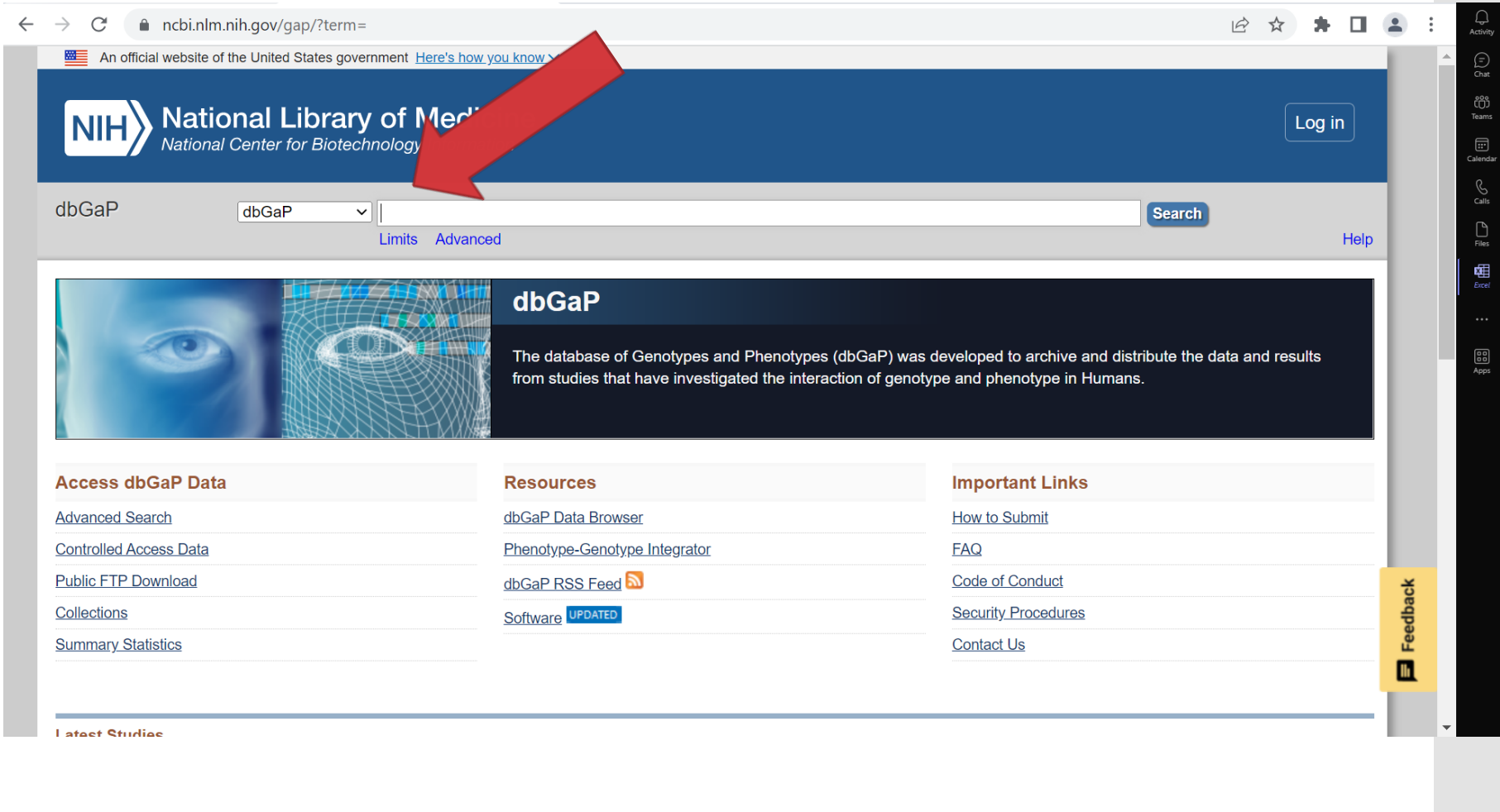
Finding Cancer Genomic Summary Results (GSR) in dbGaP

- When Accession number is known vs. not known
- Navigating dbGaP Public Study page materials and FTP site for GSR and related information

