OMB Control Number: 0925-XXXX
Expiration Date: XX/XX/XXXX

NCI dbGaP Initial Study Registration Information

In order for the National Cancer Institute to begin registration of your study in the dbGaP Submission System, please complete this form and the Institutional Certification form and send both to your Genomic Program Administrator (GPA).
Information about the Institutional Certification form is available on the NIH Scientific Data Sharing website: https://sharing.nih.gov/genomic-data-sharing-policy/institutional-certifications/completing-an-institutional-certification-form
Information about registering and submitting data into dbGaP, including contacts for data submission questions, is available at the NIH dbGaP Study Submission Guide: https://www.ncbi.nlm.nih.gov/gap/docs/submissionguide/

PART I – PI and PI Assistant/Data Submitter Information

Principal Investigator (PI): Provide the study PI information. This person must have an eRA Commons account to access the study registration (https://www.era.nih.gov/).

PI Name:

PI Email:

PI Institution:

eRA Commons Username (if available):

PI Assistant/Submitter: If someone other than the PI should be the main point of contact for the study and receive the invitation to the dbGaP Submission Portal, please provide that contact information here. This person must have an eRA Commons account to access the study registration (https://www.era.nih.gov/).

PI Assistant/Submitter Name:

PI Assistant/Submitter Email:

PI Assistant/Submitter Institution:

eRA Commons Username (if available):

PART II – dbGaP Study Name and Source of Study Samples

Study Name (this should match the study name on the Institutional Certification form, 255 character limit in the system):

Source(s) of Study Samples – check all that apply:

☐ **Clinical trial(s) – check all that apply:**
- ☐ NCI CTEP-supported clinical trial (including trials conducted in the NCTN, ETCTN, or other CTEP networks)
- ☐ Other NCI-supported clinical trial
- ☐ NCI Intramural clinical trial
- ☐ Other NIH-supported clinical trial
- ☐ Clinical trial not supported or funded by NIH, specify:

☐ **Source(s) other than a clinical trial – check all that apply:**
- ☐ NCI-supported study collecting specimens
- ☐ Other NIH-supported study collecting specimens
- ☐ Study not supported or funded by NIH, specify:

For all clinical trials from which specimens and/or data are being used, provide the clinical trial name and the ClinicalTrials.gov NCT number:

For all studies from which specimens and/or data are being used, provide the study name:

For all NCI- or NIH-supported clinical trials from which specimens and/or data are being used, provide the Grant, Contract, and/or Intramural Project numbers supporting the clinical trials:

For all NCI- or NIH-supported studies from which specimens and/or data are being used, provide the Grant, Contract, and/or Intramural Project numbers supporting the studies:

Briefly describe (including relevant grant/contract/project number(s)) any other NCI or NIH support for this study:

PART III – Acknowledgment Statement

Acknowledgment Statement: This is how future users will be expected to acknowledge this dataset in publications once it is released. Please provide specific points that should be included in the Acknowledgment, such as sources of support or collaborators who provided subjects or samples. **NIH support must be specifically acknowledged by including the grant number.** Consider citing a publication that comprehensively describes the origin of the dataset.

Acknowledgment Statement for this study:

Do you plan to deposit data using dbGaP for this study? ◯ **Yes** ◯ **No**

If **YES**, go to **Part IV**.
If **NO**, skip Part IV and go to **Part V**.

PART IV – Plans for Data Deposition Using dbGaP (Initial Study Data Outline)

Part IV should only be completed if you plan to deposit data using dbGaP for this study. For each category below, indicate whether you intend to deposit that type of data using dbGaP for this study (Yes or No). If yes, provide any further details requested within that category.

Subject and Sample IDs: these data will be submitted if any data are being deposited using dbGaP. The answer to this question will typically be "Yes."

Do you intend to deposit Subject and Sample IDs in dbGaP? ◯ **Yes** ◉ **No**

If "Yes," identify the sample types that data are expected from (check all that apply):

☐ Germline	☐ DNA	☐ Microbiome
☐ Tumor	☐ RNA	☐ From repository
☐ Normal	☐ Mitochondria	☐ Other, please specify:

Phenotype Data: these datasets include data associated with the subject, including clinical and demographic data, and data associated with the sample. The answer to this question will typically be "Yes."

