Supplementary Online Content

Wingo TS, Cutler DJ, Wingo AP, et al. Association of early-onset Alzheimer disease with elevated low-density lipoprotein cholesterol levels and rare genetic coding variants of *APOB*. *JAMA Neurol*. Published online May 28, 2019. doi:10.1001/jamaneurol.2019.0648

This supplementary material has been provided by the authors to give readers additional information about their work.

eMethods. Participants and Data Analysis

Study Population

All individuals were recruited from U.S. ADRCs or ADCs using established comprehensive research evaluations and diagnostic criteria[2-4]. EOAD cases were defined as having either probable or definite AD using the NINCDS-ADRDA Work Group criteria[4] with symptoms beginning on or before age 65, and controls were cognitively normal individuals over 60 years old. Samples available from Emory University and UCSF ADRCs had both genetic material frozen plasma available for analysis. Additional samples with available genetic material were obtained from the other ADRCs and ADCs from the National Cell Repository for Alzheimer's Disease (NCRAD) biorepository bank. Samples were restricted to Caucasians because only this group had sufficient sample size for meaningful testing. For the samples obtained from NCRAD, we obtained demographic data and *APOE* genotyping on available samples from the National Alzheimer's Coordinating Center (NACC). All participants or their legal guardians gave informed consent. Samples from each center were obtained using research protocols approved by their respective institutional review boards. All genetic analyses for this work were approved by the Emory University Institutional Review Board. The Emory ADRC samples constituted the discovery sample. There were two replication datasets that were sequenced independently – the UCSF ADRC samples and the ADRC/ADCs samples obtained from NCRAD.

LDLC measurement

Plasma samples that had not previously been thawed were available for 44% (267 / 667) of the Emory and UCSF samples for LDLc measurement. Samples were randomized with respect to sex, disease status, and age-at-draw for the Emory and UCSF, which were analyzed together. Lipid levels were determined by colorimetric method on the Beckman AU480 chemistry analyzer

at the Emory Lipid Research Laboratory (Atlanta, GA) for lipid levels analysis. Total cholesterol and total triglycerides were assayed using reagents from Beckman Diagnostics and LDLc and HDLc levels were determined by the homogeneous assay methods using reagents from Sekisui Diagnostics. ApoB levels were determined by immunoturbidometric methods using reagents from Sekisui Diagnostics.

Targeted resequencing

Genomic DNA was obtained from human blood for the Emory samples using the Gentra Puregene kit (Qiagen) following the manufacturer's protocols. For the UCSF and NCRAD samples extracted genomic DNA was provided by their respective centers. Two sets of capture primers were designed using conditions recommended for the Access Array System (Fluidigm, San Francisco, CA) by the manufacturer using the MPD software[5]. The first set was designed to capture all exonic regions and 15 bp of flanking intronic regions of *APOB*, *APP*, *PSEN1*, *PSEN2,* and 25 common X chromosome markers. The second set was designed to capture 400 ancestrally informative and 25 common X chromosome markers. Sequence capture was performed using the Access Array System (Fluidigm, San Francisco, CA) per the manufacturer's protocol in batches of 48 samples. Within each dataset, samples were randomly plated with respect to affectation, sex, education, and *APOE* genotype to minimize the influence of batch effects. All samples were barcoded using the manufacturer's protocol and 150bp (first primer set) and 250bp (second primer set) paired-ended sequencing was performed on an MiSeq (Illumina, San Diego, CA).

Base calling and quality control

Raw sequencing reads were mapped to the GRCh38 build of the human genome using PEMapper and variant sites called using PECaller using default values[6]. All variant annotation

and filtering was performed using Bystro[7]. Quality control occurred in two phases. First, samples were examined within batches and amplicons with >3 standard deviation (SD) missing sites were dropped. Samples with >3 SD missing data were dropped, and batches with >3 SD sample failures were excluded. Samples from passing batches were combined and samples with >3 SD missing data or >3 SD excess heterozygosity were dropped. In total, 2 batches and 71 individual samples were removed from further analysis. Genotyping from all remaining batches were merged and variant sites with >5% missing data and those that failed Hardy-Weinberg filtering at 10^{-7} were excluded. Samples showing discordance between demographic and genetic sex, defined as within 3 SD of the X chromosome heterozygosity for the recorded sex, were removed (n=19). Cryptically related or duplicate samples were identified by identity-by-state sharing using PLINK[8] and removed (n=32). Unlinked ancestrally informative markers were used to infer eigenvectors for principal-component analysis using EIGENSTRAT [9] and >6 SD outliers (n=1) were removed. Summaries of the sample demographics were made using samples that passed quality control measures to this point.