Start at the Main dbGaP website: https://www.ncbi.nlm.nih.gov/gap/

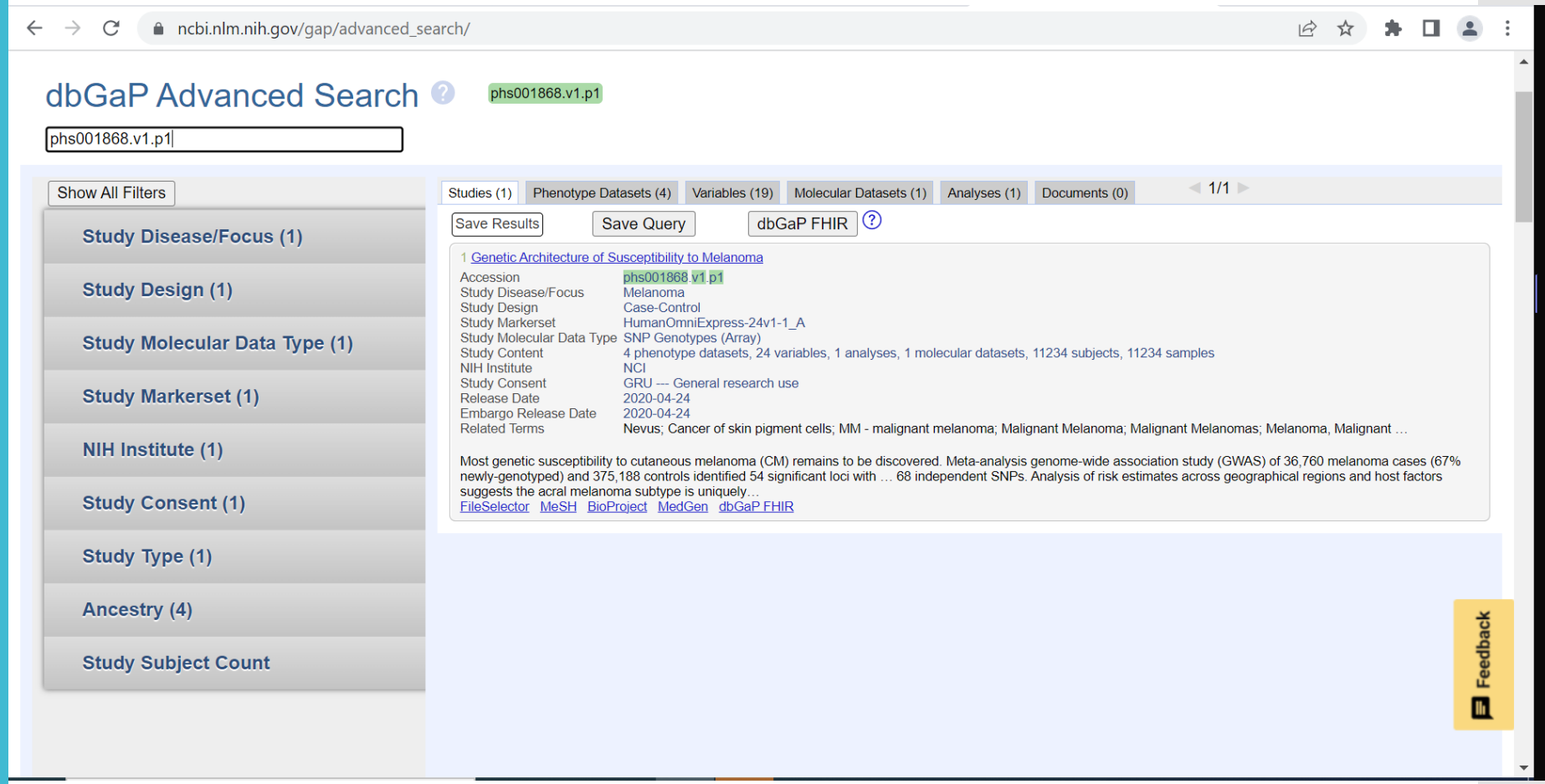
If the dbGaP Accession number is known: Enter the number starting with "phs" in the search bar (for the most comprehensive results, do not include .v# or .p# at the end)