Do you intend to deposit Phenotype Data in dbGaP? ◯ **Yes** ◉ **No**

Molecular Data: This does **not** include raw sequencing data and alignment information, which will be described below under "Sequence Data." This **does include** any GWAS, SNP array, imputations, transcriptomic, epigenomic, gene expression, variant calls from WGS, WXS, and targeted sequencing data**.**

Do you intend to deposit Molecular Data in dbGaP? ◯ **Yes** ◉ **No**

If "Yes," identify the molecular data that are expected to be submitted (check all that apply):

☐ Sequence-based multi sample genotype files	☐ CNV calls from microarray
☐ Array-based multi sample genotype files	☐ Methylation array data or summaries
☐ Single sample genotype files derived from sequencing	☐ Expression array data or summaries
	☐ CNV calls derived from sequencing
☐ Imputed genotypes (IMPUTE, etc.)	☐ Other, please specify:
☐ Tumor/normal variations / somatic SNV (.MAF)	

Sequence Data: Any high throughput sequence data (WGS, WXS, RNA-Seq, etc.) in BAM, CRAM, FASTQ formats**.**

Do you intend to deposit Sequence Data and/or Sequence Metadata in dbGaP? ◯ **Yes** ◉ **No**

If "Yes," please indicate how you intend to store the sequence data:

☐ Sequence metadata and sequence data will be stored using NCBI storage for sequence data *(most common)*
☐ Sequence metadata will be submitted to NCBI, but the sequence data will be stored in a separate NCI cloud storage location.

If sequence data will be stored in an NCI cloud storage location, please indicate the NCI cloud storage location:

☐ Google Cloud ☐ Amazon Cloud ☐ Other, please specify:

If sequence data will be stored in NCI cloud storage, who is the NCI contact for this cloud storage location?

Cloud Storage Contact Name: Contact Email:

If Sequence Data and/or Metadata will be deposited, identify the sequencing data expected to be submitted:

☐ Whole or targeted genome
☐ Whole or targeted exome
☐ RNAseq – whole or targeted transcriptome
☐ MethylSeq / Epigemonic Marks
☐ Microbiome sequence (e.g., metagenome, 16S rRNA)
☐ Other, please specify:

Links to Other Study Data in Unrestricted-Access NCBI Databases: *(not typical)*.

Do you intend to link to other study data in unrestricted-access NCBI databases? ◯ **Yes** ◉ **No**

If "Yes," check all that apply:

☐ **GEO:** repository of high-throughput gene expression data and hybridization arrays, chips, microarrays
☐ **SRA:** archive of raw sequencing data and alignment information from high-throughput sequencing platforms
☐ **GenBank:** genetic sequence database comprising an annotated collection of publicly available DNA sequences
☐ **Other**, please specify:

Association Analysis: Any aggregated genomic level data *(not typical).*

Do you intend to deposit Association Analysis Data in dbGaP? ◯ **Yes** ◉ **No**

If "Yes," check all that apply:

☐ SNP Based Association Results
☐ Gene Based Association Results

If "Yes," should the analyses be included in the Compilation of Aggregate Genomic Data (CADA), a collection of analyses across many dbGaP studies that can be accessed with a single Data Access Request? ◯ **Yes** ◉ **No**

Study Documents: Any consent forms, protocols, questionnaires, etc., that correspond to the data *(not typical)*.

Do you intend to deposit Study Documents in dbGaP? ◯ **Yes** ◉ **No**

Medical Images: Any scan images, etc. *(not typical)*.

Do you intend to deposit Medical Images in dbGaP? ◯ **Yes** ◉ **No**

Other: Any other files not covered above or notes about the expected data *(not typical)*.

Do you intend to deposit other files not covered above? ◯ **Yes** ◉ **No**
Is there other information the dbGaP curators should have about the expected data?

PART V – Plans for Data Deposition Using External Data Sources

External Data Source (EDS): A non-dbGaP NIH Institute or Center-Supported Repository that meets core NIH standards for establishing data quality and data management service protocols for NIH, based on the programmatic need of the NIH Institute or Center. Studies with data in the EDS will require credentialed users to apply for access to the data through dbGaP Authorized Access. Use of an External Data Source to deposit data may require prior approval from NCI.

Do you plan to deposit data in an external data source for this study? ◯ **Yes** ◉ **No**

If "Yes," identify the external data source:

Who is the NCI or NIH contact for this external data source?

Contact Name:

Contact Email: