |
| Mental Disorders | |
| Behavioral Disciplines and Activities | |
| Population, Convenience, Spouse, Controls | |

Appendix B

# NCI dbGaP Data Submission Information

**(Basic Study Information form)**

In order for the **NIH NATIONAL CANCER INSTITUTE** to register your data into the dbGaP Submission System, please provide the information listed below and return to your NIH Program Officer (PO), or intramural Genomic Program Administrator (GPA). You may use the sample documents or any other format.

Checklist for required documents:
☐Institutional Certification
☐dbGaP Data Submission Information

## PART I – Study Registration Information

| | | |
|---|---|---|
| Study name: | | |
| Is this a multi-center study? (Y/N)<br>**If YES, please list participating sites:** | | |
| Target data delivery date: (YYYY-MM-DD) | Target public release date: (YYYY-MM-DD) | |
| Estimated number of study participants: | | |
| Is data submission expected to the following repositories? (Description of repositories) | Sequence Read Archive (SRA): (Y/N) | Trusted Partner (e.g. Bionimbus, GDC) (Y/N/NA) |

## PART II – Principal Investigator (PI) and Funding Information

| | |
|---|---|
| PI name: | PI e-mail: |
| PI institution: | |
| PI assistant/submitter name: | PI assistant/submitter e-mail: |
| Do you have an eRA Commons or an NIH account? (Y/N)<br>If **YES**, go to next question.<br>If **NO**, please register at https://commons.era.nih.gov/commons/registration/registrationInstructions.jsp. | |
| NIH Intramural Project (Z01), Grant or Contract Number: | NIH PO (if applicable): |
| NIH Institutes/Centers supporting the study: | |

## PART III – Policy

**Do you have Institutional Certification (IC) to submit these data?** (Y/N)

The IC should include the **Data Use Limitations (DUL)**, which are based on the informed consent given by each research subject. For every research subject, his/her corresponding data will be tagged with the appropriate DUL. Each study may have multiple DULs, based on the informed consent in the study.

If **YES**, send attachment to NIH PO or intramural GPA, along with this form.

If **NO**, please obtain the Institutional Certification from your Institutional Official. dbGaP requires that the sponsoring IC verifies that this certification has been met. A description of the requirements for the Institutional Certification and an example may be found in the accompanying "Submission into the NIH Database of Genotypes and Phenotypes (dbGaP)" guide.

## PART IV – Study Description

**Study type(s) (e.g., longitudinal, case-control, case set, control set, parent-offspring trios, cohort):**

**Is aggregate-level data appropriate for General Research Use?** (Y/N)

If **YES**, aggregate data will be included in the Compilation of Aggregate Genomic Data, a collection of analyses across many dbGaP studies that can be accessed with a single Data Access Request.
**NOTE:** This should be consistent with the Institutional Certification

Samples genotyped/sequenced:

Please check all data types expected for this study:

| **General** | **Sample Types** | **Array Data** |
|---|---|---|
| ☐Individual Phenotype<br>☐Individual Genotype<br>☐Individual Sequencing<br>☐Supporting Documents<br>☐Metagenomic<br>☐Protomic/Metabolomic<br>☐Images | ☐Germline<br>☐Tumor/Normal<br>☐DNA<br>☐RNA<br>☐Mitochondria<br>☐Microbiome<br>☐From Repository | ☐SNP Array<br>☐Expression Array<br>☐Methylation Array |
| **Genotypes** | **Sequencing** | **Analyses** |
| ☐Array derived Genotypes<br>☐CNV calls from miroarray<br>☐CNV calls derived from Sequencing<br>☐Genotype calls derived from Sequence<br>☐Somatic SNV (.MAF)<br>☐Array CGH CNVs | ☐Whole Genome<br>☐Whole Exome<br>☐Targeted Genome<br>☐Targeted Exome<br>☐Whole Transcriptome<br>☐Targeted Transcriptome<br>☐Epigenomic Marks<br>☐Sanger<br>☐16S rRNA | ☐Association/Linkage Results<br>☐Array derived Expression<br>☐RNA Seq derived Expression<br>☐Array derived Methylation |

**Genotype/Sequence platform information**

| **Name and version** | **Vendor** | **# Probes** | **URL** | **Description (optional)** |
|---|---|---|---|---|
| *Example: [GenomeWideSNP_6] Affymetrix Genome-Wide Human SNP 6.0 Array* | *Affymetrix* | *1880794* | *http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL6801* | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |

| **PART V – Acknowledgement Statement(s)*** ** |
|---|
| The submitting PI should provide specific points that should be included in an acknowledgement, such as sources of support or collaborators who have made subjects or samples available. Any NIH support must be specifically acknowledged by including the grant number. Consider citing a specific publication that comprehensively describes the origin of the dataset.<br><br>The suggested Acknowledgement Statement to accompany the dataset is:<br><br>Click here to enter text. |
| **PART VI – Original Summary of Study** |
| Please provide an original description of the study.[1]<br><br>Click here to enter text. |

[1] If the submitting institution certifies that aggregate data from a project can be included in the Compilation, then a study description for the aggregate data should be provided in addition to a description for the individual data.