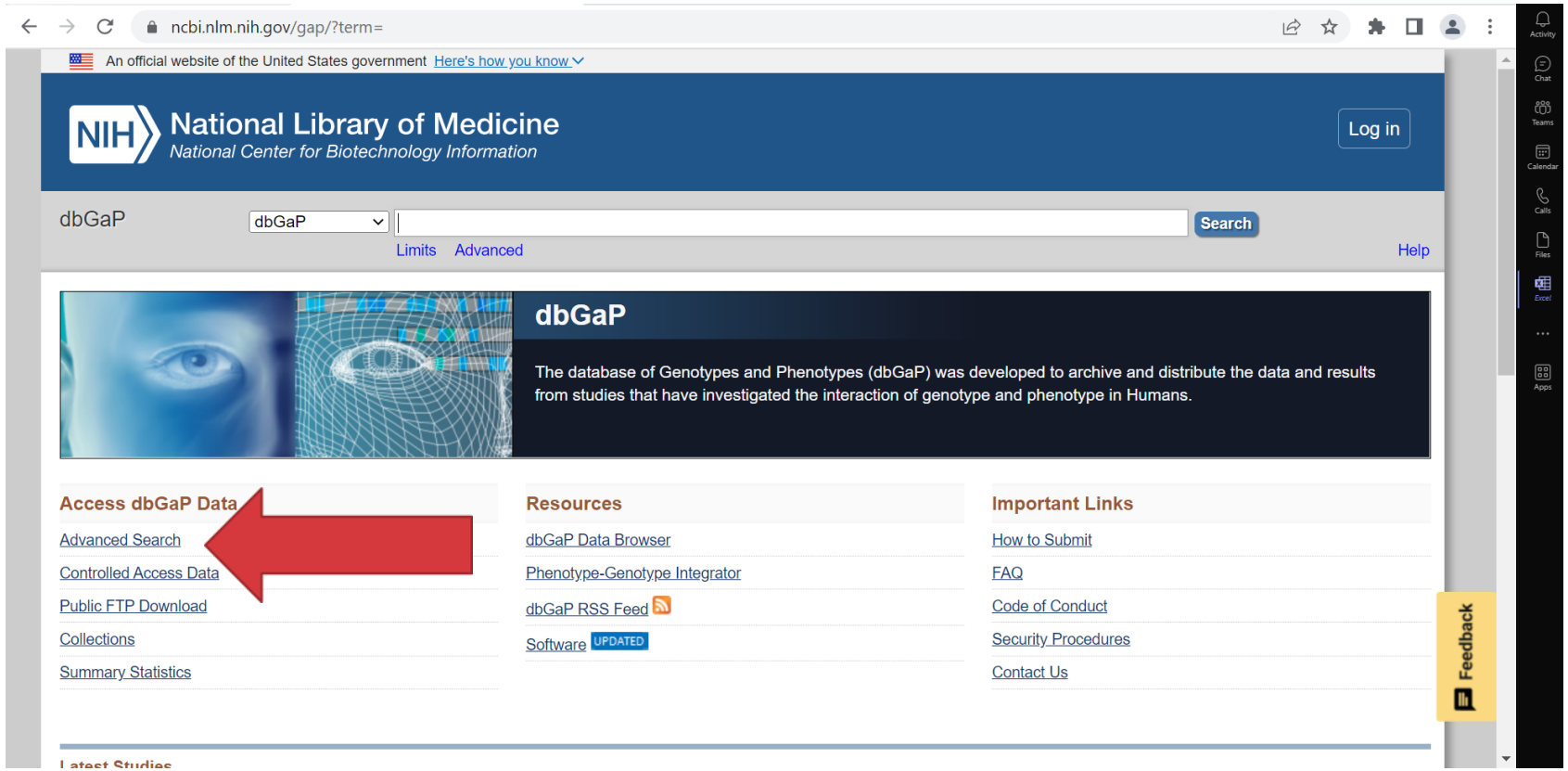
Accession Number: Search Results

Datasets associated with the accession number will be listed

GSR formatted and provided by the PI are under the "Analyses" tab

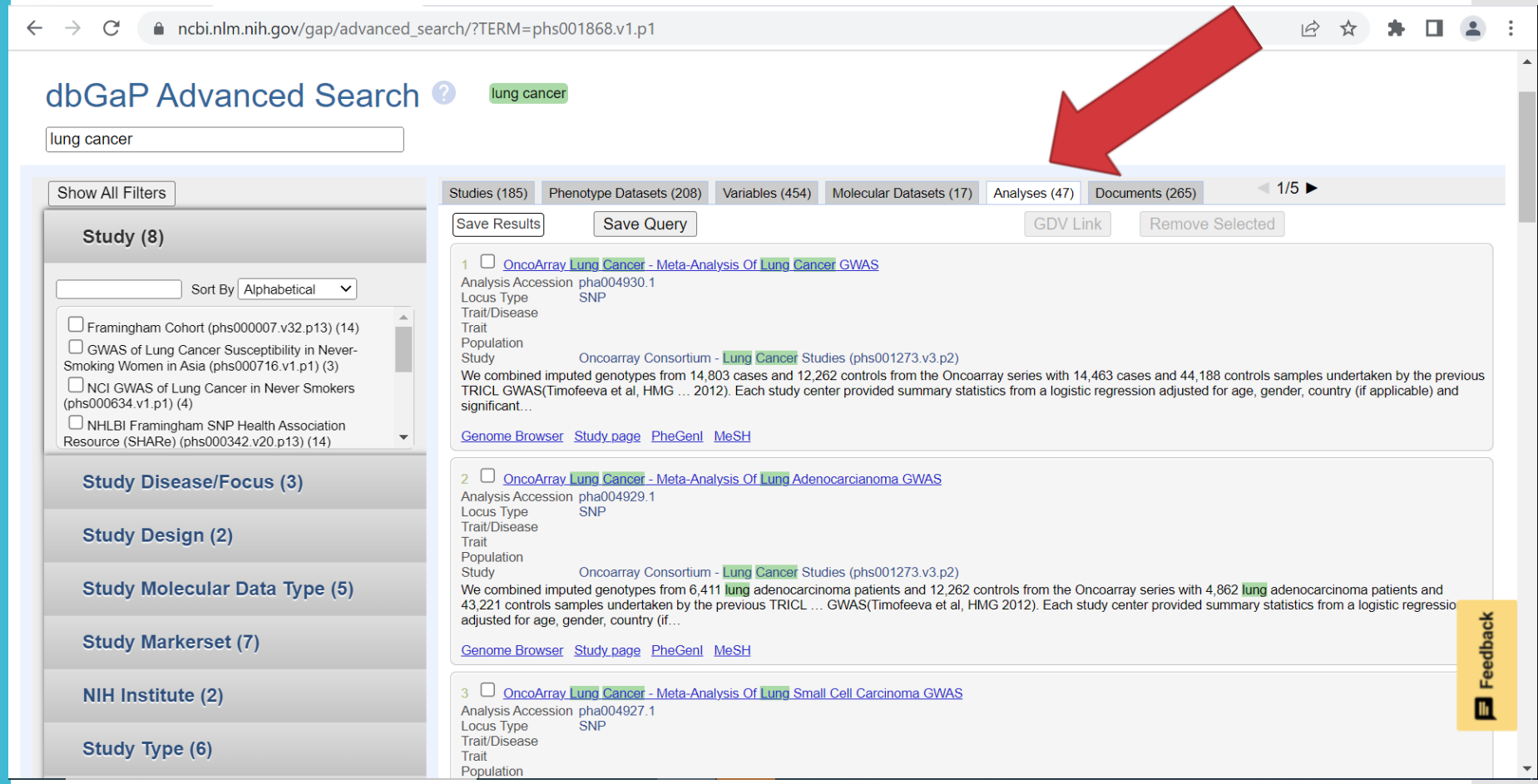


If the dbGaP Accession number is not known: Search cancer type in dbGaP Advanced Search



