

# Genetic associations with age of menopause in familial longevity

Harold Bae, PhD,<sup>1</sup> Kathryn L. Lunetta, PhD,<sup>2</sup> Joanne M. Murabito, MD,<sup>3</sup> Stacy L. Andersen, PhD,<sup>4</sup> Nicole Schupf, PhD,<sup>5</sup> Thomas Perls, MD, MPH,<sup>4</sup> Paola Sebastiani, PhD,<sup>2</sup> on behalf of the Long Life Family Study

---

## Abstract

**Objective:** We hypothesize that mechanisms associated with extended reproductive age may overlap with mechanisms for the selection of genetic variants that slow aging and decrease risk for age-related diseases. Therefore, the goal of this analysis is to search for genetic variants associated with delayed age of menopause (AOM) among women in a study of familial longevity.

**Methods:** We performed a meta-analysis of genome-wide association studies for AOM in 1,286 women in the Long Life Family Study (LLFS) and 3,151 women in the Health and Retirement Study, and then sought replication in the Framingham Heart Study (FHS). We used Cox proportional hazard regression of AOM to account for censoring, with a robust variance estimator to adjust for within familial relations.

**Results:** In the meta-analysis, a single nucleotide polymorphism (SNP) previously associated with AOM reached genome-wide significance (rs16991615; HR = 0.74,  $P = 6.99 \times 10^{-12}$ ). A total of 35 variants reached  $>10^{-4}$  level of significance and replicated in the FHS and in a 2015 large meta-analysis (ReproGen Consortium). We also identified several novel SNPs associated with AOM including rs3094005: *MICB*, rs13196892: *TXNDC5* | *MUTED*, rs72774935: *SSBP2* | *ATG10*, rs9447453: *COL12A1*, rs114298934: *FHL2* | *NCK2*, rs6467223: *TNPO3*, rs9666274 and rs10766593: *NAV2*, and rs7281846: *HSPA13*.

**Conclusions:** This work indicates novel associations and replicates known associations between genetic variants and AOM. A number of these associations make sense for their roles in aging.

**Key Words:** Association – Evolution – Genetics – Longevity – Menopause.

**Video Summary:** Supplemental Digital Content 1, <http://links.lww.com/MENO/A420>.

---

Natural menopause occurs with a sharp decline in ovarian production of estradiol and progesterone resulting in the cessation of ovarian follicle formation, menstruation, and the ability to bear children.<sup>1</sup> The age of menopause (AOM) is usually clinically defined as 1 year after the final menstruation and is on average, about 51 years.<sup>2</sup> Menopause seems to be under tight genetic control<sup>3</sup> with numerous known genetic associations for both premature AOM (premature ovarian failure)<sup>1,3,4</sup> and normal AOM.<sup>1,5-12</sup> There are also conditions and exposures associated with earlier AOM, including smoking,<sup>13-15</sup> chemotherapy, and elevated body mass index.<sup>16,17</sup>

Menopause is practically unique to humans, with notable exceptions being two species of toothed whale (resident killer whale and short-finned pilot whale), which, like human females, have the potential to live for a prolonged period of time beyond AOM.<sup>18</sup> An evolutionary mechanism for the selection of genetic variants that cause menopause in humans is posited by the adaptive theory of menopause and is based upon the fact that at some age and thereafter, a mother's increasing frailty with increasing age is associated with greater and greater childbirth-associated mortality risk.<sup>19,20</sup> Female mammals other than humans have straight birth canals and as a result negligible mortality risk is associated with the

---

Received February 15, 2019; revised and accepted April 11, 2019.

From the <sup>1</sup>College of Public Health and Human Sciences, Oregon State University, Corvallis, OR; <sup>2</sup>Department of Biostatistics, Boston University School of Public Health, Boston, MA; <sup>3</sup>Section of General Internal Medicine, Department of Medicine, and the Framingham Heart Study, Boston University School of Medicine, Boston, MA; <sup>4</sup>Geriatrics Section, Department of Medicine, Boston Medical Center and Boston University School of Medicine, Boston, MA; and <sup>5</sup>Department of Epidemiology, Mailman School of Public Health, Columbia University, NY.

Funding/support: This work was supported by the National Institute on Aging (NIA cooperative agreements U01-AG023712, U01-AG23744, U01-AG023746, U01-AG023749, and U01-AG023755) as well as NIA funded grants: U19-AG023122 TP, and R21AG056630 PS. Funding also provided by a National Institute of General Medical (NIGMS)

Interdisciplinary Training Grant for Biostatisticians [T32 GM74905], the William M. Wood Foundation (TP), and the Paulette and Marty Samowitz Family Foundation (TP). The Health and Retirement Study, whose data we used through dbGaP, is funded by the National Institute on Aging (grant numbers U01AG009740, RC2AG036495, and RC4AG039029) and is conducted by the University of Michigan.

Financial disclosure/conflicts of interest: None reported.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Website ([www.menopause.org](http://www.menopause.org)).

