Anonymized Genetic Data Available with Data Use Agreement

Access to the anonymized data sets does not require IRB approval because they are not linkable to any other data sets. These data sets contain a fixed number of variables which are outlined below each of the data sets. For access to any of these data sets, please contact us at NHANESGenetics@CDC.gov to obtain a release form.

Please scroll down or click the link in the following dbSNP IDs to view the genetic data set summaries and HWE results

Gene Name	dbSNP ID
HFE	rs1799945
HFE	rs1800562

Summary of Table Terms:

HUGO Gene ID:

For each known human gene the HUGO Gene Nomenclature Committee approve a gene name and symbol (short-form abbreviation). All approved symbols are stored in Genew, the Human Gene Nomenclature Database. Each symbol is unique and the HGNC ensure that each gene is only given one approved gene symbol.

dbSNP ID/CDC ID:

This identifier is used to refer to the specific genetic variation that was tested in the data set. In most cases, the genetic variation tested is a single nucleotide polymorphism (SNP). SNPs are genetic variations that involve a single base change in the genetic code and are generally defined as occurring at greater than 1% prevalence. In some cases, the genetic variation tested is not classified as a SNP because it is rare (i.e. <1%) or is a genetic variation such as a large insertion or deletion.

There are three possible prefixes within this category:
rs – dbSNP identifier that is a prefix to a reference variation.
ss – dbSNP identifier that is a prefix to a submitted variation.
cs – CDC identifier that is a prefix to those genetic variations without a dbSNP identifier.

dbSNP identifiers are found in the NCBI's database and contain the prefixes 'rs' or 'ss'. An identifier that begins with an 'rs' or 'ss' will contain a link to the NCBI SNP database where you can find additional information concerning that genetic variation. If the genetic variation has not been assigned a dbSNP ID, a CDC ID is assigned with the prefix 'cs'. The CDC ID is a temporary identifier that the CDC will update with the appropriate dbSNP ID if it becomes available through the NCBI.

Variation Type:

Types of genetic variations:

SNP: Single nucleotide polymorphism
Possible SNP: Polymorphism lacking the 1% frequency to define it as a SNP
Insertion: Genetic insertion
Deletion: Genetic deletion

CDC Data Submission ID:

Unique identifier assigned by the CDC for each data set submitted.

Duplicate Data:

This field denotes the availability of other data sets submitted by other labs for the same genetic variation. In the event that multiple labs test the same genetic variation, this field will contain the CDC data submission identifiers of the duplicate data sets. It will be up to the researcher to determine which of the data sets they will use. Quality control summaries may be provided to assist the researcher in choosing one duplicate data set over another.

Race-Ethnicity:

This analytic variable, based on the NHANES III survey design, was derived from many sources of data and is based on reported race and ethnicity. The 'other' category includes all Hispanics, regardless of race, who were not Mexican-American and also includes all non-Hispanics from racial groups other than white or black.

Genetic Variations:

A = Adenine
C = Cytosine
T = Thymine
G = Guanine
N = No result reported
Ins = Insertion
Del = Deletion
Wt = Wild type (Used in some cases where an insertion or deletion is reported)

Frequency:

(n) - Number of participants in that race-ethnicity with a particular genotype
(%) - Percent of all participants in that race-ethnicity with that genotype (includes null results)

HWE Pr > Chi Sq:

These are the results of Hardy-Weinberg equilibrium (HWE) analysis using the PROC ALLELE procedure from the software package SAS/Genetics. Chi-square goodness-of-fit results are reported.

Gene Name: Hemochromatosis					
HUGO Gene ID	dbSNP ID	Variation Type	CDC Data Submission ID		Duplicate Data
HFE	rs1799945	SNP	ANON1		None
Variation Sequence: agctgttcgtgttctatgat [C/G] atgagagtcgccgtgtggag					
Race-Ethnicity		Genetic Variation	Frequency (n)	Frequency (%)	HWE Pr > Chi Sq
Non-Hispanic Whites		C/C	1444	0.549	0.9548
		C/G	524	0.199	
		G/G	48	0.018	
		N/N	0	0.000	
Non-Hispanic Blacks		C/C	1504	0.713	0.0053
		C/G	91	0.043	
		G/G	5	0.002	
		N/N	0	0.000	
Mexican-Americans		C/C	1215	0.586	<.0001
		C/G	326	0.157	
		G/G	14	0.007	
		N/N	2	0.001	
Other		C/C	NA	NA	NA
		C/G	NA	NA	
		G/G	NA	NA	
		N/N	NA	NA	

Variables that are linked these data:

Variable	Label
RANDOMID	Random Identifier
HSSEX	Sex
HSAGEGR	10 Year age groups
DMARETHN	Race-ethnicity
WEIGHT	Sampling weight
RTS10	Deciles for Transferrin Saturation
H63D	rs1799945

Gene Name: Hemochromatosis					
HUGO Gene ID	dbSNP ID	Variation Type	CDC Data Submission ID		Duplicate Data
HFE	rs1800562	SNP	ANON2		None
Variation Sequence: ggaagagcagagatatacgt [A/G] ccaggtggagcacccaggcc					
Race-Ethnicity		Genetic Variation	Frequency (n)	Frequency (%)	HWE Pr > Chi Sq
Non-Hispanic Whites		A/A	1765	0.671	0.4136
		A/G	245	0.093	
		G/G	6	0.002	
		N/N	0	0.000	
Non-Hispanic Blacks		A/A	1560	0.740	0.1450
		A/G	39	0.019	
		G/G	1	0.000	
		N/N	0	0.000	
Mexican-Americans		A/A	1507	0.727	<.0001
		A/G	47	0.023	
		G/G	1	0.000	
		N/N	2	0.001	
Other		A/A	NA	NA	NA
		A/G	NA	NA	
		G/G	NA	NA	
		N/N	NA	NA	

Variables that are linked these data:

Variable	Label
RANDOMID	Random Identifier
HSSEX	Sex
HSAGEGR	10 Year age groups
DMARETHN	Race-ethnicity
WEIGHT	Sampling weight
RTS10	Deciles for Transferrin Saturation
C282Y	rs1800562