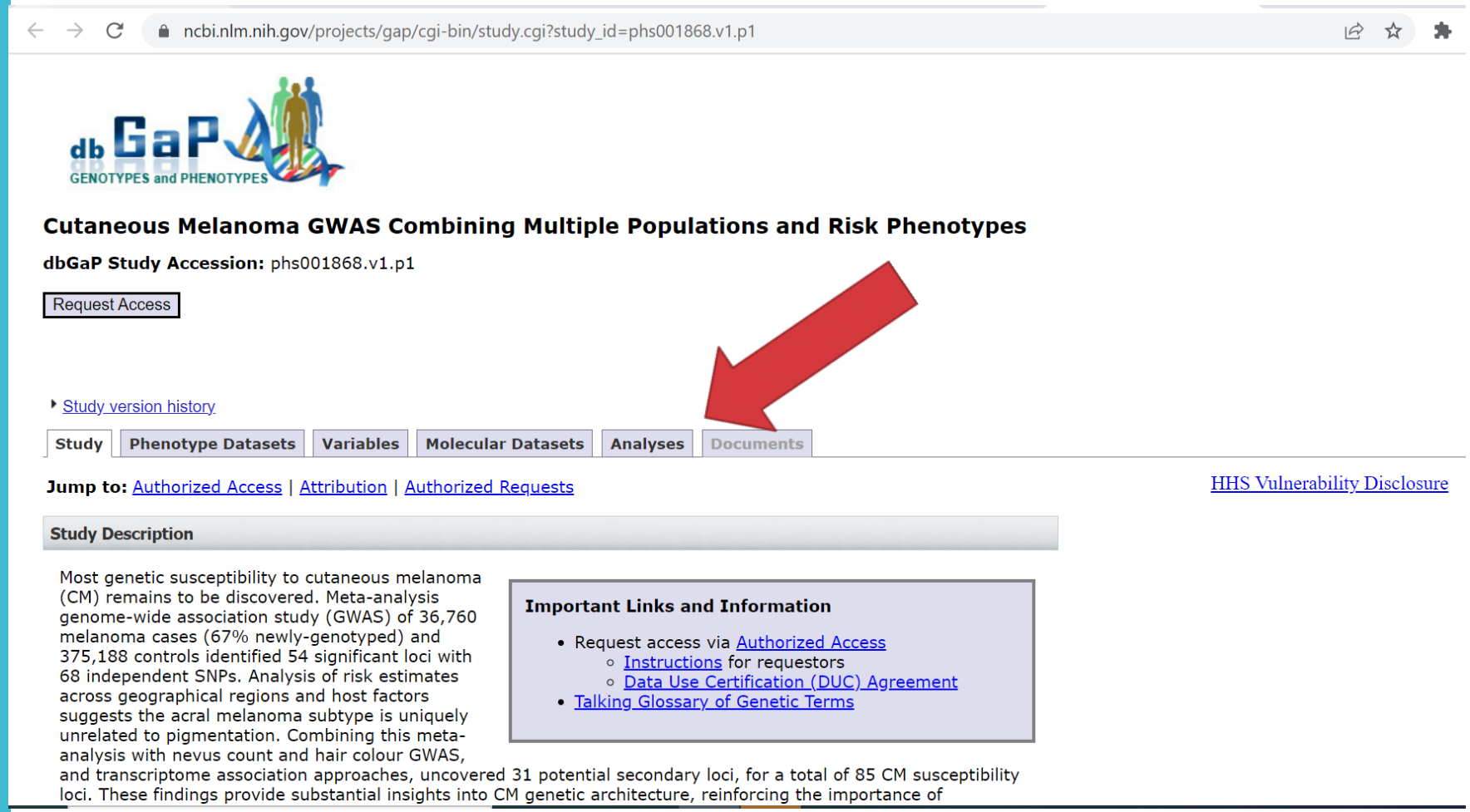
Cancer Type: Advanced Search Results

All investigator-provided GSR that match the cancer type searched will be listed under the "Analyses" tab