Address correspondence to: Harold Bae, PhD, 161 Milam Hall, School of Biological and Population Health Sciences, College of Public Health and Human Sciences, Oregon State University, Corvallis, OR 97331. E-mail: [harold.bae@oregonstate.edu](mailto:harold.bae@oregonstate.edu)

## METHODS

act of bearing young. When humans evolved an upright bipedal posture, the female birth canal, however, became the shape of a question mark making it much more difficult for the young to exit the uterus. Successfully traversing the human birth canal was made all the more difficult with the increasing size of the baby's brain and cranium. The result was higher mortality risk to both mother and child when the mother was too frail to successfully expel her newborn. Thus, the adaptive theory of menopause posits that genetic mutations were evolutionarily selected to induce menopause before such aging-associated frailty led to the mother's death and, as a result, also the deaths of any existing infants that depended upon her for survival.

From an evolutionary advantage point of view though, it would also make sense to delay menopause as long as possible when the mother is still strong enough to bear children, so that she has more time to bear children and therefore have more of them. Having more children that she is able to raise to reproductive age translates into a greater opportunity to pass down one's genes to the next generation, evolution's ultimate goal. Such a delay in menopause could result from the selection for genetic variants that contribute to slower aging and decreased risk for aging-related diseases that adversely impact upon fertility and fecundity. In what can be called the "delayed menopause selection theory" of longevity-associated genes, Perls and Fretts posited that such genes also allow for women to live well beyond the age of reproduction to raise their children and perhaps, consistent with Kristen Hawke's "grandparent hypothesis," their grandchildren, to reproductive age.<sup>19,21</sup> They posited that the by-product of some optimal combination of these and perhaps other genetic variations could enable some people to survive to the oldest ages that we observe in humans, 50 or more years beyond menopause.

Consistent with the delayed menopause selection theory, Perls, Fretts and colleagues found that women who were able to have children beyond the age of 40 years were 4 times more likely than average women from the same birth cohort, of living to 100 years or older, and those who had children at age 35 or older were 1.5 times more likely to live to 100 years or older.<sup>22</sup> The Long Life Family Study (LLFS) also found an association between a woman's age when she had her last child and survival to older age.<sup>23</sup> Although previous genetic studies have sought to discover genetic variants associated with AOM in normally aging populations, or in premature menopause, our current analysis aims to discover additional genetic variants that are associated with AOM in women with familial longevity. It is encouraging that some genetic studies of AOM have noted associations with genes that have also been implicated in slower aging such as DNA repair and immune function genes.<sup>1,8</sup> In this study of LLFS participants, a cohort of families that demonstrate unusual clustering for survival to very older ages, combined with Health and Retirement Study (HRS) participants, we performed a meta-analysis for genetic variants associated with AOM in individuals who ultimately live to a very older age. We then relied upon Framingham Heart Study (FHS) data as a replication set.

**Discovery cohorts**

The LLFS is a family based study of longevity and healthy aging in 539 families and 4,953 family members recruited from the United States and Denmark during 2006 to 2009. The study includes two generations, comprising probands and their siblings, their offspring and spouses. Details of the study design and phenotypic measures are described in Sebastiani et al.<sup>24</sup> and Newman et al.<sup>25</sup> In the LLFS, AOM was defined as self-reported age at natural menopause between the ages of 40 and 60 years<sup>11</sup>, with no history of estrogen use or hysterectomy. AOM was censored at age of hysterectomy or estrogen use if hysterectomy or estrogen use occurred before menopause. If information was not available for menopause, hysterectomy, or estrogen use, maternal age at birth of her last child was used as a conservative proxy for AOM.

The HRS is a study of healthy agers that includes 10,468 participants, and 87 individuals who survived to the oldest one percentile of their birth year cohort.<sup>26</sup> In the HRS, AOM was defined in the same manner as above, age at last period. For 205 participants there was no information on the age they began taking estrogen and so these individuals were excluded from the analysis. For participants reporting discordant AOM's at two different survey years, we selected the earlier response. AOM was censored at age of hysterectomy if hysterectomy occurred before menopause. AOM of women without menopause or hysterectomy was censored at age of the survey administration year.

**Replication cohort**

The FHS is an ongoing family based longitudinal study initiated in 1948 to identify risk factors of cardiovascular disease.<sup>27</sup> The study consists of three cohorts. The original cohort consisted of 5,209 participants aged between 28 and 62 years old. Participants were then examined every 2 years since 1948 for a total of up to 32 examinations. The offspring cohort, recruited in 1971, consisted of 5,124 offspring of the original cohort and their spouses and they have been examined every 4 to 8 years.<sup>28</sup> The third generation, recruited in 2002, consists of 4,095 children of offspring aged between 19 to 72 years old.<sup>29</sup> In the FHS, AOM was defined as self-reported age at natural menopause between the ages of 40 and 60 years. AOM was censored at age of hysterectomy if hysterectomy occurred before menopause, or age at examination where the individual reported estrogen use if estrogen use occurred before menopause.

**Genotyped and imputed data**

The LLFS DNA samples were genotyped using the Illumina Omni 2.5 or 2.5 million single nucleotide polymorphisms (SNPs) array. The details of the genotype data and quality controls are described in Bae et al.<sup>30</sup> Genome-wide genotype data were imputed to the 1,000 Genomes Haplotype Phase I integrated variant set release (in NCBI build 37 coordinates) as the reference panel using IMPUTE2.<sup>31</sup> Imputation was preceded by prephasing with the ShapeIT

program.<sup>32</sup> Only SNPs with an imputation quality score  $>0.9$  were retained, and there were approximately 7.5 million SNPs available for analysis. All participants provided informed consent. Phenotype and genetic data are available via dbGaP (dbGaP Study Accession: phs000397.v1.p1).

