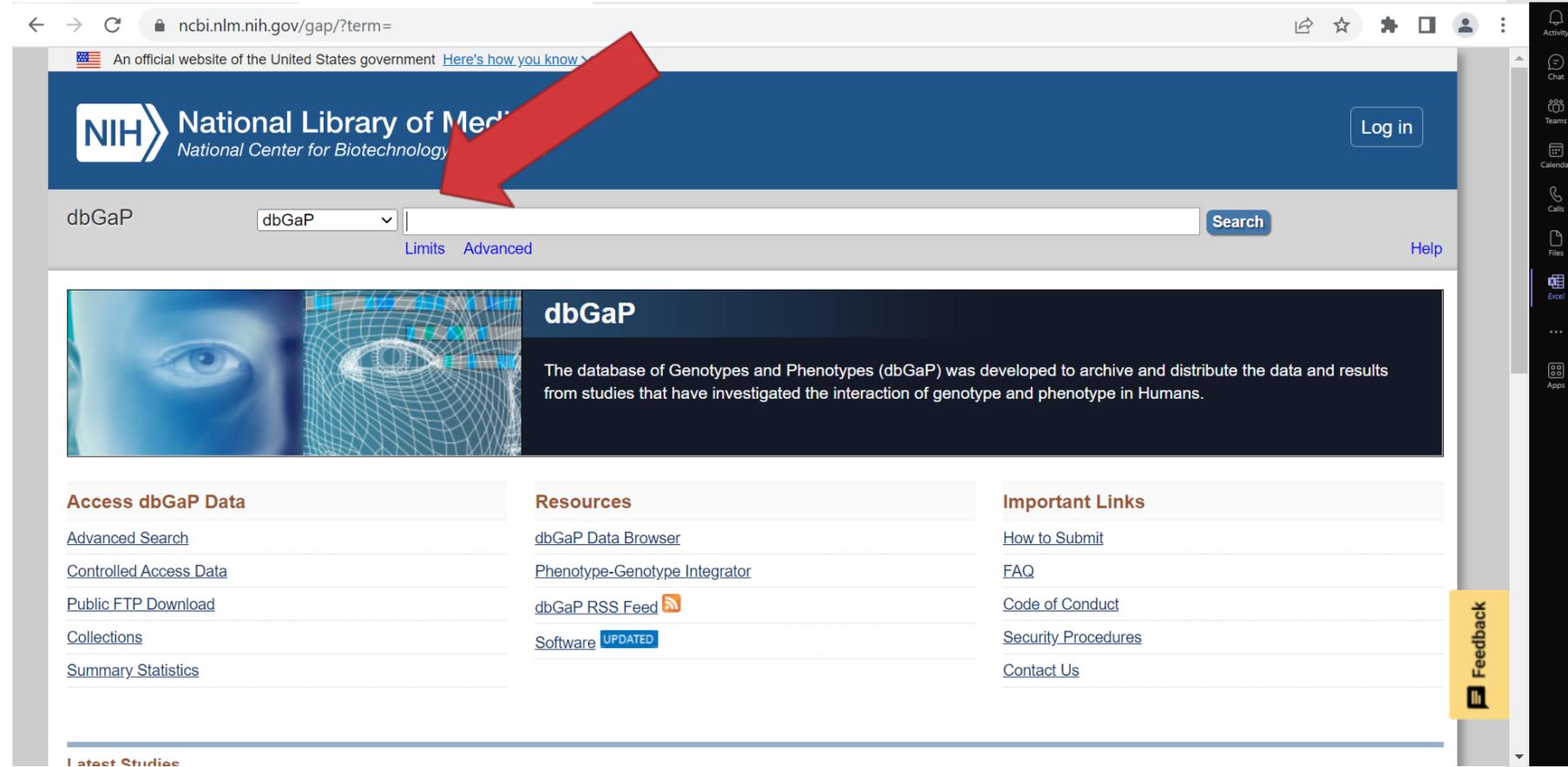


Finding Cancer Genomic Summary Results (GSR) in dbGaP

- 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)