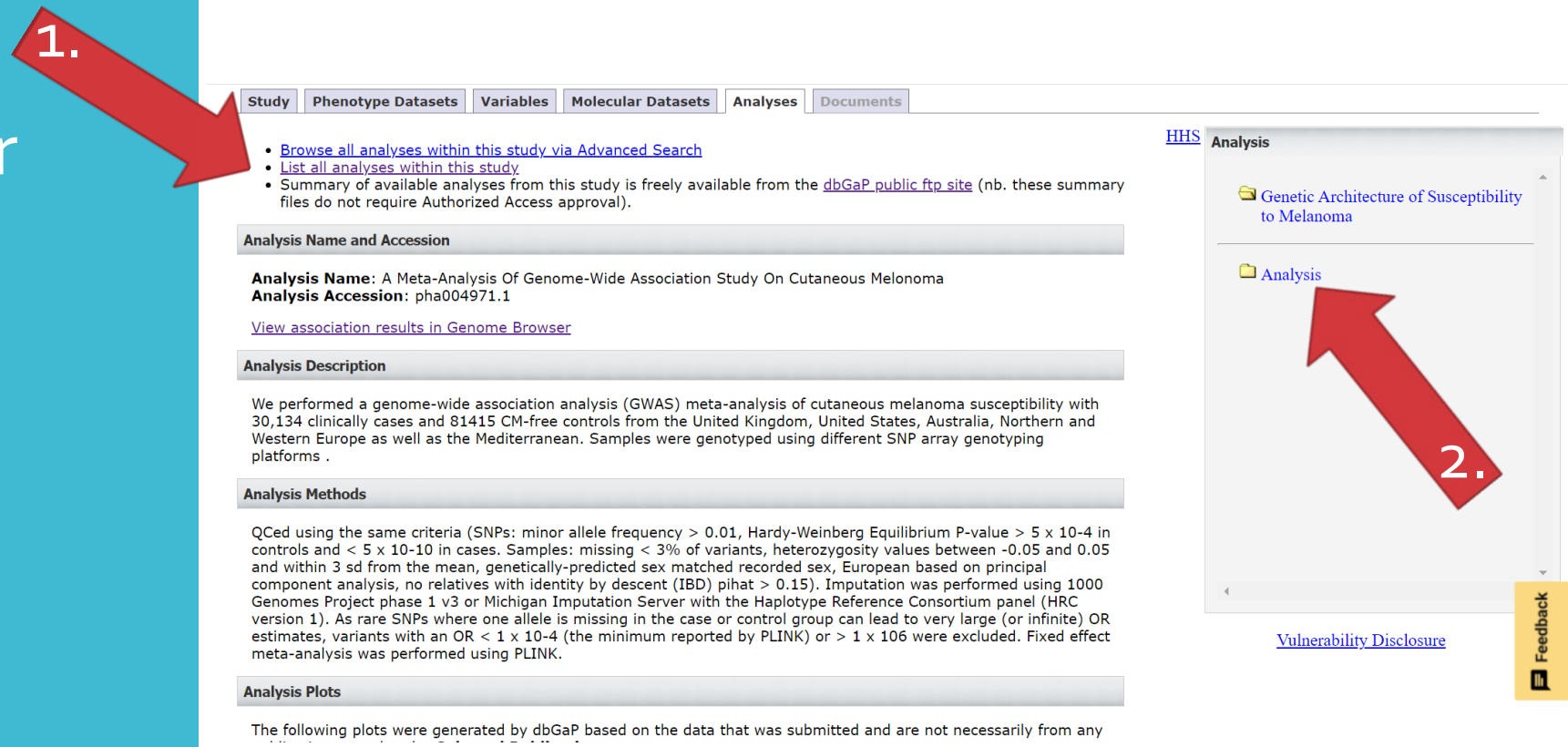
Navigating the dbGaP Study Page

Look for the "Analyses" tab



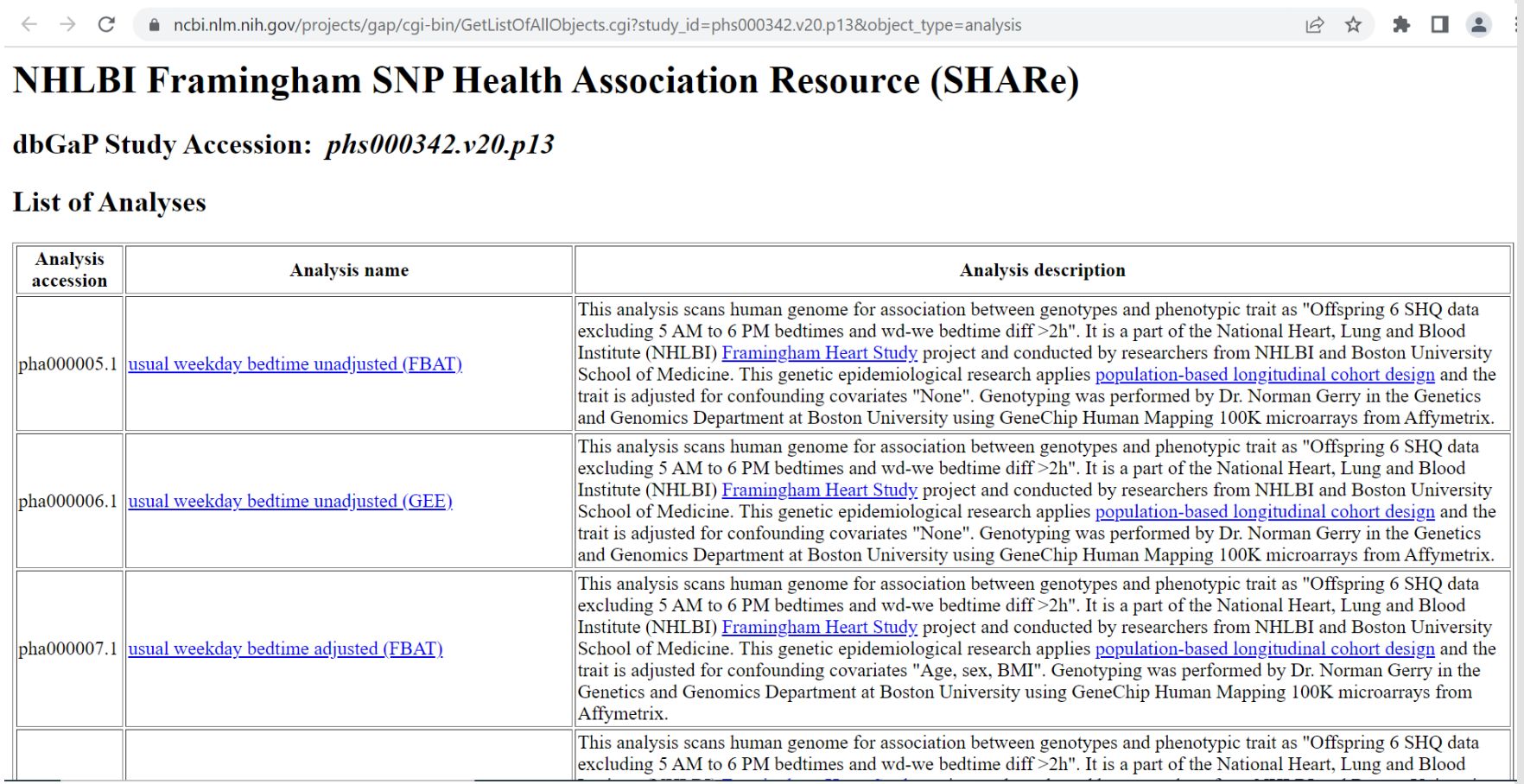
List of Analyses for the Dataset

1. Best for sharing
2. Best for navigating in browser



Example: "List all analyses within this study"