The HRS genotype and imputed data (from the 1000 Genomes Project) were downloaded from dbGaP (accession number: phs000428.v1.p1). Imputed SNPs with quality score  $R^2 > 0.7$  were used in the analysis. There were approximately 15 million variants that include both genotyped and imputed SNPs.

FHS genotype and imputed data: FHS DNA samples were genotyped using the Affymetrix GeneChip Human Mapping 500K array set and the 50K supplemental array set focused on coding SNPs and SNPs tagging protein-coding genes (Santa Clara, CA). Samples with call rate  $<97\%$  ( $n=767$ ), per participant heterozygosity  $> \pm 5$  SDs from the mean ( $n=24$ ) or excessive Mendelian errors ( $N=2$ ) were removed. From a total of 546,344 genotyped autosomal and X chromosome SNPs, 412,049 SNPs that satisfied Hardy–Weinberg  $P \geq 1e-6$ , MAF  $\geq 0.01$ , Mendelian errors  $<1,000$  and mapped to GRCh37 were used to impute to the Haplotype Reference Consortium release 1.1 reference panel.

### Statistical analysis

Genome-wide principal components of the LLFS data were computed with Eigensoft V.5.4<sup>33</sup> using approximately 80,000 common, independent SNPs as described in Sebastiani et al.<sup>34</sup> We also independently calculated genome-wide principal components in the HRS using the genome-wide genotype data available from dbGaP. Based on the computed principal components, we excluded participants with non-European ancestry to avoid population stratification. The association between each SNP and AOM was tested using a Cox proportional hazards model with a robust variance estimate to account for within-familial relations and an additive genetic effect. The same analysis was performed for SNPs on chromosome X given that all participants were females. The models were adjusted for four genome-wide principal components. SNPs with genotype counts of 2 or less and minor allele frequency (MAF)  $<5\%$  were filtered out. Standard estimates of log-hazard ratios and standard errors from the Cox regression in each study (LLFS and HRS) were meta-analyzed using fixed effect meta-analysis with inverse variance weighting implemented in METAL.<sup>35</sup> The genome wide level of significance was defined as  $5 \times 10^{-8}$ . SNPs with

$P < 10^{-4}$  in the meta-analysis were sought for replication in the FHS. We defined replication as SNPs having  $P < 0.05$  in the replication cohort with direction of effects consistent with those observed in the discovery set. In addition, we tested for replication of the top SNPs with publicly available data from Day et al.<sup>11</sup> (ReproGen consortium, [http://www.reprogen.org/data\\_download.html](http://www.reprogen.org/data_download.html)). Significant expression quantitative trait loci (eQTLs) were identified using the GTEx portal.<sup>36</sup>

## RESULTS

Characteristics of each cohort are summarized in Table 1. There were 1,286 (mean AOM: 50.9 y) and 3,151 (mean AOM: 50.6 y) participants whose AOM were available in the LLFS and HRS, respectively. A total of 1,108 and 2,066 participants were censored in the LLFS and HRS. In the replication cohort (FHS), there were 2,101 events and 981 censored participants.

The QQ-plots and Manhattan plots of the individual GWASs from the LLFS and HRS are shown in Supplementary Figures S1 and S2, <http://links.lww.com/MENO/A421>. In the meta-analysis, only SNP rs16991615 reached genome-wide significance (HR = 0.74,  $P = 6.99 \times 10^{-12}$ ), this is a previously published variant that reached genome-wide significance with consistent effect in a GWAS of AOM published by Day et al.<sup>11</sup> In addition, three SNPs reached  $P = 5 \times 10^{-7}$ , 19 SNPs reached  $P = 5 \times 10^{-6}$ , 329 reached  $P = 5 \times 10^{-5}$  and 576 reached  $P = 10^{-4}$ . As the sample size in our study is substantially smaller than the  $\sim 70,000$  used in Day et al.,<sup>11</sup> we investigated the set of 576 SNPs for additional replication.

The results of the top 576 SNPs and their results in the individual GWASs are summarized in Supplementary Table S1, <http://links.lww.com/MENO/A422>. These 576 SNPs correspond to 121 loci and include variants associated with higher risk for, and therefore earlier AOM and variants associated with reduced risk for, and hence delayed AOM. Nineteen of the 576 SNPs showed consistent effects and reached genome-wide statistical significance in the GWAS of AOM published by Day et al.<sup>11</sup> (Table 2). An additional 14 SNPs reached nominal level of significance ( $P < 0.05$ ) in Day et al.<sup>11</sup> The minor alleles of 13 of these 33 SNPs in Table 2 were associated with delayed AOM, whereas the remaining SNPs were associated with earlier AOM.

### Annotation and replication

Thirty-five SNPs representing 10 loci among the top 576 SNPs of the discovery meta-analysis replicated in the FHS

TABLE 1. Cohort characteristics

	LLFS		HRS		FHS	
	Events	Censored	Events	Censored	Events	Censored
<i>N</i>	1,286	1,108	3,151	2,066	2,101	981
Mean AOM $\pm$ SD	50.85 $\pm$ 3.95	NA	50.56 $\pm$ 4.42	NA	<sup>a</sup>	NA

AOM, age of menopause; FHS, Framingham Heart Study; HRS, Health and Retirement Study; LLFS, Long Life Family Study; *N*, sample size; SNP, single nucleotide polymorphisms.