For the *APOB* gene-based analyses, samples with probable or possible AD-causing mutations in *APP*, *PSEN1*, or *PSEN2* were removed from further analysis because these mutations presumably explain EOAD in those individuals. Probable or possible AD-causing mutations were defined by ClinVar and the Alzheimer's Disease and Frontotemporal Dementia Mutation Database[10,11].

APOE Genotyping

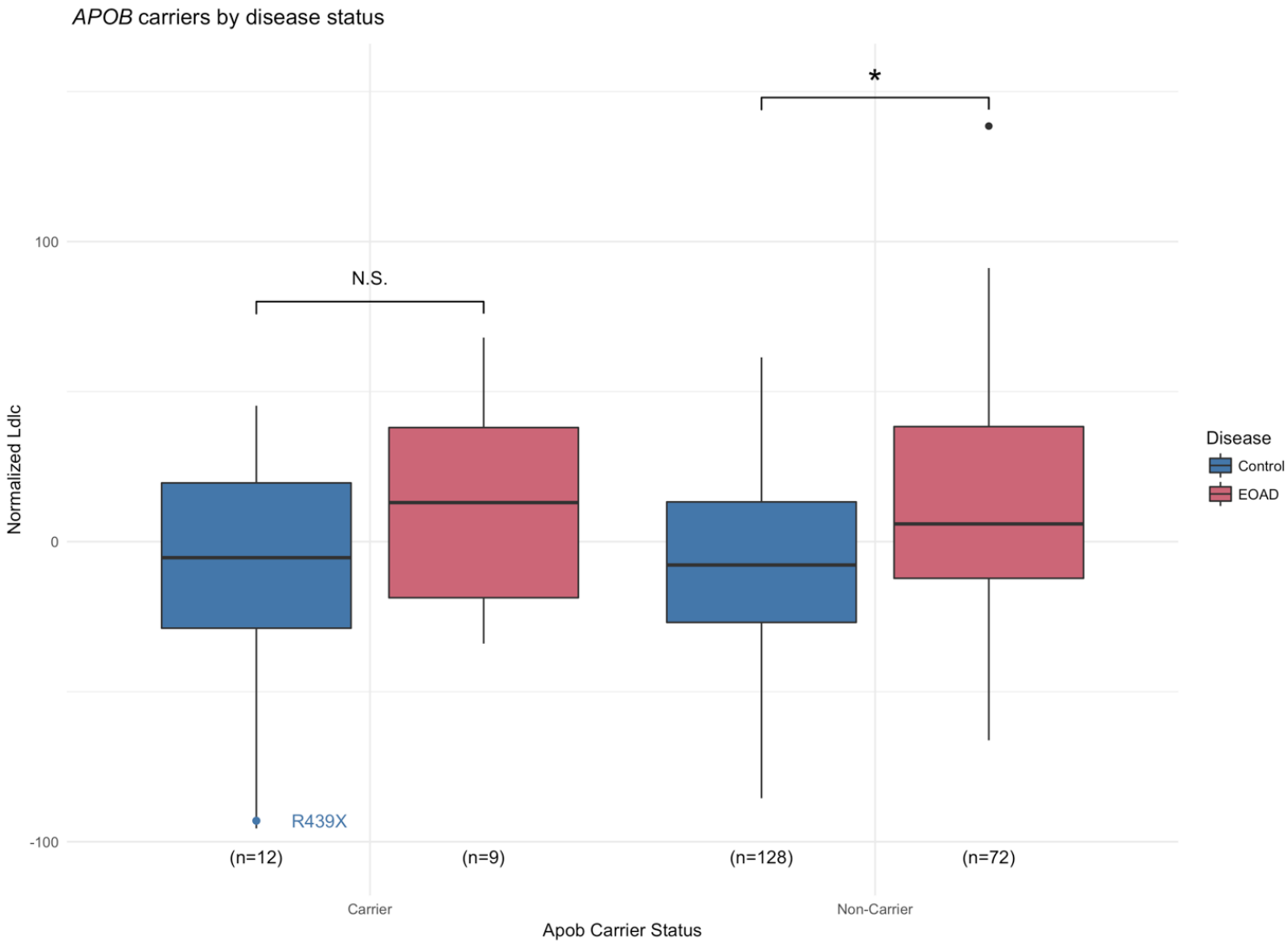
APOE genotyping was performed in triplicate using custom TaqMan assays to rs7412 and rs426758 following the manufacturer's protocol using an ABI 7300 Real-Time PCR machine.

