Restricted Genetic Data Available for Analysis at the NCHS

The data sets summarized below are available for analysis in the Research Data Center at the National Center for Health Statistics in Hyattsville, MD. Prior IRB approval is required before these data sets can be accessed. For more information on how to apply for access, please see our 'Guidelines for secondary data analysis of NHANES III genetic data.'

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

List of Available Genetic Variation Data Sets

Gene Name	dbSNP ID
ApoE	rs7412
ApoE	rs429358
CRP	cs1*
CRP	cs2*
CRP	rs1205
CRP	rs1417938
CRP	rs2808630
CRP	rs3093066
CRP	rs3093058
CRP	rs3091244
FABP2	rs1799883
PPARG	rs1801282
***cs IDs are variations that have not been assigned a dbSNP ID.**	

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 that they will use. Quality control summaries may be provided to assist the researcher in choosing one duplicate data set over another.

Race-Ethnicity:

This key 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 Variation:

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: Apolipoprotein E				
HUGO Gene ID	dbSNP ID	Variation Type	CDC Data Submission ID	Duplicate Data
APOE	rs7412	SNP	S0021741201	None
Variation Sequence: atgccgatgacctgcagaag [C/T] gcctggcagtgtaccaggcc				

Race-Ethnicity	Genetic Variation	Frequency (n)	Frequency (%)	HWE Pr > Chi Sq
Non-Hispanic Whites	C/C	2180	0.829	0.9904
	C/T	417	0.159	
	T/T	20	0.008	
	N/N	13	0.005	
Non-Hispanic Blacks	C/C	1694	0.804	0.9058
	C/T	381	0.181	
	T/T	22	0.010	
	N/N	11	0.005	
Mexican-Americans	C/C	1925	0.929	0.731
	C/T	131	0.063	
	T/T	1	0.000	
	N/N	16	0.008	
Other	C/C	307	0.882	0.3186
	C/T	35	0.101	
	T/T	0	0.000	
	N/N	6	0.017	

Gene Name:				
HUGO Gene ID	dbSNP ID	Variation Type	CDC Data Submission ID	Duplicate Data
APOE	rs429358	SNP	S002142935801	None
Variation Sequence: gcgcggacatggaggacgtg [C/T] gcggccgcctggtgcagtac				

Race-Ethnicity	Genetic Variation	Frequency (n)	Frequency (%)	HWE Pr > Chi Sq
Non-Hispanic Whites	C/C	58	0.022	0.7212
	C/T	677	0.257	
	T/T	1870	0.711	
	N/N	25	0.010	
Non-Hispanic Blacks	C/C	103	0.049	0.8942
	C/T	717	0.340	
	T/T	1269	0.602	
	N/N	19	0.009	
Mexican-Americans	C/C	33	0.016	0.0135
	C/T	365	0.176	
	T/T	1666	0.804	
	N/N	9	0.004	
Other	C/C	2	0.006	0.372
	C/T	65	0.187	
	T/T	273	0.784	
	N/N	8	0.023	

Gene Name: C-reactive Protein, Pentraxin-Related				
HUGO Gene ID	CDC SNP ID	Variation Type	CDC Data Submission ID	Duplicate Data
CRP	cs1	Possible SNP	S0013101	None
Variation Sequence: cttcggtgggaactttgaag [G/A] aagccagtccctggtgggag				

Race-Ethnicity	Genetic Variation	Frequency (n)	Frequency (%)	HWE Pr > Chi Sq
Non-Hispanic Whites	A/A	0	0.000	NA
	A/G	0	0.000	
	G/G	2564	0.975	
	N/N	66	0.025	
Non-Hispanic Blacks	A/A	0	0.000	0.9912
	A/G	1	0.000	
	G/G	2042	0.969	
	N/N	65	0.031	
Mexican-Americans	A/A	0	0.000	0.9911
	A/G	1	0.000	
	G/G	2031	0.980	
	N/N	41	0.020	
Other	A/A	0	0.000	NA
	A/G	0	0.000	
	G/G	336	0.966	
	N/N	12	0.034	

Gene Name: C-reactive Protein, Pentraxin-Related				
HUGO Gene ID	dbSNP ID	Variation Type	CDC Data Submission ID	Duplicate Data
CRP	cs2	Possible SNP	S0013201	None
Variation Sequence: gttctgggtagatgggaagc [C/T] cagggtgaggaagagtctga				

Race-Ethnicity	Genetic Variation	Frequency (n)	Frequency (%)	HWE Pr > Chi Sq
Non-Hispanic Whites	C/C	2564	0.975	NA
	C/T	0	0.000	
	T/T	0	0.000	
	N/N	66	0.025	
Non-Hispanic Blacks	C/C	2033	0.964	NA
	C/T	0	0.000	
	T/T	0	0.000	
	N/N	75	0.036	
Mexican-Americans	C/C	2012	0.971	NA
	C/T	0	0.000	
	T/T	0	0.000	
	N/N	61	0.029	
Other	C/C	0	0.000	NA
	C/T	0	0.000	
	T/T	336	0.966	
	N/N	12	0.034	