The screenshot shows the dbGaP website homepage. At the top, there is a navigation bar with the NIH logo and the text "National Library of Medicine National Center for Biotechnology". A red arrow points to the search bar, which contains the text "dbGaP" and a dropdown menu. Below the search bar, there are links for "Limits" and "Advanced". The main content area features a large banner with the text "dbGaP The database of Genotypes and Phenotypes (dbGaP) was developed to archive and distribute the data and results from studies that have investigated the interaction of genotype and phenotype in Humans." Below the banner, there are three columns of links: "Access dbGaP Data" (Advanced Search, Controlled Access Data, Public FTP Download, Collections, Summary Statistics), "Resources" (dbGaP Data Browser, Phenotype-Genotype Integrator, dbGaP RSS Feed, Software), and "Important Links" (How to Submit, FAQ, Code of Conduct, Security Procedures, Contact Us). A "Feedback" button is visible in the bottom right corner.

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

The screenshot shows the dbGaP Advanced Search interface. The search term 'phs001868.v1.p1' is entered in the search box. The results are displayed under the 'Analyses' tab, showing one result: 'Genetic Architecture of Susceptibility to Melanoma'. The result details include:

- Accession: phs001868.v1.p1
- Study Disease/Focus: Melanoma
- Study Design: Case-Control
- Study Markerset: HumanOmniExpress-24v1-1_A
- Study Molecular Data Type: SNP Genotypes (Array)
- Study Content: 4 phenotype datasets, 24 variables, 1 analyses, 1 molecular datasets, 11234 subjects, 11234 samples
- NIH Institute: NCI
- Study Consent: GRU --- General research use
- Release Date: 2020-04-24
- Embargo Release Date: 2020-04-24
- Related Terms: Nevus; Cancer of skin pigment cells; MM - malignant melanoma; Malignant Melanoma; Malignant Melanomas; Melanoma, Malignant ...

The abstract text reads: "Most genetic susceptibility to cutaneous melanoma (CM) remains to be discovered. Meta-analysis genome-wide association study (GWAS) of 36,760 melanoma cases (67% newly-genotyped) and 375,188 controls identified 54 significant loci with ... 68 independent SNPs. Analysis of risk estimates across geographical regions and host factors suggests the acral melanoma subtype is uniquely ...".

Navigation tabs at the top include: Studies (1), Phenotype Datasets (4), Variables (19), Molecular Datasets (1), Analyses (1), Documents (0). The current page is 1/1.

Buttons for 'Save Results', 'Save Query', and 'dbGaP FHIR' are visible. A 'Feedback' button is located in the bottom right corner.

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

The screenshot shows the dbGaP website interface. At the top, there is a navigation bar with the NIH logo and the text "National Library of Medicine National Center for Biotechnology Information". A "Log in" button is located in the top right corner. Below the navigation bar, there is a search bar with the text "dbGaP" and a dropdown menu set to "dbGaP". A "Search" button is to the right of the search bar. Below the search bar, there are links for "Limits" and "Advanced".

The main content area features a banner with the text "dbGaP" and a description: "The database of Genotypes and Phenotypes (dbGaP) was developed to archive and distribute the data and results from studies that have investigated the interaction of genotype and phenotype in Humans." Below the banner, there are three columns of links:

- Access dbGaP Data**
 - [Advanced Search](#)
 - [Controlled Access Data](#)
 - [Public FTP Download](#)
 - [Collections](#)
 - [Summary Statistics](#)
- Resources**
 - [dbGaP Data Browser](#)
 - [Phenotype-Genotype Integrator](#)
 - [dbGaP RSS Feed](#)
 - [Software](#) **UPDATED**
- Important Links**
 - [How to Submit](#)
 - [FAQ](#)
 - [Code of Conduct](#)
 - [Security Procedures](#)
 - [Contact Us](#)

A red arrow points to the "Advanced Search" link in the "Access dbGaP Data" section. On the right side of the page, there is a vertical sidebar with various utility icons and a "Feedback" button at the bottom.

