Navigating dbGaP Public Page / Advanced Search

July 1, 2025

Overview

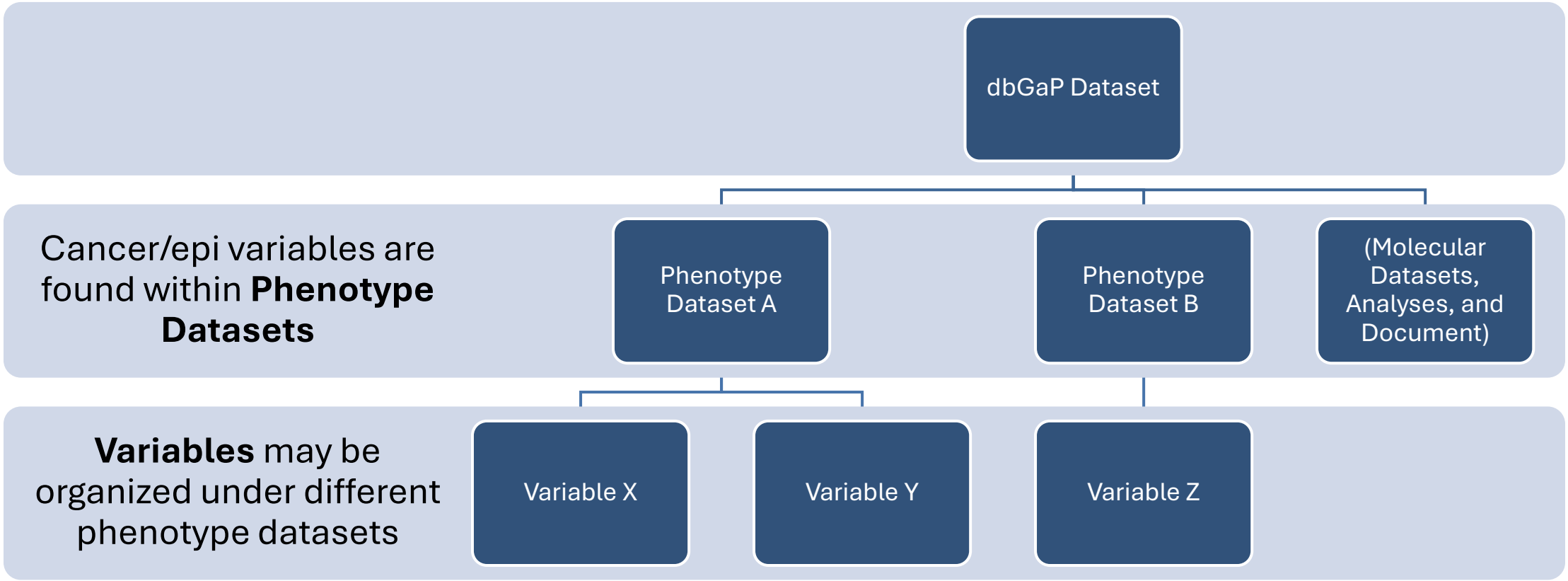
dbGaP Main Study Page

- Study description
- List of Consent Groups and the number of participants per group
- Links to other NCBI/dbGaP resources related to this study

Available Tabs

- **Phenotype Datasets**
- **Variables**
- **Molecular Datasets**
- **Documents**