Statistical Analysis

Plasma lipid analysis was performed in R (v3.3.1)[12], using glm for regression and built-in functions unless otherwise specified. All variables were tested for missingness that disproportionately affected either cases or controls within each center (e.g., Emory, UCSF, and NCRAD) by either Fisher's Exact Test when groups included counts less than 10 or Chi-squared test when all groups included 10 or more counts. Estimates of the liability to AD were generated using age-at-onset for cases and age-at-enrollment for controls, sex, and affectation status by linear interpolation, as described previously[13]. Linear regression was used to test for association between liability to EOAD and copies of *APOE E4* alleles adjusting for sex and sample source. To estimate the variance explained by *APOE E4*, we used liability to EOAD as the outcome and fitted one linear model with copies of *APOE E4* and another without copies of *APOE E4*, adjusting for the same covariates (i.e., sex and sample source). The proportion of the variance explained by *APOE E4* was calculated as the difference in R^2 between the two models.

To estimate the variance of AD liability explained by LDLc, we fitted one linear model with LDLc and another without LDLc, adjusting for sex, education, and copies of *APOE E4* alleles, and calculated the difference in R^2 between the two models.

eFigure. Normalized LDLC for Carriers of *APOB* Variants



Normalized LDLc levels are adjusted for sex, batch, education, and copies of *APOE E4* alleles. **A.** Normalized LDLc levels among carriers of *APOB* variants for EOAD cases and controls. EOAD cases with rare coding *APOB* variants tested in this study do not have a statistically significant difference in normalized LDLc compared to controls (N=21; p=0.28) by logistic regression, which is likely due to limited power; however, we continue to observe higher LDLc among EOAD cases compared to controls suggesting the involvement of other genetic or environmental factors (N=200; $p=2.6\times10^{-4}$). We also find that not all *APOB* variants act to increase LDLc as evidenced by a control with low LDLc (red point) who carries a stop-gain variant in *APOB* (p.439R>X) previously associated with low LDLc[1].

eTable 1. Known Pathogenic Mutations Identified			
Chromosome	**Position**	**Exon**	**Mutation**
chr1	226888112	EX6 TM-III	g.6206A>G(Met174Val)
chr1	226888975	EX7 TM-V	g.7069T>C(Leu238Pro)
chr14	73170945	EX4 N-Term	g.22921C>T(Ala79Val)
chr14	73186877	EX6 TM-III	g.38802T>C(Ser169Pro)
chr14	73186881	EX6 TM-III	g.38806C>T(Ser170Phe)
chr14	73192712	EX7 TM-IV	g.44636G>C(Gly206Ala)
chr14	73198052	EX8 HL-VI a	g.49976C>T(Pro264Leu)
chr14	73198067	EX8 HL-VI a	g.49991G>A(Arg269His)
chr14	73211906	EX10 HL-VI b	g.63824T>G(Ser365Ala)
chr14	73217170	EX11 TM-VII	g.69088C>G(Leu392Val)
chr14	73219161	EX12 TM-VIII	g.71080G>C(Ala426Pro)
chr14	73219177	EX12 HL-VIII	g.71096C>A(Ala431Glu)
chr21	25891784	EX17 TM-I	g.275341G>A(Val717Ile;_London_APP)
chr21	25891789	EX17 TM-I	g.275336T>C(Val715Ala;_German_APP)

eTable 2. Results of Gene-Based Burden *t* Test for *APP*					
Set	Emory	UCSF	ADCs	**Meta-Analysis**	**Adjusted Meta-Analysis***
Samples (n)	381	181	1563	2125	2088
Variants (n)	8	6	14	26	26
Average Allele Frequency (%)	0.13	0.27	0.05	0.032	0.033
Min MAF (%)	0.13	0.27	0.032	0.024	0.024
Effect Size	0.0057	0.0088	0.0018	0.0024	0.0021
P-value	0.32	0.4	0.38	0.076	0.055
All analyses were adjusted for Sex and the first 4 PCs					
*adjusted for Sex, first 4 PCs, Apoe E4, ADC Center, and Sequencing Batch					