Gene Name: C-reactive Protein, Pentraxin-Related				
HUGO Gene ID	CDC SNP ID	Variation Type	CDC Data Submission ID	Duplicate Data
CRP	rs1205	SNP	S0011120501	None
Variation Sequence: ttgccacatggagagagact [A/G] tgaggacagaagccaaactg				

Race-Ethnicity	Genetic Variation	Frequency (n)	Frequency (%)	HWE Pr > Chi Sq
Non-Hispanic Whites	A/A	280	0.106	0.0543
	A/G	1178	0.448	
	G/G	1044	0.397	
	N/N	128	0.049	
Non-Hispanic Blacks	A/A	109	0.052	0.0084
	A/G	627	0.297	
	G/G	1262	0.599	
	N/N	110	0.052	
Mexican-Americans	A/A	245	0.118	0.6696
	A/G	916	0.442	
	G/G	821	0.396	
	N/N	91	0.044	
Other	A/A	47	0.135	0.2796
	A/G	144	0.414	
	G/G	142	0.408	
	N/N	15	0.043	

Gene Name: C-reactive Protein, Pentraxin-Related				
HUGO Gene ID	dbSNP ID	Variation Type	CDC Data Submission ID	Duplicate Data
CRP	rs1417938	SNP	S0011141793801	None
Variation Sequence: cccccatacctcagatcaaa [A/T] ctctcccatagcctggggtg				

Race-Ethnicity	Genetic Variation	Frequency (n)	Frequency (%)	HWE Pr > Chi Sq
Non-Hispanic Whites	A/A	1246	0.474	0.6401
	A/T	1085	0.413	
	T/T	226	0.086	
	N/N	73	0.028	
Non-Hispanic Blacks	A/A	1586	0.752	0.6151
	A/T	437	0.207	
	T/T	27	0.013	
	N/N	58	0.028	
Mexican-Americans	A/A	848	0.409	0.4003
	A/T	919	0.443	
	T/T	229	0.110	
	N/N	77	0.037	
Other	A/A	201	0.578	0.1995
	A/T	113	0.325	
	T/T	23	0.066	
	N/N	11	0.032	

Gene Name: C-reactive Protein, Pentraxin-Related				
HUGO Gene ID	dbSNP ID	Variation Type	CDC Data Submission ID	Duplicate Data
CRP	rs1800947	SNP	S0011180094701	None

Variation Sequence: aacatgtgggactttgtgct [C/G] tcaccagatgagattaacac

Race-Ethnicity	Genetic Variation	Frequency (n)	Frequency (%)	HWE Pr > Chi Sq
Non-Hispanic Whites	C/C	1	0.000	0.4932
	C/G	134	0.051	
	G/G	2269	0.863	
	N/N	226	0.086	
Non-Hispanic Blacks	C/C	0	0.000	0.7672
	C/G	26	0.012	
	G/G	1929	0.915	
	N/N	153	0.073	
Mexican-Americans	C/C	0	0.000	0.3663
	C/G	78	0.038	
	G/G	1863	0.899	
	N/N	132	0.064	
Other	C/C	0	0.000	0.7312
	C/G	12	0.034	
	G/G	305	0.876	
	N/N	31	0.089	

Gene Name: C-reactive Protein, Pentraxin-Related				
HUGO Gene ID	dbSNP ID	Variation Type	CDC Data Submission ID	Duplicate Data
CRP	rs2808630	SNP	S0011280863001	None
Variation Sequence: ttgcttgcatcttactatac [A/G] tagtctggtagacagcctct				

Race-Ethnicity	Genetic Variation	Frequency (n)	Frequency (%)	HWE Pr > Chi Sq
Non-Hispanic Whites	A/A	1222	0.465	0.8422
	A/G	972	0.370	
	G/G	197	0.075	
	N/N	239	0.091	
Non-Hispanic Blacks	A/A	1422	0.675	0.1307
	A/G	553	0.262	
	G/G	41	0.019	
	N/N	92	0.044	
Mexican-Americans	A/A	951	0.459	0.6298
	A/G	501	0.242	
	G/G	71	0.034	
	N/N	550	0.265	
Other	A/A	195	0.560	0.2705
	A/G	113	0.325	
	G/G	11	0.032	
	N/N	29	0.083	

Gene Name: C-reactive Protein, Pentraxin-Related				
HUGO Gene ID	dbSNP ID	Variation Type	CDC Data Submission ID	Duplicate Data
CRP	rs3091244	SNP	S0011309124401	None