Cancer/Epidemiological Variables in dbGaP Datasets

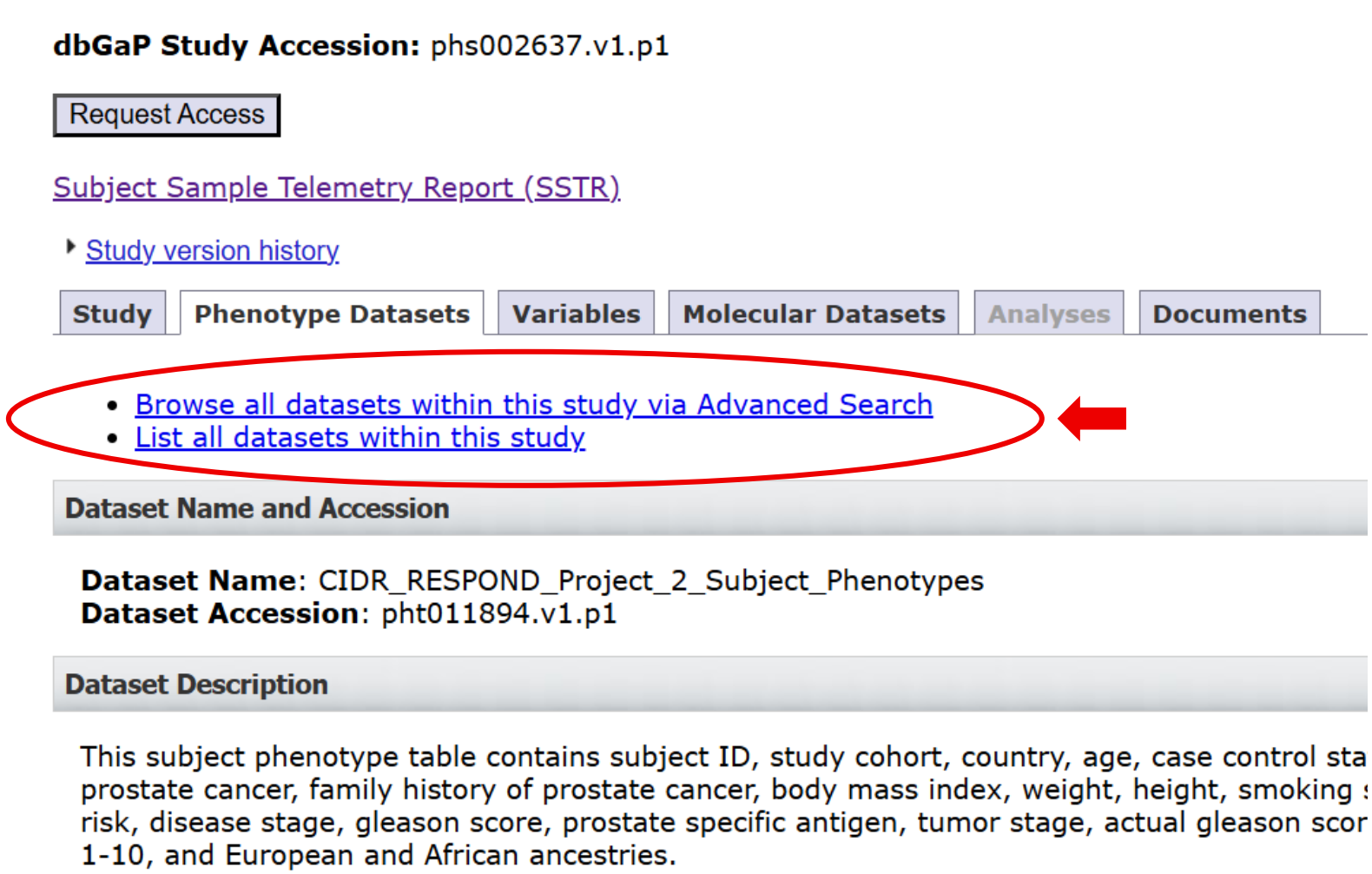


Phenotype Datasets Tab

Phenotype Datasets dbGaP Page provides Dataset description and list of variables

- Links to the Variable Report and Data Dictionary are also provided
- To view all datasets, you must select one of the links in the red circle

