MIDUS Genomics Project

APOE Data Documentation

For File:

M2M3MR1_GEN_APOE_N1535_20250512

May 2025

MIDUS Apolipoprotein E (APOE) Data Documentation

Overview

APOE genotyping was conducted on 1,535 Biomarker project participants from MIDUS 2 (M2), MIDUS 3 (M3), and MIDUS Refresher 1 (MR1). Specifically, 567 from M2, 148 from M3, and 820 from MR1 Biomarker participants who consented to genomic study and provided blood sample were included in the APOE dataset. Note that for 148 Core cases whose consent to genomic study and blood sample were available at M3, but not at M2, M3 blood DNA sample was used for APOE genotyping. The data file includes only whose APOE data is available.

MIDUS follows variable naming and coding conventions (seen Naming and Coding Conventions included with the Survey documentation), thus the variable names for the APOE data begin with the unique 6 characterset BCRA6D based on the source of the blood DNA sample (i.e., B for M2, C for M3, RA for Refresher 1, 6 for Genomics project, D for data derived from DNA). The SPSS data file "M2M3MR1_GEN_APOE_N1535_20250512.sav" contains 6 variables as follows:

- Administrative:
    - M2MRID – contains the public identifiers for the MIDUS core sample and the MIDUS Refresher (for details, see README of APOE data)
    - SAMPLMAJ – the identifier created by the MIDUS Administrative Core to indicate the participants' sample of origin (e.g. MIDUS Refresher, Twin, etc.),
    - M2M3MRCASE – indicates source of blood DNA sample (e.g., MIDUS 2, MIDUS 3 or MIDUS Refresher 1)
- APOE
    - BCRA6DRS429358 – Rs429358
    - BCRA6DRS7412 – Rs7412
    - BCRA6DAPOE – APOE genotype

DNA Samples and Isolation

Whole-EDTA blood was collected from each participant and aliquoted into cryovials that were later stored at –80°C. QIAamp 96 DNA Blood Isolation kits (Qiagen, Germantown, MD, USA; cat. no. 51162) were used for extraction of genomic DNA according to the manufacture protocol. DNA concentrations were determined using the QuantiFluor ONE dsDNA System (Promega, Madison, WI, USA; cat. no. E4870) according to manufacturer's protocol and quantified using the Agilent Biotek Synergy H1 fluorometer. Following determination of DNA concentrations, 20ul samples in 96 well PCR plates were submitted in batches to the University of Wisconsin-Madison Biotechnology Center for APOE haplotype sequencing.

Construction and Sequencing of APOE libraries

Concentration of genomic DNA submitted to the University of Wisconsin-Madison Biotechnology Center was further verified fluorometrically using either the Qubit® dsDNA HS Assay Kit or Quant-iT™ PicoGreen® dsDNA Assay Kit (ThermoFisher Scientific, Waltham, MA, USA). Samples were there prepared in a similar process to the one described in Illumina's 16s Metagenomic Sequencing Library

Preparation Protocol, Part #15044223 Rev. B (Illumina Inc., San Diego, California, USA) with the following modifications:

The APOE gene was amplified with fusion primers:
forward primer:
5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCTACAGAATTCGCCCCGGCCTGGTACAC-3'

reverse primer:
5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAAGCTTGGCACGGCTGTCCAAGGA-3'

Region specific primers, previously described in Hixson and Vernier 1990 (underlined sequences above) were modified to add Illumina adapter overhang nucleotide sequences 5' of the gene-specific sequences. Following initial amplification, reactions were cleaned using a 1x volume of AxyPrep Mag PCR clean-up beads (Axygen Biosciences, Union City, CA). In a subsequent PCR, Illumina dual indexes and Sequencing adapters were added using the following primers

Forward primer:
5'-AATGATACGGCGACCACCGAGATCTACAC[55555555]ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'

Reverse Primer:
5'-CAAGCAGAAGACGGCATACGAGAT[77777777]GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT -3', Bracketed sequences are equivalent to the Illumina Dual Index adapters D501-D508 and D701-D712,N716,N718-N724,N726-N729).

Following PCR, reactions were cleaned using a 0.8x volume of AxyPrep Mag PCR clean-up beads (Axygen Biosciences). Quality and quantity of the finished libraries were assessed using an Agilent DNA 1000 kit (Agilent Technologies, Santa Clara, CA) and Qubit® dsDNA HS Assay Kit (ThermoFisher Scientific), respectively. Libraries were pooled in an equimolar fashion and appropriately diluted prior to sequencing. Paired end, 150 bp sequencing was performed using the Illumina NovaSeq X Plus (Illumina, San Diego, CA).