<sup>a</sup>Mean AOMs in the FHS were 49.69  $\pm$  3.58 in the original cohort, 50.00  $\pm$  3.99 in the offspring generation, 50.85  $\pm$  4.62 in the new offspring spouses, and 49.44  $\pm$  4.07 in the third generation.

TABLE 2. Summary of SNPs that replicated in Day et al (2015)

SNP	Chr:Pos	CA	NCA	Gene	Discovery meta-analysis				Day et al <sup>11</sup> (2015)			
					Beta <sup>a</sup>	SE	P	Direction	Beta_LR <sup>b</sup>	SE	P	CAF
rs11851797	14:95169718	t	c	AL049839.1   RPSAP4	-0.13	0.033	9.00E-05	-	0.07	0.03	0.033	0.14
rs12448714	16:88989299	t	c	CBFA2T3	-0.10	0.024	3.62E-05	-	0.15	0.03	6.70E-07	0.78
rs1172822 <sup>c</sup>	19:55819845	t	c	BRSK1	0.10	0.022	6.05E-06	++	-0.39	0.02	1.70E-77	0.36
rs4806660 <sup>c</sup>	19:55824634	t	c	TMEM150B	-0.10	0.023	6.24E-06	-	0.39	0.02	3.30E-79	0.63
rs4806661 <sup>c</sup>	19:55824661	a	t	TMEM150B	-0.09	0.023	3.63E-05	-	0.39	0.02	3.70E-79	0.62
rs2384687 <sup>c</sup>	19:55831188	a	g	TMEM150B	-0.10	0.023	1.38E-05	-	0.39	0.02	3.60E-79	0.62
rs1109368 <sup>c</sup>	19:55832567	a	g	TMEM150B	0.10	0.023	1.54E-05	++	-0.39	0.02	7.30E-76	0.38
rs11668309 <sup>c</sup>	19:55833460	t	c	TMEM150B	0.11	0.023	4.87E-07	++	-0.41	0.02	9.20E-84	0.36
rs11668344 <sup>c</sup>	19:55833664	a	g	TMEM150B	-0.11	0.023	5.47E-07	-	0.41	0.02	5.50E-85	0.65
rs10425848 <sup>c</sup>	19:55834448	c	g	TMEM150B	-0.10	0.023	8.61E-06	-	0.39	0.02	1.50E-77	0.62
rs11676185	2:36971344	t	c	MCM8	0.09	0.022	2.15E-05	++	-0.06	0.02	0.0088	0.37
rs12712520	2:36974088	t	g	VIT	0.09	0.022	8.05E-05	++	-0.05	0.02	0.02	0.37
rs1072218	2:36974472	t	c	VIT	0.10	0.022	1.15E-05	++	-0.06	0.02	0.014	0.37
rs16991615 <sup>c</sup>	20:5948227	a	g	VIT	-0.30	0.044	6.99E-12	-	0.88	0.04	1.60E-89	0.08
rs1387818	3:21751600	a	t	ZNF385D	-0.09	0.023	4.06E-05	-	0.05	0.02	0.013	0.65
rs2130505	3:21752966	a	g	ZNF385D	-0.09	0.023	8.58E-05	-	0.05	0.02	0.02	0.64
rs6813104	4:5816938	t	c	CRMP1	0.10	0.024	4.97E-05	++	-0.06	0.03	0.035	0.22
rs1472920	5:176334601	t	g	UIMC1	0.11	0.025	1.31E-05	++	-0.22	0.02	3.60E-21	0.71
rs2962845	5:176336189	a	t	UIMC1	-0.11	0.025	9.35E-06	-	0.22	0.02	3.00E-21	0.29
rs2860534	5:176343061	a	g	UIMC1	-0.11	0.025	9.58E-06	-	0.22	0.02	3.20E-21	0.29
rs11739147	5:176359335	t	c	UIMC1	0.11	0.025	9.30E-06	++	-0.22	0.02	2.50E-21	0.71
rs2454951	5:176362137	t	c	UIMC1	-0.11	0.025	9.88E-06	-	0.21	0.02	1.10E-20	0.29
rs353465	5:176396888	t	c	UIMC1	0.11	0.025	1.22E-05	++	-0.22	0.02	1.90E-21	0.71
rs352943 <sup>c</sup>	5:176464738	a	c	ZNF346	-0.11	0.027	8.62E-05	-	0.21	0.03	2.10E-16	0.25
rs166138	5:176472339	a	g	ZNF346	-0.11	0.024	1.35E-05	-	0.22	0.02	2.10E-21	0.28
rs353482	5:176498882	t	c	ZNF346	-0.11	0.024	1.57E-05	-	0.23	0.02	4.70E-21	0.28
rs6929983	6:43249411	t	c	TTBK1	0.11	0.030	9.92E-05	++	-0.06	0.03	0.027	0.16
rs2185062	6:78908259	a	t	MEI4	-0.16	0.036	1.41E-05	-	0.07	0.04	0.042	0.9
rs2896309	7:121431954	a	g		-0.17	0.039	2.32E-05	-	0.08	0.04	0.038	0.09
rs2272347	7:128619415	a	t	TNPO3	0.09	0.022	9.20E-05	++	-0.08	0.02	0.00018	0.54
rs11761356	7:21243518	t	c		0.10	0.024	5.16E-05	++	-0.06	0.02	0.012	0.36
rs3904459	9:3032897	a	t	CARM1P1	0.11	0.028	5.36E-05	++	-0.07	0.03	0.0064	0.17
rs3842928	9:3036913	a	t	CARM1P1	-0.10	0.026	6.66E-05	-	0.07	0.03	0.01	0.82

CA, coded allele; CAF, coded allele frequency; NCA, noncoded allele; SNP, single nucleotide polymorphism.