NOTE: dbGaP page only shows one dataset at a time



List of all Phenotype Datasets

****Each Dataset name link takes you to the dbGaP page for that dataset**

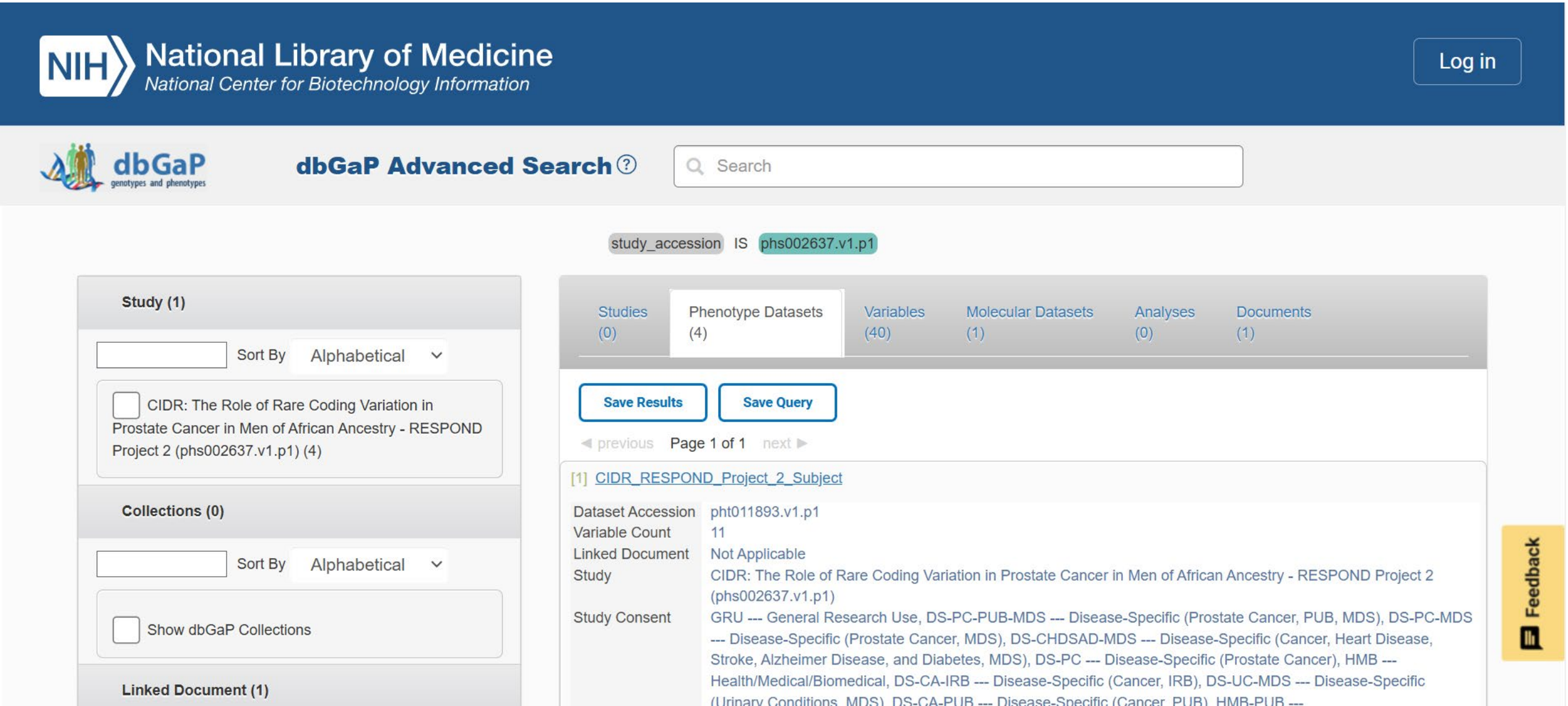
CIDR: The Role of Rare Coding Variation in Prostate Cancer in Men of African Ancestry - RESPOND Project 2

