Health and Retirement Study:
APOE and Serotonin Transporter Alleles – Early Release
August, 2021

Jessica Faul, Stacey Collins, Jennifer Smith, Wei Zhao, Sharon Kardia and David Weir
Health and Retirement Study, University of Michigan, Ann Arbor, MI

Table of Contents

Description of File

This data package contains data for the APOE isoform, directly genotyped where available and imputed otherwise, as well as human serotonin transporter (5HTTLPR) short and long alleles measured using polymerase chain reaction (PCR). Details on the lab procedures and data contents are described below.

Sample

HRS participants who consented and completed salivary DNA collection in 2006 (Phase 1), 2008 (Phase 2), 2010 (Phase 3), or 2012 (Phase 4) were eligible for inclusion. Collection procedures are described in detail elsewhere. In total, there are 19,193 HRS participants in the resulting data file: 17,237 with directly genotyped data for APOE and 1,956 additional participants with imputed data. There are 17,364 participants with valid values for 5HTTLPR.

APOE

The APOE protein has three major isoforms (ε2, ε3, and ε4). These isoforms are formed from two SNPs, rs7412 and rs429358. The resulting APOE isoform is inferred according to the following algorithm:

rs7412 genotype	rs429358 genotype	APOE genotype
T/T	T/T	ε2/ε2
C/T	T/T	ε2/ε3
C/T	C/T	ε2/ε4
C/C	T/T	ε3/ε3
C/C	C/T	ε3/ε4
C/C	C/C	ε4/ε4

Genotyping Processes

The DNA samples were genotyped and analyzed at the Center for Inherited Disease Research (CIDR) Genetic Resources Core Facility (GRCF) and Fragment Analysis Facility (FAF) at Johns Hopkins University (https://cidr.jhmi.edu).

TaqMan Genotyping

TaqMan allelic discrimination assays consists of two PCR primers that flank the single nucleotide polymorphism (SNP) of interest and two minor groove binder (MGB) probes 5' labeled with fluorescent molecules VIC and FAM that are specific for wild type and variant alleles. Each probe has a non-fluorescent quencher molecule (NFQ) at the 3' end to provide a greater signal-to-noise ratio. Each reporter probe anneals specifically to complimentary sequence in the PCR fragment and due to the proximity of the NFQ molecule to the reporter dye fluorescence is suppressed. As amplification precedes, the exonuclease activity of AmpliTaq Gold DNA polymerase cleaves the probes that are hybridized to complimentary target sequence, separating the reporter dye and the NFQ molecule, resulting in increased fluorescence of the reporter. Probes that are mismatched to the target sequence

have a low binding efficiency and are more readily displaced by the AmpliTaq Gold DNA polymerase without cleavage. The fluorescent signal generated by PCR amplification is then collected on a 7900HT PCR System and quantified using SDS software (Applied Biosystems, Foster City, CA).

APOE TaqMan

Genotyping is performed using predesigned TaqMan allelic discrimination SNP assays C___3084793 (rs429358) and C____904973 (rs7412) to determine APOE genotypes. Each SNP assay kit consists of two PCR primers that flank the SNP of interest and two fluorogenic allele-specific minor groove binder (MGB) probes, specific for wild type and variant alleles, with a non-fluorescent quencher molecule (NFQ) at the 3' end of each probe. The fluorescent signal generated by PCR amplification is then collected on a 7900HT Sequence Detection System and quantified using SDS v2.1software (Applied Biosystems, Foster City, CA). The GRCF follows the manufacturers supplied protocols. PCR amplification reactions are performed in a 5 µl volume containing 0. 25µl (20X assay stock) primer/probe mix, 2.5 µl 2X Genotyping Master Mix, and 20 ng genomic DNA. DNA is incubated at 95 °C for 10 min, followed by 40 cycles at 95 °C for 15 seconds and 60 °C for 1 min. Allele-specific fluorescence measurement is distinguished by measuring endpoint 6-FAM or VIC fluorescence intensities normalized to ROX signal for each well. Genotypes are determined using TaqMan Genotyper software.

Quality Control Procedures

Duplicates are included for analyses so that a genotyping/measurement reproducibility can be examined during processing. Duplicate samples were distributed across all plates so that each duplicate is on a different plate than its paired sample. Each duplicate pair is used only once.

CIDR FAF Supplied Controls: The FAF reserves 3 wells on each 96-well plate for controls that are added to the plate after it arrives at the FAF. As part of the QC process, control sample genotypes are compared to previously-determined genotypes confirmed by Sanger sequencing or Hapmap-genotypes when available.

Plates with discordant duplicate sample genotyping calls are examined in detail for plating or identity issues. Samples/plates that can't be resolved are removed from the data release.

APOE Imputation

For participants for whom direct genotyping was not available, either because there was insufficient sample or the sample did not pass QC (where one or both of the SNPs that determine the APOE isoform failed), APOE status was imputed from previously genotyped array data.

All of the HRS genotype datasets were genotyped on the Illumina HumanOmni2.5 array: Phases 1-2 on HumanOmni2.5-4v1, Phase 3 on HumanOmni2.5-8v1, and Phase 4 on HumanOmni2.5-8v1.1. SNPs retained in the combined dataset for imputation were those common to all versions of the array and with matching rsID, chromosome, and position in both array annotations. SNPs with any discordant calls between phases in control samples were excluded from imputation. Imputation to the 1000 Genomes Cosmopolitan Reference Panel (phase 3) was performed. Dosage data for rs7412 and rs429358 was used

to assign the “best guess genotypes” for each respondent. Then, we used the “best guess genotypes” to infer the APOE isoform. A flag is included in the data file to indicate where the APOE call is imputed rather than measured directly.