Note: dbGaP assigns each analysis its own accession number (pha), separate from the dataset accession number (phs)



ncbi.nlm.nih.gov/projects/gap/cgi-bin/GetListOfAllObjects.cgi?study_id=phs000342.v20.p13&object_type=analysis

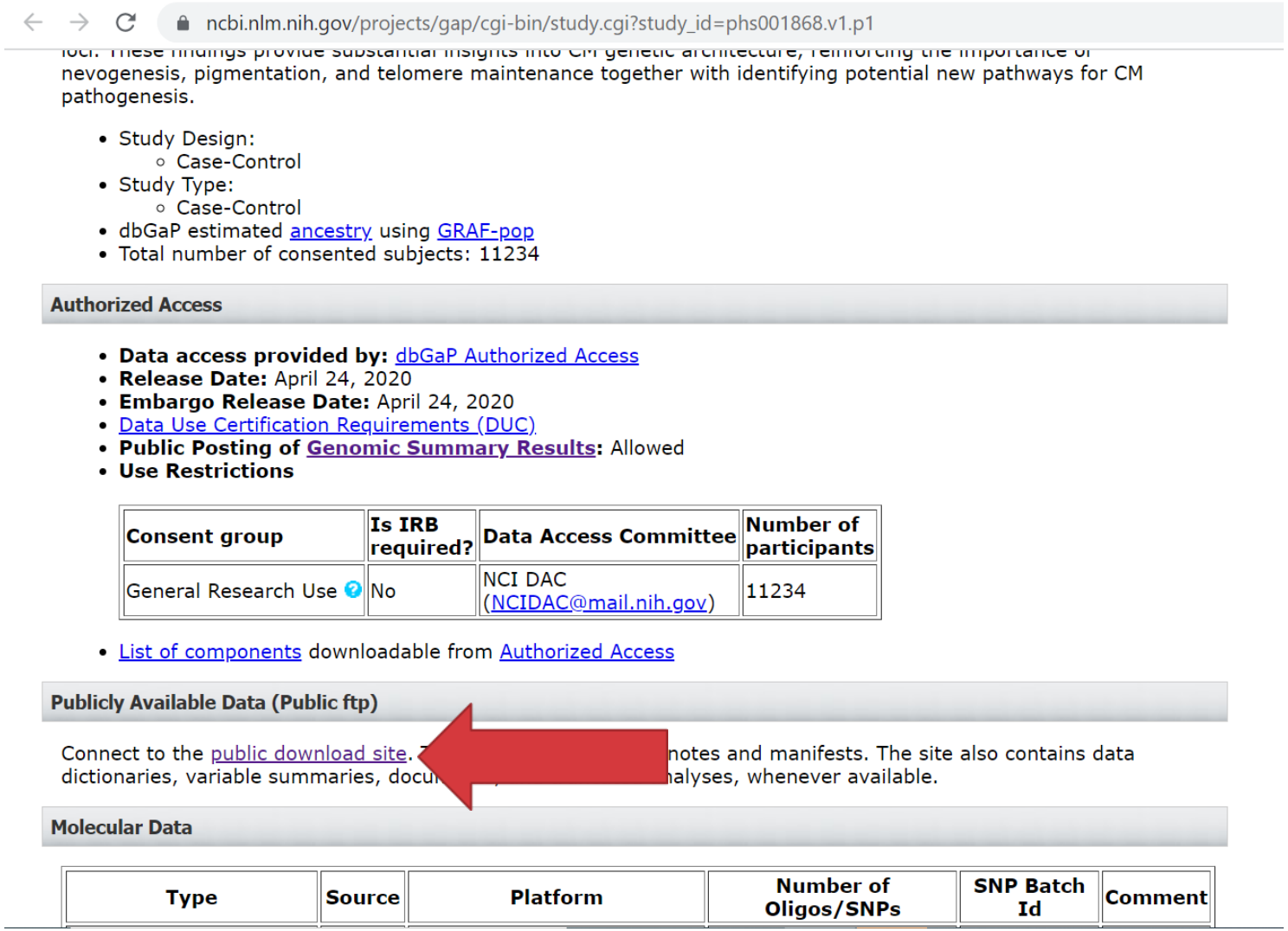
NHLBI Framingham SNP Health Association Resource (SHARe)