dbGaP Study Accession: *phs002637.v1.p1*

List of Datasets

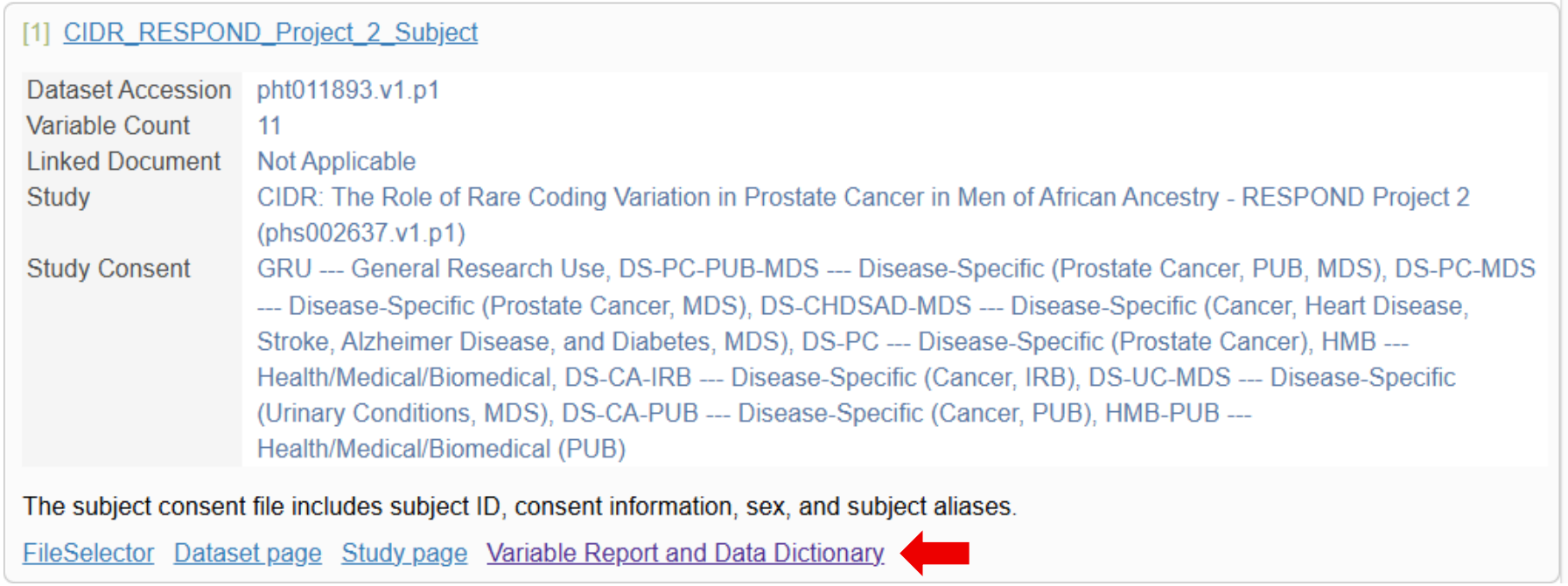
Dataset accession	Dataset name	Dataset description
pht011893.v1.p1	CIDR_RESPOND_Project_2_Subject	The subject consent file includes subject ID, consent information, sex, and subject aliases.
pht011894.v1.p1	CIDR_RESPOND_Project_2_Subject_Phenotypes	This subject phenotype table contains subject ID, study cohort, country, age, case control status of the subject for prostate cancer, family history of prostate cancer, body mass index, weight, height, smoking status, aggressiveness risk, disease stage, gleason score, prostate specific antigen, tumor stage, actual gleason score, principal components 1-10, and European and African ancestries.
pht012994.v1.p1	CIDR_RESPOND_Project_2_Sample	This data table contains a mapping of study subject IDs to sample IDs. Samples are the final preps submitted for genotyping, sequencing, and/or expression data. For example, if one patient (subject ID) gave one sample, and that sample was processed differently to generate 2 sequencing runs, there would be two rows, both using the same subject ID, but having 2 unique sample IDs.
pht012995.v1.p1	CIDR_RESPOND_Project_2_Sample_Attributes	This sample attributes table contains sample ID, body site where sample was collected, analyte type, and tumor status.

Browse Phenotype Datasets via Advanced Search



Browse Phenotype Datasets via Advanced Search

"Variable Report and Data Dictionary" is the direct link for the index of the phenotype dataset's phenotype variables summaries. This folder is also available through the main dataset's public FTP.



[1] CIDR_RESPOND_Project_2_Subject

Dataset Accession	pht011893.v1.p1
Variable Count	11
Linked Document	Not Applicable
Study	CIDR: The Role of Rare Coding Variation in Prostate Cancer in Men of African Ancestry - RESPOND Project 2 (phs002637.v1.p1)
Study Consent	GRU --- General Research Use, DS-PC-PUB-MDS --- Disease-Specific (Prostate Cancer, PUB, MDS), DS-PC-MDS --- Disease-Specific (Prostate Cancer, MDS), DS-CHDSAD-MDS --- Disease-Specific (Cancer, Heart Disease, Stroke, Alzheimer Disease, and Diabetes, MDS), DS-PC --- Disease-Specific (Prostate Cancer), HMB --- Health/Medical/Biomedical, DS-CA-IRB --- Disease-Specific (Cancer, IRB), DS-UC-MDS --- Disease-Specific (Urinary Conditions, MDS), DS-CA-PUB --- Disease-Specific (Cancer, PUB), HMB-PUB --- Health/Medical/Biomedical (PUB)

The subject consent file includes subject ID, consent information, sex, and subject aliases.