<sup>a</sup>Beta = log-transformed hazard ratio for age of menopause of one copy of the coded allele relative to the noncoded allele.

<sup>b</sup>Beta\_LR = genetic effect on age of menopause of one copy of the coded allele. Note that higher risk for menopause in the survival analysis (beta >0) corresponds to earlier age of menopause (beta\_LR <0).

<sup>c</sup>SNPs with <sup>a</sup> are also in Table 3.

with the same genetic effect using the same survival regression model and nominal level of significance (Table 3). Of these 35 SNPs, 5 SNPs at 3 loci (rs16991615: *MCM8*, rs1172822: *BRSK1*, rs11668309 and rs2384687: *TMEM224*, rs352943: *ZNF346*) are variants that reached genome-wide significance in the GWAS of AOM published by Day et al.<sup>11</sup> The nine previously unreported variants that replicated in the FHS include rs114298934, located between genes *FHL2* and *NCK2*, rs13196892, located between genes *TXNDC5* and *MUTED*, rs3094005 in *MICB*, rs9447453 in *COL12A1*, rs6467223 in *TNPO3*, rs9666274 and rs10766593 in *NAV2*, rs7281846 in *HSPA13* and rs72774935 located between genes *SSBP2* and *ATG10*. The coded alleles of rs3094005, rs9447453, rs114298934, and rs72774935 are uncommon in the population and are associated with earlier AOM. The additional previously unreported variant rs13196892 is an uncommon variant and its coded allele is associated with delayed AOM. The coded alleles of rs9666274, rs10766593, and rs7281846 are associated with earlier AOM, whereas the coded allele of rs6467223 is associated with delayed AOM. Functional annotation with expression data using the GTEx database<sup>36</sup> shows that rs3094005 is a significant eQTL of

*MICB* in 25 different tissue types (Fig. 1) and in most of these tissues, the T allele, associated with increased risk for earlier AOM, increases the expression levels of *MICB*. SNP rs6467223 is a significant eQTL of *IRF5* in 19 different tissue types, and in most tissues, the A allele, associated with delayed AOM, increases the expression levels of *IRF5*. SNP rs13196892 is also a significant eQTL for *TXNDC5* in the esophagus. None of the remaining SNPs (rs9447453, rs72774935, rs9666274, rs10766593, rs7281846, rs114298934) seemed to be significant eQTLs based on GTEx data.

The list of 576 SNPs that reached a significance level <10<sup>-4</sup> includes 186 SNPs in the region between *HTR1B* and *IRAK1BP1* in the locus on chromosome 6 that is adjacent to *COL12A1*, but the regional plot in Figure 2 suggests these two signals are independent. An additional result of interest in the list of suggestive associations is the cluster of SNPs in the *INSIG1* gene in chromosome 7. Figure 3 presents a regional association plot on chromosome 7 from ~154.8 to 155.2 Mb that shows enrichment of variants near *INSIG1* that correspond with two independent loci. One locus includes a group of less common SNPs (top SNP: rs141214584) and the uncommon allele C of rs141214584 is associated with