eTable 3. Results of Gene-Based Burden *t* Test for *PSEN1*					
Set	Emory	UCSF	ADCs	**Meta-Analysis**	**Adjusted Meta-Analysis***
Samples (n)	381	181	1563	2125	2088
Variants (n)	3	4	15	20	20
Average Allele Frequency (%)	0.13	0.34	0.046	0.033	0.034
Min MAF (%)	0.13	0.28	0.032	0.024	0.024
Effect Size	-0.11	0.42	0.41	0.37	0.21
P-value	0.7	0.056	0.000016	0.000011	0.0026
All analyses were adjusted for Sex and the first 4 PCs					
*adjusted for Sex, first 4 PCs, Apoe E4, ADC Center, and Sequencing Batch					

eTable 4. Results of Gene-Based Burden *t* Test for *PSEN2*					
Set	Emory	UCSF	ADCs	**Meta-Analysis**	**Adjusted Meta-Analysis***
Samples (n)	381	181	1563	2125	2088
Variants (n)	6	6	14	22	22
Average Allele Frequency (%)	0.22	0.28	0.057	0.044	0.045
Min MAF (%)	0.13	0.28	0.032	0.024	0.024
Effect Size	-0.13	-0.042	0.062	0.012	-0.031
P-value	0.39	0.84	0.45	0.86	0.58
All analyses were adjusted for Sex and the first 4 PCs					
*adjusted for Sex, first 4 PCs, Apoe E4, ADC Center, and Sequencing Batch					

eTable 5. Variants Identified in *APOB*

chrom	pos	type	ref allele	alt allele	phastCons	phyloP	cadd	refSeq Allele Function	refSeq ref Codon	refSeq alt Codon	refSeq ref Amino Acid	refSeq alt Amino Acid	refSeq codon Position	refSeq codon Number	dbSNP name	gnomad genomes Allele Frequency	gnomad exomes Allele Frequency
chr2	21002683	SNP	G	A	1	2.8	47	stopGain	CAA	TAA	Q	*	1	4247	rs907126709	!	2.84998E-05
chr2	21002979	SNP	G	T	0.26	2	23.9	nonSynonymous	GCC	GAC	A	D	2	4148	rs756476242	3.23102E-05	2.15159E-05
chr2	21003079	SNP	C	A	0.48	0.51	22.5	nonSynonymous	GTT	TTT	V	F	1	4115	rs765681925	!	!
chr2	21004405	SNP	G	T	1	4	23.7	nonSynonymous	ACC	AAC	T	N	2	3984	!	!	!
chr2	21004414	SNP	G	A	1	4.78	25.8	nonSynonymous	CCA	CTA	P	L	2	3981	rs886055575	!	8.12513E-06
chr2	21005113	SNP	A	T	1	6.51	24.6	nonSynonymous	TCA	ACA	S	T	1	3919	!	!	!
chr2	21005124	SNP	G	A	1	6.08	24.6	nonSynonymous	TCC	TTC	S	F	2	3915	rs201990496	3.2281E-05	2.03164E-05
chr2	21005559	SNP	A	C	1	8.05	25.7	nonSynonymous	ATC	AGC	I	S	2	3770	rs199954127	6.45911E-05	4.0699E-05
chr2	21005813	SNP	C	T	1	7.61	46	stopGain	TGG	TGA	W	*	3	3685	!	!	!
chr2	21006088	SNP	A	G	0.99	2.32	21.4	nonSynonymous	TGG	CGG	W	R	1	3594	rs61744288	0.00245446	0.000610128
chr2	21006397	SNP	A	G	1	5.05	24.7	nonSynonymous	TCC	CCC	S	P	1	3491	!	!	!
chr2	21006876	SNP	A	G	1	9.32	25.2	nonSynonymous	ATT	ACT	I	T	2	3331	!	!	!
chr2	21006988	SNP	A	G	1	5.14	23.6	nonSynonymous	TCA	CCA	S	P	1	3294	rs12720855	0.0146631	0.00388678
chr2	21007080	SNP	A	G	1	5.39	22.9	nonSynonymous	TTC	TCC	F	S	2	3263	!	!	4.06623E-06
chr2	21007648	SNP	A	G	1	3.61	24.4	nonSynonymous	TAT	CAT	Y	H	1	3074	!	!	!
chr2	21007776	SNP	C	T	1	7.91	27.9	nonSynonymous	GGA	GAA	G	E	2	3031	!	!	4.06739E-06
chr2	21008055	SNP	T	C	1	8.02	26	nonSynonymous	GAT	GGT	D	G	2	2938	rs377687939	3.22914E-05	4.06603E-06
chr2	21008655	SNP	A	T	1	9.32	25.9	nonSynonymous	ATA	AAA	I	K	2	2738	rs1032400403	!	!
chr2	21009354	SNP	A	T	0.01	2.48	36	stopGain	TTG	TAG	L	*	2	2505	!	!	!
chr2	21009538	SNP	G	A	0.85	2.15	23.9	nonSynonymous	CGT	TGT	R	C	1	2444	rs754565622	9.68617E-05	8.55321E-05
chr2	21010453	SNP	T	C	1	5.94	22.5	nonSynonymous	AAG	GAG	K	E	1	2139	!	!	!
chr2	21011296	SNP	T	C	0.98	6.22	20.6	nonSynonymous	AAG	GAG	K	E	1	1858	rs762535625	!	!
chr2	21011334	SNP	G	A	0.77	1.34	23	nonSynonymous	GCT	GTT	A	V	2	1845	!	!	!
chr2	21011427	SNP	A	G	0.94	7.57	25.3	nonSynonymous	CTA	CCA	L	P	2	1814	!	!	!