FileSelector Dataset page Study page Variable Report and Data Dictionary

Variables Tab

Variables dbGaP page proves description of a variable and lists which dataset it belongs to

- Also provides statistical summary of variable across the consent groups
- To view all variables, you must select one of the links in the red circle

NOTE: dbGaP page only shows one variable at a time

List of all Variables

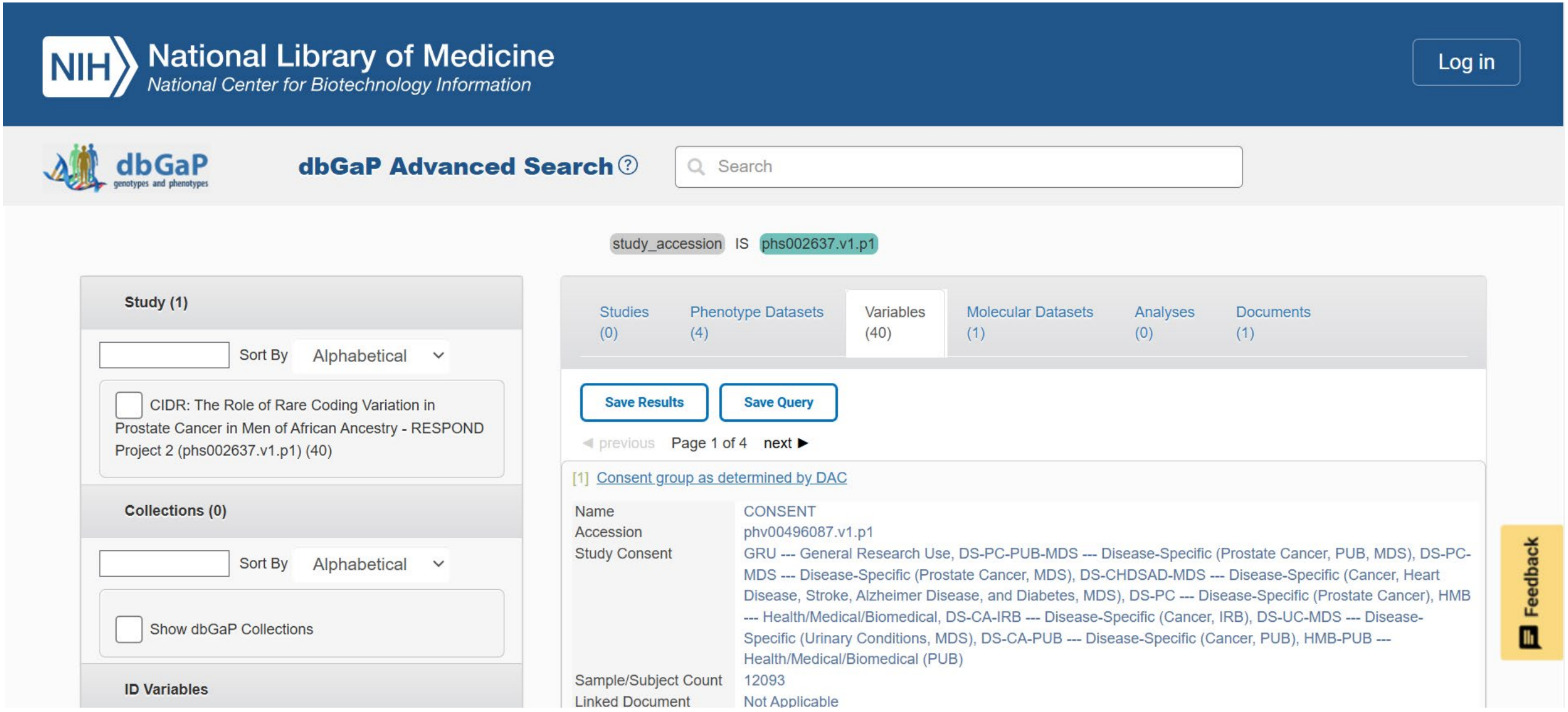
CIDR: The Role of Rare Coding Variation in Prostate Cancer in Men of African Ancestry - RESPOND Project 2