Cancer Type: Advanced Search Results

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

ncbi.nlm.nih.gov/gap/advanced_search/?TERM=phs001868.v1.p1

dbGaP Advanced Search lung cancer

lung cancer

Show All Filters

Study (8)

Sort By: Alphabetical

- Framingham Cohort (phs000007.v32.p13) (14)
- GWAS of Lung Cancer Susceptibility in Never-Smoking Women in Asia (phs000716.v1.p1) (3)
- NCI GWAS of Lung Cancer in Never Smokers (phs000634.v1.p1) (4)
- NHLBI Framingham SNP Health Association Resource (SHARe) (phs000342.v20.p13) (14)

Study Disease/Focus (3)

Study Design (2)

Study Molecular Data Type (5)

Study Markerset (7)

NIH Institute (2)

Study Type (6)

Studies (185) | Phenotype Datasets (208) | Variables (454) | Molecular Datasets (17) | **Analyses (47)** | Documents (265) | 1/5

Save Results | Save Query | GDV Link | Remove Selected

- [OncoArray Lung Cancer - Meta-Analysis Of Lung Cancer GWAS](#)
Analysis Accession: phs004930.1
Locus Type: SNP
Trait/Disease: **Lung Cancer**
Trait: **Lung Cancer**
Population Study: Oncoarray Consortium - **Lung Cancer** Studies (phs001273.v3.p2)
We combined imputed genotypes from 14,803 cases and 12,262 controls from the Oncoarray series with 14,463 cases and 44,188 controls samples undertaken by the previous TRICL GWAS(Timofeeva et al, HMG ... 2012). Each study center provided summary statistics from a logistic regression adjusted for age, gender, country (if applicable) and significant...
[Genome Browser](#) [Study page](#) [PheGenI](#) [MeSH](#)
- [OncoArray Lung Cancer - Meta-Analysis Of Lung Adenocarcinoma GWAS](#)
Analysis Accession: phs004929.1
Locus Type: SNP
Trait/Disease: **Lung Cancer**
Trait: **Lung Cancer**
Population Study: Oncoarray Consortium - **Lung Cancer** Studies (phs001273.v3.p2)
We combined imputed genotypes from 6,411 **lung** adenocarcinoma patients and 12,262 controls from the Oncoarray series with 4,862 **lung** adenocarcinoma patients and 43,221 controls samples undertaken by the previous TRICL ... GWAS(Timofeeva et al, HMG 2012). Each study center provided summary statistics from a logistic regression adjusted for age, gender, country (if...
[Genome Browser](#) [Study page](#) [PheGenI](#) [MeSH](#)
- [OncoArray Lung Cancer - Meta-Analysis Of Lung Small Cell Carcinoma GWAS](#)
Analysis Accession: phs004927.1
Locus Type: SNP
Trait/Disease: **Lung Cancer**
Trait: **Lung Cancer**
Population Study: Oncoarray Consortium - **Lung Cancer** Studies (phs001273.v3.p2)

Feedback

Navigating the dbGaP Study Page

Look for the "Analyses" tab

ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001868.v1.p1

dbGaP
GENOTYPES and PHENOTYPES

Cutaneous Melanoma GWAS Combining Multiple Populations and Risk Phenotypes
dbGaP Study Accession: phs001868.v1.p1

Request Access

▸ [Study version history](#)

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

Jump to: [Authorized Access](#) | [Attribution](#) | [Authorized Requests](#) [HHS Vulnerability Disclosure](#)

Study Description

Most genetic susceptibility to cutaneous melanoma (CM) remains to be discovered. Meta-analysis genome-wide association study (GWAS) of 36,760 melanoma cases (67% newly-genotyped) and 375,188 controls identified 54 significant loci with 68 independent SNPs. Analysis of risk estimates across geographical regions and host factors suggests the acral melanoma subtype is uniquely unrelated to pigmentation. Combining this meta-analysis with nevus count and hair colour GWAS, and transcriptome association approaches, uncovered 31 potential secondary loci, for a total of 85 CM susceptibility loci. These findings provide substantial insights into CM genetic architecture, reinforcing the importance of

Important Links and Information

- Request access via [Authorized Access](#)
 - [Instructions](#) for requestors
 - [Data Use Certification \(DUC\) Agreement](#)
- [Talking Glossary of Genetic Terms](#)



List of Analyses for the Dataset

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



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

- [Browse all analyses within this study via Advanced Search](#)
- [List all analyses within this study](#)
- Summary of available analyses from this study is freely available from the [dbGaP public ftp site](#) (nb. these summary files do not require Authorized Access approval).