TABLE 3. Summary of SNPs that replicated in the FHS

SNP	Chr:Pos	CA	NCA	Gene	Discovery meta-analysis				FHS			
					Beta	SE	P	Direction	Beta	SE	P	CAF
rs114298934	2:106219192	a	c	FHL2   NCK2	0.168	0.042	6.11E-05	++	0.126	0.064	0.0486	0.06
rs34100300	5:176410753	a	g	UIMC1	-0.125	0.027	4.39E-06	-	-0.092	0.043	0.0336	0.22
rs71601327	5:176427065	t	c	UIMC1	0.114	0.027	2.12E-05	++	0.093	0.043	0.0310	0.78
rs35798539	5:176455269	a	g	ZNF346	-0.108	0.027	7.75E-05	-	-0.097	0.043	0.0244	0.22
rs352943 <sup>a</sup>	5:176464738	a	c	ZNF346	-0.107	0.027	8.62E-05	-	-0.097	0.043	0.0236	0.22
rs72774935	5:81148952	a	g	SSBP2   ATG10	-0.131	0.033	7.84E-05	-	-0.107	0.048	0.0264	0.87
rs3094005	6:31465047	a	c	MICB	0.144	0.037	9.26E-05	++	0.146	0.049	0.0031	0.12
rs9447453	6:75878425	a	c	COL12A1	0.140	0.036	8.42E-05	++	0.108	0.053	0.0409	0.11
rs13196892	6:7924126	a	c	TXNDC5   MUTED	-0.138	0.029	1.53E-06	-	-0.102	0.045	0.0229	0.16
rs6467223	7:128674666	a	g	TNPO3	-0.092	0.023	6.34E-05	-	-0.070	0.034	0.0390	0.34
rs9666274	11:19802659	a	g	NAV2	0.089	0.023	8.57E-05	++	0.080	0.040	0.0420	0.32
rs10766593	11:19804434	c	g	NAV2	0.093	0.023	4.45E-05	++	0.088	0.039	0.0240	0.34
rs1172821	19:55816678	t	c	BRSK1	0.103	0.022	3.90E-06	++	0.116	0.037	0.0019	0.37
rs1172822 <sup>a</sup>	19:55819845	t	c	BRSK1	0.100	0.022	6.05E-06	++	0.119	0.037	0.0013	0.37
rs4806660	19:55824634	t	c	TMEM150B	-0.102	0.023	6.24E-06	-	-0.119	0.036	0.0012	0.63
rs4806661	19:55824661	a	t	TMEM150B	-0.093	0.023	3.65E-05	-	-0.090	0.033	0.0064	0.60
rs35515212	19:55825755	a	g	TMEM150B	-0.094	0.023	3.05E-05	-	-0.090	0.033	0.0070	0.60
rs8107664	19:55825941	a	g	TMEM150B	0.093	0.023	4.30E-05	++	0.090	0.033	0.0070	0.40
rs34962991	19:55827175	a	g	TMEM150B	0.105	0.023	4.39E-06	++	0.116	0.037	0.0016	0.36
rs2384690	19:55828794	c	g	TMEM150B	-0.092	0.023	5.65E-05	-	-0.088	0.034	0.0099	0.60
rs28875253	19:55829214	a	c	TMEM150B	-0.092	0.023	5.55E-05	-	-0.088	0.034	0.0099	0.60
rs10420693	19:55830777	a	g	TMEM150B	0.097	0.023	1.86E-05	++	0.091	0.034	0.0077	0.40
rs10401201	19:55830871	t	c	TMEM150B	-0.094	0.023	3.76E-05	-	-0.090	0.034	0.0079	0.60
rs2384687 <sup>a</sup>	19:55831188	a	g	TMEM150B	-0.098	0.023	1.38E-05	-	-0.090	0.034	0.0078	0.60
rs897796	19:55831589	a	g	TMEM150B	-0.097	0.023	1.66E-05	-	-0.090	0.035	0.0089	0.61
rs1109368	19:55832567	a	g	TMEM150B	0.098	0.023	1.54E-05	++	0.091	0.034	0.0076	0.40
rs11668309 <sup>a</sup>	19:55833460	t	c	TMEM150B	0.115	0.023	4.87E-07	++	0.119	0.038	0.0015	0.36
rs11668344	19:55833664	a	g	TMEM150B	-0.114	0.023	5.47E-07	-	-0.119	0.038	0.0015	0.64
rs10425848	19:55834448	c	g	TMEM150B	-0.101	0.023	8.61E-06	-	-0.093	0.035	0.0072	0.60
rs11084393	19:55836010	t	c	TMEM150B	0.117	0.024	6.55E-07	++	0.120	0.038	0.0016	0.36
rs77101446	19:55836085	a	g	TMEM150B	0.091	0.023	5.32E-05	++	0.127	0.039	0.0010	0.53
rs2384686	19:55837531	a	g	TMEM150B	0.116	0.024	9.94E-07	++	0.123	0.039	0.0015	0.36
rs4806664	19:55840931	t	c	TMEM150B	-0.101	0.024	2.44E-05	-	-0.122	0.042	0.0033	0.64
rs34563859	19:55842314	c	g	TMEM150B	-0.120	0.024	8.72E-07	-	-0.110	0.042	0.0093	0.61
rs16991615 <sup>a</sup>	20:5948227	a	g	MCM8	-0.303	0.044	6.99E-12	-	-0.270	0.064	2.42E-05	0.07
rs7281846	21:15755349	t	c	HSPA13	0.090	0.023	8.46E-05	++	0.077	0.034	0.0254	0.31

CA, coded allele; CAF, coded allele frequency in the Framingham Heart Study (FHS); NCA, noncoded allele.

<sup>a</sup>These variants are published SNPs that reached genome-wide significance in Day et al.<sup>11</sup>

increased risk for menopause ( $P = 6.49E-06$ ) and hence earlier AOM, whereas a second group of more common SNPs (rs36196777 and rs6422756) is associated with later menopause.