dbGaP Study Accession: *phs002637.v1.p1*

List of Variables

Variable accession	Variable name	Variable description	Dataset accession	Dataset name
phv00496086.v1.p1	SUBJECT_ID	Subject ID	pht011893.v1.p1	CIDR_RESPOND_Project_2_Subject
phv00496087.v1.p1	CONSENT	Consent group as determined by DAC	pht011893.v1.p1	CIDR_RESPOND_Project_2_Subject
phv00496088.v1.p1	SEX	SEX	pht011893.v1.p1	CIDR_RESPOND_Project_2_Subject
phv00496089.v1.p1	SUBJECT_SOURCE	Repository name in the sample report of study A	pht011893.v1.p1	CIDR_RESPOND_Project_2_Subject
phv00496090.v1.p1	SOURCE_SUBJECT_ID	Submitted subject_ID of study A	pht011893.v1.p1	CIDR_RESPOND_Project_2_Subject
phv00496091.v1.p1	SUBJECT_SOURCE2	Repository name in the sample report of study B	pht011893.v1.p1	CIDR_RESPOND_Project_2_Subject
phv00496092.v1.p1	SOURCE_SUBJECT_ID2	Submitted subject_ID of study B	pht011893.v1.p1	CIDR_RESPOND_Project_2_Subject
phv00496093.v1.p1	SUBJECT_SOURCE3	Repository name in the sample report of study C	pht011893.v1.p1	CIDR_RESPOND_Project_2_Subject
phv00496094.v1.p1	SOURCE_SUBJECT_ID3	Submitted subject_ID of study C	pht011893.v1.p1	CIDR_RESPOND_Project_2_Subject
phv00496095.v1.p1	SUBJECT_ID	Subject ID	pht011894.v1.p1	CIDR_RESPOND_Project_2_Subject_Phenotypes
phv00496096.v1.p1	STUDY	Study	pht011894.v1.p1	CIDR_RESPOND_Project_2_Subject_Phenotypes
phv00496097.v1.p1	COUNTRY	Country	pht011894.v1.p1	CIDR_RESPOND_Project_2_Subject_Phenotypes
phv00496098.v1.p1	AGE	Age	pht011894.v1.p1	CIDR_RESPOND_Project_2_Subject_Phenotypes
phv00496099.v1.p1	AFFECTION_STATUS	Case Control status of the subject for Prostate Cancer	pht011894.v1.p1	CIDR_RESPOND_Project_2_Subject_Phenotypes
phv00496100.v1.p1	FAMHIST	Family History of Prostate Cancer	pht011894.v1.p1	CIDR_RESPOND_Project_2_Subject_Phenotypes
phv00496101.v1.p1	BMI	Body Mass Index	pht011894.v1.p1	CIDR_RESPOND_Project_2_Subject_Phenotypes

Browse Variables via Advanced Search (1/2)



Browse Variables via Advanced Search (2/2)

[1] Consent group as determined by DAC

Name	CONSENT
Accession	phv00496087.v1.p1
Study Consent	GRU --- General Research Use, DS-PC-PUB-MDS --- Disease-Specific (Prostate Cancer, PUB, MDS), DS-PC-MDS --- Disease-Specific (Prostate Cancer, MDS), DS-CHDSAD-MDS --- Disease-Specific (Cancer, Heart Disease, Stroke, Alzheimer Disease, and Diabetes, MDS), DS-PC --- Disease-Specific (Prostate Cancer), HMB --- Health/Medical/Biomedical, DS-CA-IRB --- Disease-Specific (Cancer, IRB), DS-UC-MDS --- Disease-Specific (Urinary Conditions, MDS), DS-CA-PUB --- Disease-Specific (Cancer, PUB), HMB-PUB --- Health/Medical/Biomedical (PUB)
Sample/Subject Count	12093
Linked Document	Not Applicable
Study	CIDR: The Role of Rare Coding Variation in Prostate Cancer in Men of African Ancestry - RESPOND Project 2 (phs002637.v1.p1)
Dataset	CIDR_RESPOND_Project_2_Subject (pht011893.v1.p1)
Embargo Release Date	2023-06-14