Serotonin Transporter – 5-HTTLPR Long and Short Alleles

5-HTTLPR (serotonin-transporter-linked promoter region) is a degenerate repeat polymorphic region in SLC6A4, the gene that codes for the serotonin transporter. Researchers commonly refer to two variations in humans: short ("S") and a long ("L") alleles. The short allele is associated with reduced transcriptional activity of the serotonin transporter (often termed 'low-activity') compared to the long allele.

Lab Processes

The human serotonin transporter (5HTTLPR) short and long alleles were PCR amplified using forward primer 5’ TCTCCCGCCTGGCGTTGC-3’ and reverse primer 5’-GCCGGTTGGGCTGAGCGTCT-3’. PCR was performed in a 10µl reaction consisting of 0.4 µM primers, 0.15 µM 7-deaza dGTP, 1X MasterAmp™ 2X PCR PreMix K, (Epicenter Biotechnologies, Madison, WI), 1 unit Taq DNA polymerase (Applied Biosystems, Foster City, CA) and 40 ng DNA. Amplification was carried out in a Thermo Hybaid MBS 0.2S thermocycler (Thermo Electron Corporation, Waltham, MA) for 4 min at 95C, followed by 35 cycles of 94C for 30 s, 62.8C for 30 s and 72C for 30 s, followed by a final extension step of 72C for 10 min. PCR products were resolved on 2% NuSieve (FMC BioProducts, Vallensbaek, Denmark), 1% agarose (Invitrogen, Carlsbad, CA) gels and visualized by ethidium bromide staining. The short and long allele are characterized by 515 bp and 560 bp products, respectively.

Quality Control Procedures

Duplicates are included for analyses so that a genotyping/measurement reproducibility can be examined during processing. Duplicate samples were distributed across all plates so that each duplicate is on a different plate than its paired sample. Each duplicate pair is used only once.

CIDR FAF Supplied Controls: The FAF reserves 3 wells on each 96-well plate for controls that are added to the plate after it arrives at the FAF. As part of the QC process, control sample genotypes are compared to previously-determined genotypes confirmed by Sanger sequencing or Hapmap-genotypes when available.

Plates with discordant duplicate sample genotyping calls are examined in detail for plating or identity issues. Samples/plates that can’t be resolved are removed from the data release.

The L represent the long allele and the S represents the short allele. Any sample with “undetermined” listed failed PCR. Any sample that failed was repeated, so if “undetermined” is listed it means the sample failed twice for that assay.

Data File Contents

Column Name	Description	Type of Field
HHID	Subject household ID	Character
PN	Subject individual ID	Character
rs7412	Measured or best guess genotype for rs7412	C/C, C/T, or T/T
rs7412_dosage	dosage for rs7412	Numeric (ranging from 0 to 2)
rs7412_posterior_probability	posterior probability for rs7412 (a measure of SNP imputation quality)	Percentage (ranging from 0 to 1)
rs429358	Measured or best guess genotype for rs429358	C/C, T/C, or T/T
rs429358_dosage	dosage for rs429358	Numeric (ranging from 0 to 2)
rs429358_posterior_probability	posterior probability for rs429358 (a measure of SNP imputation quality)	Percentage (ranging from 0 to 1)
APOE	Measured or best guess isoforms of APOE	e2/e2, e2/e3, e2/e4, e3/e3, e3/e4, e4/e4
APOE_Imputed	Flag for whether APOE isoforms are imputed	0=APOE Measured; 1=APOE Imputed
HTTLPR Call	5HTTLPR short or long allele; L=long allele, S=short allele, V=variant	LL / LS / LV / SS

User Recommendations

Imputed results are provided as best guess genotypes, dosage, and the probability of each of the three genotypes for each genetic marker. We recommend incorporating the imputed probabilities into any downstream analysis, when imputed data are used. Quality metrics are provided and can be used to filter imputation results on a per-variant basis. For the APOE data, we recommend that users exclude subjects with imputed data with a posterior probability <0.8 for either rs7412 or rs429358.

If You Need to Know More

This document is intended to serve as a brief overview to the APOE and Serotonin Transporter Alleles data product. If you have questions or concerns that are not adequately covered here or on our Web site, or if you have any comments, please contact us. We will do our best to provide answers.

HRS Internet Site

Health and Retirement Study public release data and additional information about the study are available on HRS web site. To access public data or to find out more about sensitive health, or restricted data products and access procedures, visit the HRS web site.

Contact Information

If you need to contact us, you may do so by one of the methods listed below.

Internet: Help Desk at the HRS web site (http://hrsonline.isr.umich.edu)

E-mail: hrsquestions@umich.edu

Postal Service:

Health and Retirement Study
The Institute for Social Research
426 Thompson Street, 3050 ISR
Ann Arbor, Michigan 48104

Citing this Document

Please include the following citation in any research reports, papers, or publications based on these data:

In text:

"The HRS (Health and Retirement Study) is sponsored by the National Institute on Aging (NIA U01AG009740) and is conducted by the University of Michigan. "

In references:

"Faul JD, Collins S, Smith JA, Zhao W, Kardia S, Weir DR . APOE and Serotonin Transporter Alleles – Early release. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan; 2021."