chr2	21012272	SNP	C	G	0	-0.8	22.9	nonSynonymous	CAG	CAC	Q	H	3	1532	!	!	!
chr2	21012586	SNP	G	A	1	8.98	25.4	nonSynonymous	CAC	TAC	H	Y	1	1428	!	!	!
chr2	21013198	SNP	G	A	1	4.75	24.2	nonSynonymous	GCT	GTT	A	V	2	1393	rs143282164	0.000807285	0.000215325
chr2	21013349	SNP	G	A	1	8.01	24.8	nonSynonymous	CCC	TCC	P	S	1	1343	rs374427541	!	2.43677E-05
chr2	21013351	SNP	A	G	1	9.26	24.9	nonSynonymous	ATT	ACT	I	T	2	1342	!	!	!
chr2	21013517	SNP	A	G	1	5.73	24.4	nonSynonymous	TAT	CAT	Y	H	1	1287	!	!	!
chr2	21014512	SNP	C	T	1	3.75	23.8	nonSynonymous	GAG	AAG	E	K	1	1260	rs144892654	!	!
chr2	21015252	SNP	T	C	1	6.92	26.3	nonSynonymous	AAG	GAG	K	E	1	1173	!	!	!
chr2	21015260	SNP	T	C	1	6.92	27.7	nonSynonymous	GAT	GGT	D	G	2	1170	!	!	4.06167E-06
chr2	21015480	SNP	G	A	1	2.59	24.3	nonSynonymous	GCC	GTC	A	V	2	1133	!	!	!
chr2	21015516	SNP	T	C	1	1.35	22.8	nonSynonymous	AAG	AGG	K	R	2	1121	rs879255342	!	!
chr2	21019069	SNP	G	T	1	6.57	26	nonSynonymous	TCT	TAT	S	Y	2	1015	rs773281594	!	!
chr2	21019726	SNP	G	A	1	2.86	20.8	nonSynonymous	ACC	ATC	T	I	2	999	!	!	!
chr2	21019808	SNP	C	T	1	7.78	32	nonSynonymous	GGC	AGC	G	S	1	972	rs199893862	!	5.68574E-05
chr2	21022924	SNP	T	C	1	5.78	21.4	nonSynonymous	GAG	GGG	E	G	2	908	!	3.22726E-05	8.12387E-06
chr2	21023647	SNP	G	A	1	7.56	22.7	nonSynonymous	CAC	TAC	H	Y	1	828	!	!	!
chr2	21025036	SNP	C	T	0.86	0.22	22.4	nonSynonymous	CGC	CAC	R	H	2	778	rs201595604	3.22789E-05	5.6862E-05
chr2	21025078	SNP	T	C	1	4.82	28.9	nonSynonymous	GAT	GGT	D	G	2	764	rs781111539	!	!
chr2	21025123	SNP	T	C	1	6.57	24.2	nonSynonymous	GAT	GGT	D	G	2	749	!	!	!
chr2	21026793	SNP	C	T	1	3.42	24.7	nonSynonymous	GAG	AAG	E	K	1	747	rs756379920	!	4.06795E-06
chr2	21027902	SNP	T	A	0.99	2.92	27.3	nonSynonymous	AAC	TAC	N	Y	1	665	rs755666654	!	3.24947E-05
chr2	21027988	SNP	T	C	1	5.19	22.9	nonSynonymous	AAC	AGC	N	S	2	636	!	!	!
chr2	21028007	SNP	A	G	1	6.2	24.9	nonSynonymous	TTC	CTC	F	L	1	630	!	!	!
chr2	21028510	SNP	A	C	0.94	1.84	24.1	nonSynonymous	CTT	CGT	L	R	2	549	rs745897939	!	1.62862E-05
chr2	21029662	SNP	G	A	1	0.27	27.4	nonSynonymous	CGG	TGG	R	W	1	532	rs13306194	0.00913729	0.0108896
chr2	21029679	SNP	G	A	1	7.09	33	nonSynonymous	GCT	GTT	A	V	2	526	!	!	!
chr2	21029719	SNP	A	G	1	8.36	27.2	nonSynonymous	TGT	CGT	C	R	1	513	rs938250596	!	!