Variable page Study page Dataset page

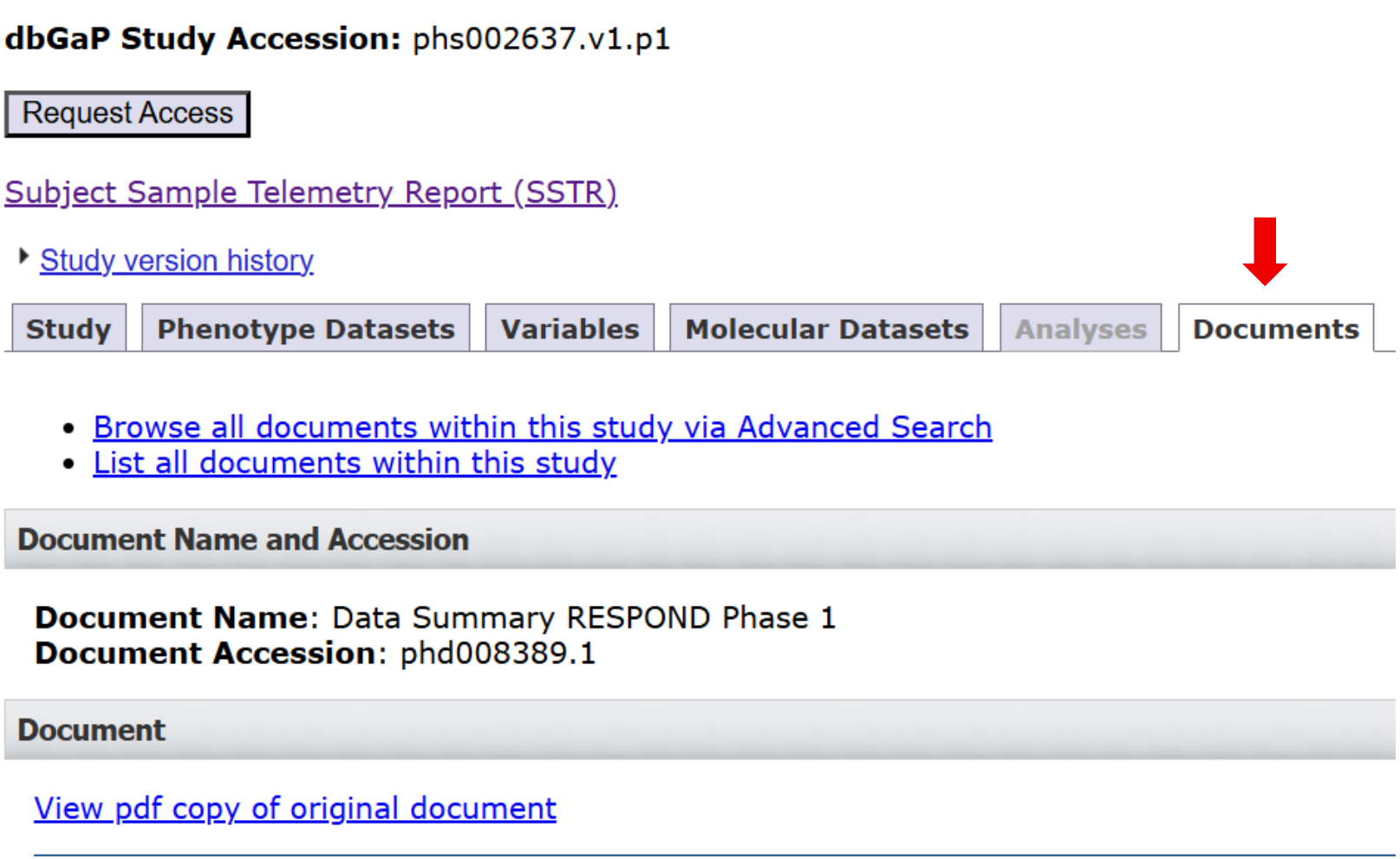
[2] SEX

Name	SEX
Accession	phv00496088.v1.p1
Study Consent	GRU --- General Research Use, DS-PC-PUB-MDS --- Disease-Specific (Prostate Cancer, PUB, MDS), DS-PC-MDS --- Disease-Specific (Prostate Cancer, MDS), DS-CHDSAD-MDS --- Disease-Specific (Cancer, Heart Disease, Stroke, Alzheimer Disease, and Diabetes, MDS), DS-PC --- Disease-Specific (Prostate Cancer), HMB --- Health/Medical/Biomedical, DS-CA-IRB --- Disease-Specific (Cancer, IRB), DS-UC-MDS --- Disease-Specific (Urinary Conditions, MDS), DS-CA-PUB --- Disease-Specific (Cancer, PUB), HMB-PUB --- Health/Medical/Biomedical (PUB)
Sample/Subject Count	12093
Linked Document	Not Applicable
Study	CIDR: The Role of Rare Coding Variation in Prostate Cancer in Men of African Ancestry - RESPOND Project 2 (phs002637.v1.p1)
Dataset	CIDR_RESPOND_Project_2_Subject (pht011893.v1.p1)
Embargo Release Date	2023-06-14