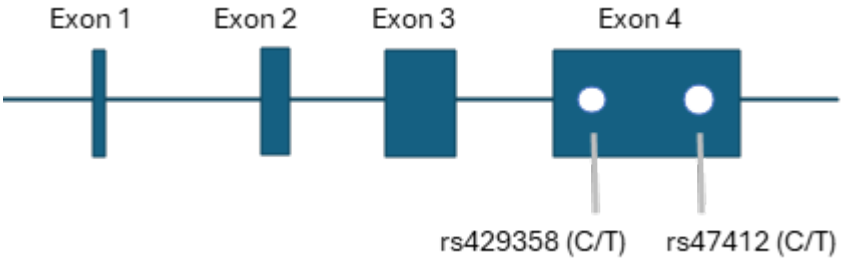
Bioinformatics

Apolipoprotein E [HGNC:613] references were obtained from Ensemble database: geneID ENSG00000130203, a region spanning chromosome 19 from positions 44,905,754 – 44,909,393 on the forward strand in assembly GRCh38. Single nucleotide variants (SNV) SNV-122 and SNV-158 are located at offsets 2931 and 3069, respectively, from the beginning of the sequence (Hixson and Vernier, 1990). The trimming software Skewer (Jiang, et al., 2014) was used to preprocess raw fastq files. End-nucleotides of each read were trimmed until a minimum Phred quality of 20 was reached. Of those, only reads with a minimum read length of 25 nucleotides were retained for further analysis. Haplotypes were determined from a custom R script developed at the Bioinformatics Resource Center, UW-Madison. Briefly, trimmed paired-end reads were mapped to the ApoE reference with the Rsubread software package (Liao, et al., 2013) allowing at most three mismatches and no insertions or deletions. An added constraint of mapping uniqueness within the reference sequence was also imposed. Soft clipping introduced during the alignment was filtered and only those mappings representing a properly mated pair were retained. A BCF file was created with the mpileup function in SAMtools (Li, et al., 2009) and used as input in freebayes (Garrison and Marth, 2012) to identify SNVs and their associated quality metrics. Alleles were obtained by querying first each nucleotide at sites SNV-122 (read1) and SNV-158

(read2) in each mapped and mated read pair. Haplotypes were assembled by combining allele calls for each matched pair in the 5' to 3' direction, using the naming convention reported in Hixson and Vernier (1990). A threshold filter eliminated any low-count haplotype instances (10% of the maximum observed haplotype count), which likely reflected sequencing errors leading to misleading low-frequency observations.

APOE Haplotype

APOE haplotypes are determined by two single nucleotide polymorphisms (SNPs), rs7412 and rs429358. Combinations of these SNPs result in the three major APOE isoforms ε2, ε3, and ε4. APOE genotype column report summarized data from Rs429358 and Rs7412 sequencing.
The summarized data results from combining the values in the SNP columns (Rs429358 and Rs7412).



Allele	**ε2**	**ε3**	**ε4**
Haplotype	Rs429358 (T) Rs7412 (T)	Rs429358 (T) Rs7412 (C)	Rs429358 (C) Rs7412 (C)

Top: Schematic representation of the APOE gene and positions of rs429358 and rs7412 in exon 4.
Bottom: The haplotype for **ε2, ε3, ε4** is dependent up on the combinations of C/T at rs429358 and rs7412.

References:

Emi M, Wu LL, Robertson MA, Myers RL, Hegele RA, Williams RR, White R, Lalouel J-M. (1988) Genotyping and sequence analysis of apolipoprotein E isoforms. Genomics. 3: 373-379.

Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907 [q-bio.GN].

Hixson, J E and Vernier, D T. 1990. Restriction isotyping of human apolipoprotein E by gene amplification and cleavage with HhaI. Journal of Lipid Research. 31:(3), 545-549.

Hongshan Jiang, Rong Lei, Shou-Wei Ding and Shuifang Zhu. 2014. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics. 15:(1), 1471-2105.

Liao Y, Smyth GK and Shi W. 2013. "The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote." Nucleic Acids Research, 41, pp. e108.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. and 1000 Genome Project Data Processing Subgroup. 2009. The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics, 25, 2078-9.