Variation Sequence: cgtgcacccagatggccact [A/C/T] gtttaatatgttaccatttc

Race-Ethnicity	Genetic Variation	Frequency (n)	Frequency (%)	HWE Pr > Chi Sq
Non-Hispanic Whites	A/A	9	0.003	0.4586
	A/C	219	0.083	
	A/T	85	0.032	
	C/C	1031	0.392	
	C/T	994	0.378	
	T/T	236	0.090	
	N/N	56	0.021	
Non-Hispanic Blacks	A/A	163	0.077	0.3302
	A/C	466	0.221	
	A/T	321	0.152	
	C/C	419	0.199	
	C/T	508	0.241	
	T/T	164	0.078	
	N/N	67	0.032	
Mexican-Americans	A/A	5	0.002	0.3634
	A/C	129	0.062	
	A/T	60	0.029	
	C/C	699	0.337	
	C/T	861	0.415	
	T/T	263	0.127	
	N/N	56	0.027	
Other	A/A	6	0.017	0.1459
	A/C	47	0.135	
	A/T	14	0.040	
	C/C	127	0.365	
	C/T	112	0.322	
	T/T	33	0.095	
	N/N	9	0.026	

Gene Name: C-reactive Protein, Pentraxin-Related				
HUGO Gene ID	dbSNP ID	Variation Type	CDC Data Submission ID	Duplicate Data
CRP	rs3093058	SNP	S0011309305801	None
Variation Sequence: atcttgtgaatagaggaaag [A/T] gtagaatcagattatcctga				

Race-Ethnicity	Genetic Variation	Frequency (n)	Frequency (%)	HWE Pr > Chi Sq
Non-Hispanic Whites	A/A	2539	0.965	0.9131
	A/T	11	0.004	
	T/T	0	0.000	
	N/N	80	0.030	
Non-Hispanic Blacks	A/A	1433	0.680	0.8857
	A/T	555	0.263	
	T/T	55	0.026	
	N/N	65	0.031	
Mexican-Americans	A/A	1974	0.952	0.3361
	A/T	55	0.027	
	T/T	1	0.000	
	N/N	43	0.021	
Other	A/A	299	0.859	0.0002
	A/T	25	0.072	
	T/T	4	0.011	
	N/N	20	0.057	

Gene Name: C-reactive Protein, Pentraxin-Related				
HUGO Gene ID	dbSNP ID	Variation Type	CDC Data Submission ID	Duplicate Data
CRP	rs3093066	SNP	S0011309306601	None
Variation Sequence: ctgggagctcgttaactatg [A/C] tgggaaacggtccaaaagaa				

Race-Ethnicity	Genetic Variation	Frequency (n)	Frequency (%)	HWE Pr > Chi Sq
Non-Hispanic Whites	A/A	0	0.000	0.7896
	A/C	26	0.010	
	C/C	2374	0.903	
	N/N	230	0.087	
Non-Hispanic Blacks	A/A	107	0.051	0.7217
	A/C	691	0.328	
	C/C	1167	0.554	
	N/N	143	0.068	
Mexican-Americans	A/A	1	0.000	0.4879
	A/C	62	0.030	
	C/C	1938	0.935	
	N/N	72	0.035	
Other	A/A	2	0.006	0.2712
	A/C	32	0.092	
	C/C	297	0.853	
	N/N	17	0.049	

Gene Name: Fatty Acid Binding Protein 2, Intestinal				
HUGO Gene ID	dbSNP ID	Variation Type	CDC Data Submission ID	Duplicate Data
FABP2	rs1799883	SNP	S0021179988301	None
Variation Sequence: tcacagtcaaagaatcaagc [A/G] cttttcgaaacattgaagtt				

Race-Ethnicity	Genetic Variation	Frequency (n)	Frequency (%)	HWE Pr > Chi Sq
Non-Hispanic Whites	C/C	1414	0.538	0.1174
	C/T	1032	0.392	
	T/T	160	0.061	
	N/N	24	0.009	
Non-Hispanic Blacks	C/C	1240	0.588	0.7614
	C/T	739	0.351	
	T/T	106	0.050	
	N/N	23	0.011	
Mexican-Americans	C/C	1107	0.534	0.7572
	C/T	807	0.389	
	T/T	142	0.068	
	N/N	17	0.008	
Other	C/C	186	0.534	0.9919
	C/T	131	0.376	
	T/T	23	0.066	
	N/N	8	0.023	

Gene Name: Peroxisome Proliferative Activated Receptor, Gamma					
HUGO Gene ID	dbSNP ID	Variation Type	CDC Data Submission ID		Duplicate Data
PPARG	rs1801282	SNP	S0021180128202		None
Variation Sequence: tgggagattctcctattgac [C/G] cagaaagcgattccttcact					
Race-Ethnicity		Genetic Variation	Frequency (n)	Frequency (%)	HWE Pr > Chi Sq
Non-Hispanic Whites		C/C	1973	0.750	0.8197
		C/G	601	0.229	
		G/G	44	0.017	
		N/N	12	0.005	
Non-Hispanic Blacks		C/C	2001	0.949	0.9395
		C/G	93	0.044	
		G/G	1	0.000	
		N/N	13	0.006	
Mexican-Americans		C/C	1575	0.760	0.9993
		C/G	456	0.220	
		G/G	33	0.016	
		N/N	9	0.004	
Other		C/C	292	0.839	0.4739
		C/G	47	0.135	
		G/G	3	0.009	
		N/N	6	0.017	