dbGaP Study Accession: ***phs000342.v20.p13***

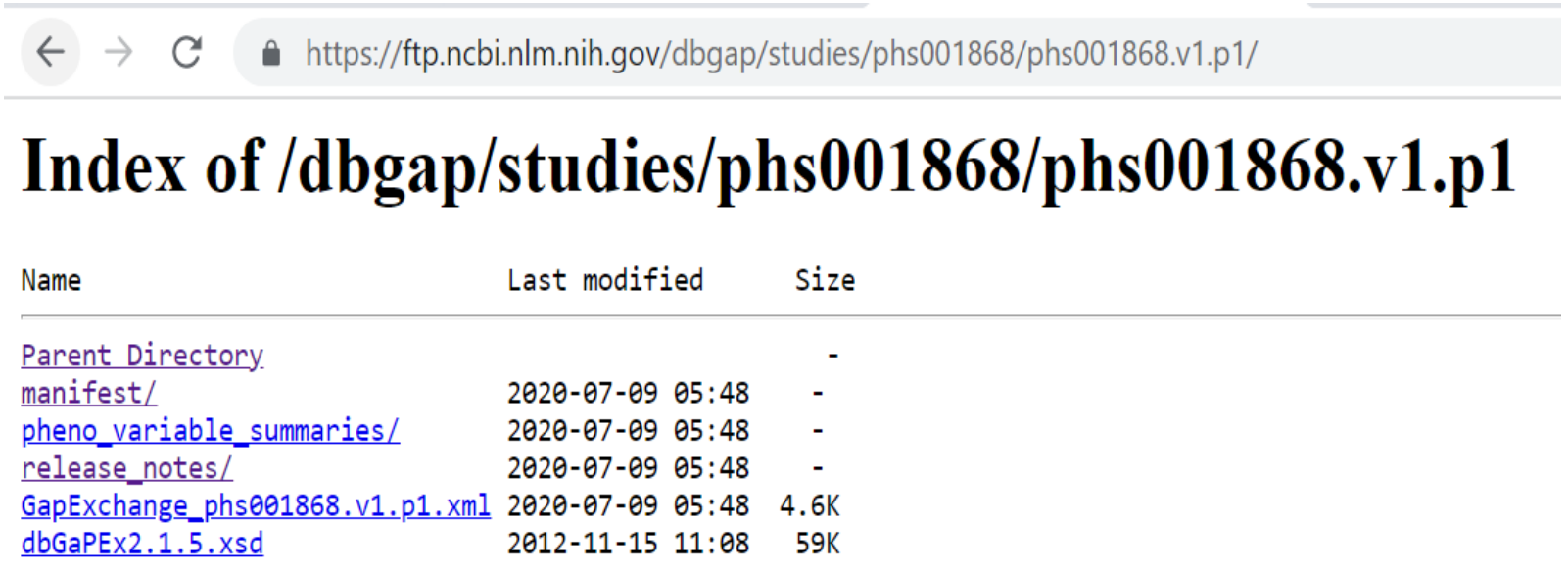
List of Analyses

Analysis accession	Analysis name	Analysis description
pha000005.1	usual weekday bedtime unadjusted (FBAT)	This analysis scans human genome for association between genotypes and phenotypic trait as "Offspring 6 SHQ data excluding 5 AM to 6 PM bedtimes and wd-we bedtime diff >2h". It is a part of the National Heart, Lung and Blood Institute (NHLBI) Framingham Heart Study project and conducted by researchers from NHLBI and Boston University School of Medicine. This genetic epidemiological research applies population-based longitudinal cohort design and the trait is adjusted for confounding covariates "None". Genotyping was performed by Dr. Norman Gerry in the Genetics and Genomics Department at Boston University using GeneChip Human Mapping 100K microarrays from Affymetrix.
pha000006.1	usual weekday bedtime unadjusted (GEE)	This analysis scans human genome for association between genotypes and phenotypic trait as "Offspring 6 SHQ data excluding 5 AM to 6 PM bedtimes and wd-we bedtime diff >2h". It is a part of the National Heart, Lung and Blood Institute (NHLBI) Framingham Heart Study project and conducted by researchers from NHLBI and Boston University School of Medicine. This genetic epidemiological research applies population-based longitudinal cohort design and the trait is adjusted for confounding covariates "None". Genotyping was performed by Dr. Norman Gerry in the Genetics and Genomics Department at Boston University using GeneChip Human Mapping 100K microarrays from Affymetrix.
pha000007.1	usual weekday bedtime adjusted (FBAT)	This analysis scans human genome for association between genotypes and phenotypic trait as "Offspring 6 SHQ data excluding 5 AM to 6 PM bedtimes and wd-we bedtime diff >2h". It is a part of the National Heart, Lung and Blood Institute (NHLBI) Framingham Heart Study project and conducted by researchers from NHLBI and Boston University School of Medicine. This genetic epidemiological research applies population-based longitudinal cohort design and the trait is adjusted for confounding covariates "Age, sex, BMI". Genotyping was performed by Dr. Norman Gerry in the Genetics and Genomics Department at Boston University using GeneChip Human Mapping 100K microarrays from Affymetrix.
		This analysis scans human genome for association between genotypes and phenotypic trait as "Offspring 6 SHQ data excluding 5 AM to 6 PM bedtimes and wd-we bedtime diff >2h". It is a part of the National Heart, Lung and Blood