chr2	21032490	SNP	G	A	0.3	1.21	23.7	nonSynonymous	CTT	TTT	L	F	1	406	!	!	!
chr2	21033498	SNP	C	T	0.28	0.71	21.5	nonSynonymous	GCA	ACA	A	T	1	309	rs141888564	0.000226025	0.000194959
chr2	21034830	SNP	C	T	1	3.48	34	nonSynonymous	CGC	CAC	R	H	2	297	rs147223101	0.000225967	7.30923E-05
chr2	21035590	SNP	G	A	1	7.63	24.9	nonSynonymous	TCC	TTC	S	F	2	271	rs750907137	!	4.06187E-06
chr2	21038001	SNP	A	G	1	4.59	26.9	nonSynonymous	CTG	CCG	L	P	2	165	rs375285536	!	4.06058E-06
chr2	21038047	SNP	G	A	1	2.4	24.5	nonSynonymous	CCT	TCT	P	S	1	150	!	!	4.06062E-06

References

1. Consortium UK, Walter K, Min JL, et al. The UK10K project identifies rare variants in health and disease. *Nature.* 2015;526(7571):82-90.
2. NACC. Description of available data. http://www.alz.washington.edu/NONMEMBER/INVEST/desc.html. Accessed 2015-06-08.
3. Beekly DL, Ramos EM, Lee WW, et al. The National Alzheimer's Coordinating Center (NACC) database: the Uniform Data Set. *Alzheimer Dis Assoc Disord.* 2007;21(3):249-258.
4. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology.* 1984;34(7):939-944.
5. Wingo TS, Kotlar A, Cutler DJ. MPD: multiplex primer design for next-generation targeted sequencing. *BMC bioinformatics.* 2017;18(1):14.
6. Johnston HR, Chopra P, Wingo TS, et al. PEMapper and PECaller provide a simplified approach to whole-genome sequencing. *Proc Natl Acad Sci U S A.* 2017.
7. Kotlar AV, Trevino CE, Zwick ME, Cutler DJ, Wingo TS. Bystro: rapid online variant annotation and natural-language filtering at whole-genome scale. *Genome biology.* 2018;19(1):14.
8. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81(3):559-575.
9. Patterson N, Price AL, Reich D. Population Structure and Eigenanalysis. *PLoS Genet.* 2006;2(12):e190.
10. Cruts M. Alzheimer Disease & Frontotemporal Dementia Mutation Database. http://www.molgen.ua.ac.be/ADMutations. Accessed 08-30-2016, 2016.
11. Landrum MJ, Lee JM, Benson M, et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* 2016;44(D1):D862-868.
12. *R: A Language and Environment for Statistical Computing* [computer program]. R Foundation for Statistical Computing; 2016.
13. Wingo TS, Lah JJ, Levey AI, Cutler DJ. Autosomal recessive causes likely in early-onset Alzheimer disease. *Archives of neurology.* 2012;69(1):59-64.