## DISCUSSION

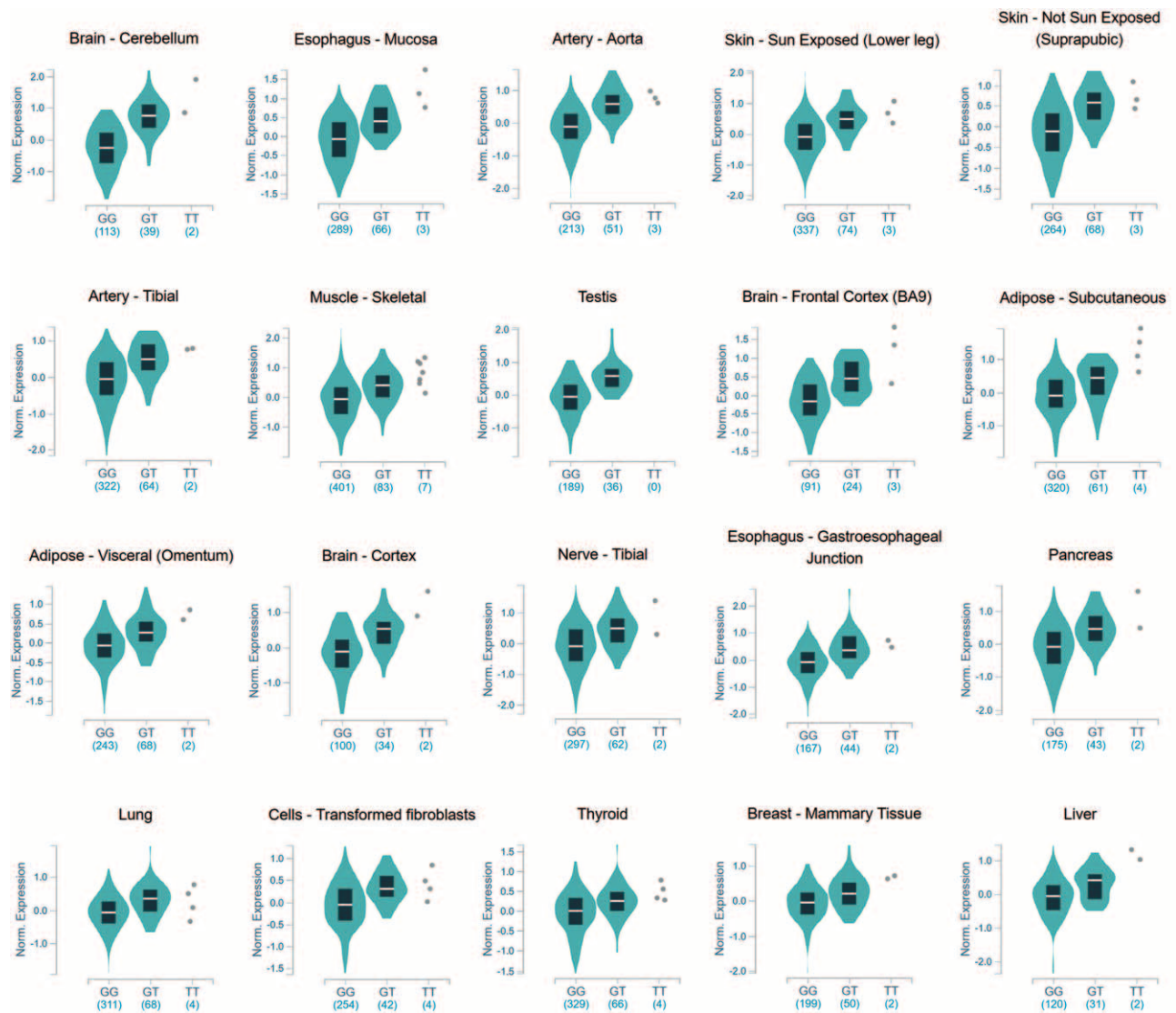
### Summary

We examined the associations of genetic variants with AOM in participants of two studies enriched for healthy agers. Meta-analysis of the two cohorts identified a genome-wide significant variant, which was previously known to be associated with delayed AOM. Several top SNPs in our analysis, although not genome-wide significant, replicated in the GWAS of AOM published by Day et al.<sup>11</sup> Moreover, we identified SNPs that associate with AOM, replicate in the FHS, and have not been previously reported for their association with AOM. These include rs3094005 in *MICB* (earlier AOM), rs9447453 in *COL12A1* (earlier AOM), rs72774935 located between genes *SSBP2* and *ATG10* (earlier AOM), rs13196892 located between genes *TXNDC5* and *MUTED* (delayed AOM), rs114298934 located between genes *FHL2* and *NCK2* (earlier AOM), rs6467223 in *TNPO3*

(delayed AOM), rs9666274 and rs10766593 in *NAV2* (earlier AOM), and rs7281846 in *HSPA13* (earlier AOM).

### Discussion of the genetic findings

Our analysis identified novel associations between AOM and variants of genes that have not been previously associated with this trait. The nine variants that replicated in the FHS, but have not been previously reported are rs3094005: *MICB*, rs13196892: *TXNDC5* | *MUTED*, rs72774935: *SSBP2* | *ATG10*, rs9447453: *COL12A1*, rs114298934: *FHL2* | *NCK2*, rs6467223: *TNPO3*, rs9666274 and rs10766593: *NAV2*, and rs7281846: *HSPA13*. *MICB* is one of the major histocompatibility complex class I chain-related genes and has been shown to be associated with multiple health conditions, including lupus,<sup>37</sup> dengue shock syndrome,<sup>38</sup> some cancers,<sup>39</sup> and a recent study reported that estrogen upregulates *MICB* expression levels in lung cancer.<sup>40</sup> In addition to *MICB*, rs3094005 is a significant eQTL of multiple genes that include *C4A*, *C4B*, *CCHCR1*, *CYP21A1P*, *FLOT1*, *NOTCH4*, and *PSORSIC1* across various tissue types. Genetic variations in *MICB* have been associated with susceptibility to cytomegalovirus.<sup>41</sup> There has been substantial interest in the role