Available analyses may also be on the Public FTP site



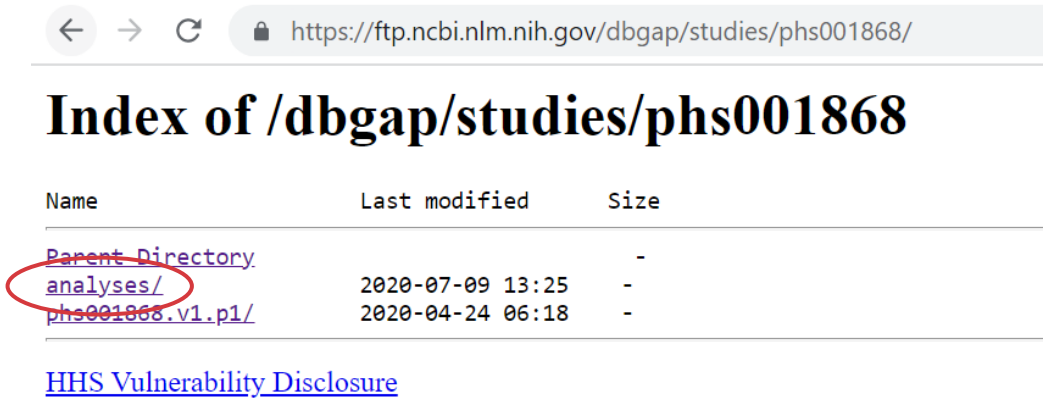
Public FTP Site Breakdown

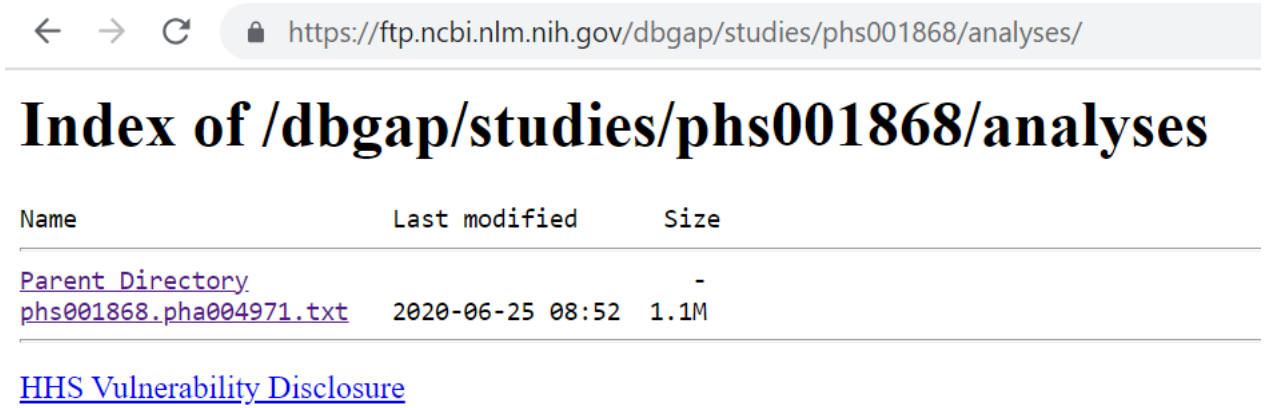


- FTP stands for File Transfer Protocol
- Folders contain publicly accessible data from the study
- Select "manifest" folder for a report document showing the breakdown of data in all the folders
- Folder "release notes" sometimes contain information on genomic summary results

Public FTP Site with Analyses

(available analysis format may differ)





No Analyses in Tab or Public FTP? As a last resort, also check dbGaP Study Release Notes

One possible location is under the "Molecular Data" section

dbGaP Study Release Notes

Release Notes for NCI MADCaP Sub-Saharan Africa, phs002718.v1.p1
"*Genetics of Prostate Cancer in Africa*"

For any questions or comments, please contact: dbgap-help@ncbi.nlm.nih.gov.

November 5, 2022 Version 1 Data set release date

2022-11-05
Version 1 Data set release for NCI MADCaP Sub-Saharan Africa now available

This release includes phenotype tables and SNP array (Array_SNP), and imputed genotype (Imputation_SNP_CNV) data. Please refer to the latest study configuration report for a detailed description of each download component.

There are no overlapping subjects between the three consent groups listed below.

Consent group 1 (c1): General Research Use (GRU)

Data Type	subjects	samples
Phenotype	690	690
Array_SNP	690	690
Imputation_SNP_CNV	690	690

Consent group 2 (c2): Disease-Specific (Cancer, IRB) (DS-CA-IRB)

Data Type	subjects	samples
Phenotype	2138	2138
Array_SNP	2138	2138
Imputation_SNP_CNV	2138	2138

Consent group 3 (c3): Disease-Specific (Prostate Cancer) (DS-PC)

Data Type	subjects	samples
Phenotype	1900	1900
Array_SNP	1900	1900
Imputation_SNP_CNV	1900	1900

For a description of SAMPLE_USE terms, please see:
https://www.ncbi.nlm.nih.gov/projects/gap/submission/GetSampleUseTypes.cgi

Molecular Data

1. Genotype data are accessioned under phg001706.v1 for data from MADCaP array and phg001753.v1 for imputed data files. In both cases, please see "sample-info" component for genotyped samples, consent status and mapping of sample to data files.
2. Genomic variants from array are available in originally submitted plink matrix format. They are from 4728 samples and split based on sample consent status. The data are packed separately in folders marked as "genotype-calls-matrixfmt".
3. CEL files used for image analysis of array are available in original format. They are split based on sample consent status and packed separately in folders marked as "raw-data-cel".

Example: Summary statistics are in a marked folder described in the notes

dbGaP Study Release Notes

4. QC results from dbGaP and submitter's README are in the folder marked as "genotype-qc".
5. Imputed genotypes from 4728 samples are available in originally submitted VCFv4.2 format. They are split based on sample consent status and packed into separate folders marked as "genotype-calls-vcf".
6. Summary statistical results from submitter are packed in the folder marked as "vcf-summary-data".

Authorized Access (Individual Level Data)

Individual level data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs002718/phs002718.v1.p1