Analysis Name and Accession

Analysis Name: A Meta-Analysis Of Genome-Wide Association Study On Cutaneous Melanoma
Analysis Accession: pha004971.1

[View association results in Genome Browser](#)

Analysis Description

We performed a genome-wide association analysis (GWAS) meta-analysis of cutaneous melanoma susceptibility with 30,134 clinically cases and 81415 CM-free controls from the United Kingdom, United States, Australia, Northern and Western Europe as well as the Mediterranean. Samples were genotyped using different SNP array genotyping platforms .

Analysis Methods

QCed using the same criteria (SNPs: minor allele frequency > 0.01, Hardy-Weinberg Equilibrium P-value > 5 x 10⁻⁴ in controls and < 5 x 10⁻¹⁰ in cases. Samples: missing < 3% of variants, heterozygosity values between -0.05 and 0.05 and within 3 sd from the mean, genetically-predicted sex matched recorded sex, European based on principal component analysis, no relatives with identity by descent (IBD) pi_{hat} > 0.15). Imputation was performed using 1000 Genomes Project phase 1 v3 or Michigan Imputation Server with the Haplotype Reference Consortium panel (HRC version 1). As rare SNPs where one allele is missing in the case or control group can lead to very large (or infinite) OR estimates, variants with an OR < 1 x 10⁻⁴ (the minimum reported by PLINK) or > 1 x 10⁶ were excluded. Fixed effect meta-analysis was performed using PLINK.

Analysis Plots

The following plots were generated by dbGaP based on the data that was submitted and are not necessarily from any

HHS Analysis

- Genetic Architecture of Susceptibility to Melanoma
- Analysis**

[Vulnerability Disclosure](#)

Feedback



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

ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001868.v1.p1

... these findings provide substantial insights into CM genetic architecture, reinforcing the importance of neovogenesis, pigmentation, and telomere maintenance together with identifying potential new pathways for CM pathogenesis.

- Study Design:
 - Case-Control
- Study Type:
 - Case-Control
- dbGaP estimated [ancestry](#) using [GRAF-pop](#)
- Total number of consented subjects: 11234

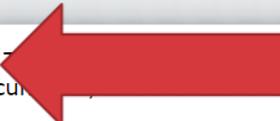
Authorized Access

- **Data access provided by:** [dbGaP Authorized Access](#)
- **Release Date:** April 24, 2020
- **Embargo Release Date:** April 24, 2020
- [Data Use Certification Requirements \(DUC\)](#)
- **Public Posting of [Genomic Summary Results](#):** Allowed
- **Use Restrictions**

Consent group	Is IRB required?	Data Access Committee	Number of participants
General Research Use 	No	NCI DAC (NCIDAC@mail.nih.gov)	11234

- [List of components](#) downloadable from [Authorized Access](#)

Publicly Available Data (Public ftp)

Connect to the [public download site](#).  notes and manifests. The site also contains data dictionaries, variable summaries, documents, and analyses, whenever available.

Molecular Data

Type	Source	Platform	Number of Oligos/SNPs	SNP Batch Id	Comment
------	--------	----------	-----------------------	--------------	---------

Public FTP Site Breakdown

← → ↻ 🔒 <https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs001868/phs001868.v1.p1/>

Index of /dbgap/studies/phs001868/phs001868.v1.p1

Name	Last modified	Size
Parent Directory		-
manifest/	2020-07-09 05:48	-
pheno_variable_summaries/	2020-07-09 05:48	-
release_notes/	2020-07-09 05:48	-
GapExchange_phs001868.v1.p1.xml	2020-07-09 05:48	4.6K
dbGaPEx2.1.5.xsd	2012-11-15 11:08	59K

[HHS Vulnerability Disclosure](#)

- 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)

← → ↻ <https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs001868/>

Index of /dbgap/studies/phs001868

Name	Last modified	Size
Parent Directory		-
analyses/	2020-07-09 13:25	-
phs001868.v1.p1/	2020-04-24 06:18	-

[HHS Vulnerability Disclosure](#)



← → ↻ <https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs001868/analyses/>

Index of /dbgap/studies/phs001868/analyses

Name	Last modified	Size
Parent Directory		-
phs001868.pha004971.txt	2020-06-25 08:52	1.1M

[HHS Vulnerability Disclosure](#)

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>