Variable page Study page Dataset page

Documents Tab

Documents dbGaP page is not always consistent but usually provides link or displays consent form, protocol, or questionnaire



Molecular Datasets Tab

Molecular Datasets dbGaP page provides a summary of all the available molecular data for the dataset, including sequencing data that is stored in SRA

Study | Phenotype Datasets | Variables | Molecular Datasets | Analyses | Documents

Summary of Molecular Data

Sample and subject counts organized by Consent Group and Sample Use

Study	Sample Use	DS-CA-IRB		DS-CA-PUB		DS-CHDSAD-MDS		DS-PC	
		samples	subjects	samples	subjects	samples	subjects	samples	subjects
phs002637.v1.p1	Seq_DNA_SNP_CNV	598	598	1855	1847	632	632	1418	1412
phs002637.v1.p1	WXS	597	597	1851	1843	632	632	1417	1411

Legend:

- **DS-CA-IRB:** Disease-Specific (Cancer, IRB)[consent code: 6]
- **DS-CA-PUB:** Disease-Specific (Cancer, PUB)[consent code: 8]
- **DS-CHDSAD-MDS:** Disease-Specific (Cancer, Heart Disease, Stroke, Alzheimer Disease, and Diabetes, MDS) [consent code: 3]
- **DS-PC:** Disease-Specific (Prostate Cancer)[consent code: 4]
- **DS-PC-MDS:** Disease-Specific (Prostate Cancer, MDS)[consent code: 2]
- **DS-PC-PUB-MDS:** Disease-Specific (Prostate Cancer, PUB, MDS)[consent code: 10]
- **DS-UC-MDS:** Disease-Specific (Urinary Conditions, MDS)[consent code: 7]
- **GRU:** General Research Use[consent code: 1]
- **HMB:** Health/Medical/Biomedical[consent code: 5]
- **HMB-PUB:** Health/Medical/Biomedical (PUB)[consent code: 9]
- **NRUP:** Subjects did not participate in the study, did not complete a consent document and are included only for genotype controls, such as HapMap subjects
- List of Sample Uses and their meaning

RunSelector (Note: use this link to get information (e.g. file size, release date etc.) about SRA data within this study. This link will not let you download the SRA data itself.)

dbGaP Accession Numbers

- Study Accession Number - Once the Study Data Outline (SDO) is completed, a study accession is assigned: phs######.v#.p#. The study accession is a unique, stable, and versioned identifier (ID) that can be used in publications. It is prefixed by "phs," indicating a phenotype study.
- The version number (.v#) and participant set number (.p#) do not change during iterations within a release cycle, but following release and only after changes have been made to existing data or new data is added. The Study v# is always incremented, while the v# for its components are only incremented when there are changes to that specific component. The p# is incremented when subjects in an existing study set changes consent status. The p# is never incremented when only new subjects are added and existing subjects have not changed consents.
- Dataset Accession Number - Each phenotype table (SC, SSM, pedigree, subject phenotypes, and sample attributes) is assigned a pht######.v#.
- Variable Accession Number - Each variable in a phenotype table (SC, SSM, pedigree, subject phenotypes, and sample attributes) is assigned a phv########.v#.
- Document Accession Number - Each study document (e.g. protocols, questionnaires, manuals of procedures and operations) is assigned a phd######.#, where .# is the version number.
- Molecular Data Accession Number - Each grouping of molecular data is assigned a phg######.v#.
- Analysis Accession Number - Each analysis is assigned a pha#######.v#.

Source: NCBI DbGaP Submission Guide

Resources

DbGaP Overview Video

- Watch the DbGAP Overview Video on YouTube

Submission Guide

- View the NCBI DbGaP Submission Guide