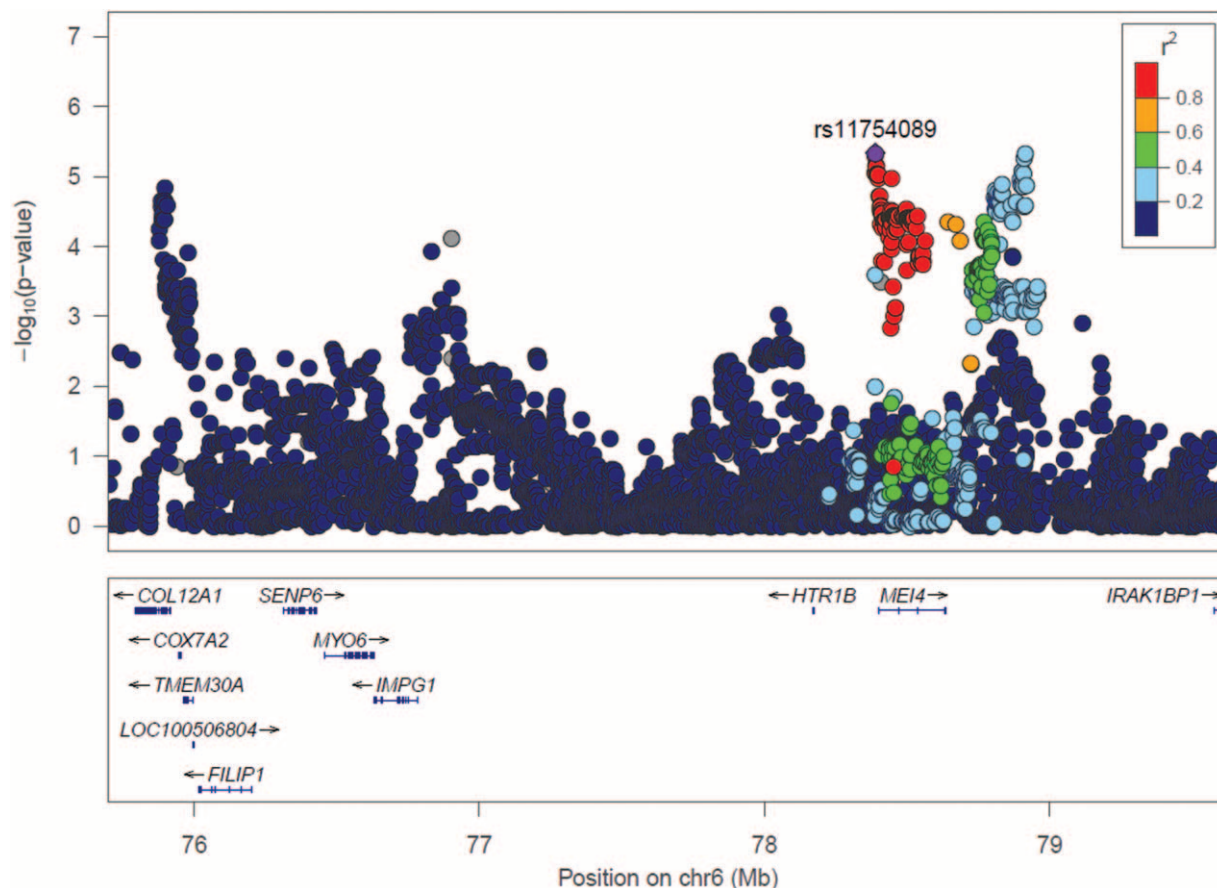
**FIG. 1.** Boxplot of gene expression levels of *MICB* in 20 tissues by genotype of rs3094005 using the genotype-tissue expression database. Reference allele G, alternative allele T. Note that boxplots for five tissues were not available.

of cytomegalovirus (CMV) in premature aging and in longevity.<sup>42</sup> In addition, two SNPs (rs9666274 and rs10766593) that replicated in the FHS are located in *NAV2*. A recent study revealed that several genetic variants in *NAV2* were associated with the risk and age at onset of Alzheimer's Disease.<sup>43</sup> For the remaining three SNPs, how functions of their nearby genes may influence premature or delayed AOM is not immediately clear. Changes of expression of *COL12A1* in bone tissues have been associated with pre- and postmenopause status,<sup>44</sup> but we were unable to determine an association between rs9447453 and *COL12A1* expression using GTEx.

We identified several previously unreported SNPs that replicated in Day et al dataset,<sup>11</sup> although they did not reach genome-wide significance. For example, the common variant of rs12448714 (meta-analysis HR = 0.9,  $P = 3.62 \times 10^{-5}$ ) was subgenome-wide significant ( $P = 6.70 \times 10^{-7}$ ) in the

Day et al data.<sup>11</sup> It is a significant eQTL of *RP11-830F9.5* in left ventricle and the common homozygote genotype is associated with increased expression of the gene. The SNP is also an eQTL for *TRAPPC2L* in brain (cerebellum), and *CBFA2T3* in skin and the uncommon allele (associated with earlier AOM) is associated with increased expression of these two genes. SNP rs1072218 is a significant eQTL of *VIT* in nerve (tibial), subcutaneous adipose, skeletal muscle, and transformed fibroblast cells. This gene is highly expressed in ovary tissue. Functional annotations with expression data from GTEx reveal rs6813104 is a significant eQTL of *EVC* in lung, rs6929983 is a significant eQTL of *PEX6* in esophagus, and rs2272347 is a significant eQTL of *IRF5* in multiple tissues.

The analysis identified multiple loci with robust associations that failed to replicate in the FHS and the GWAS of



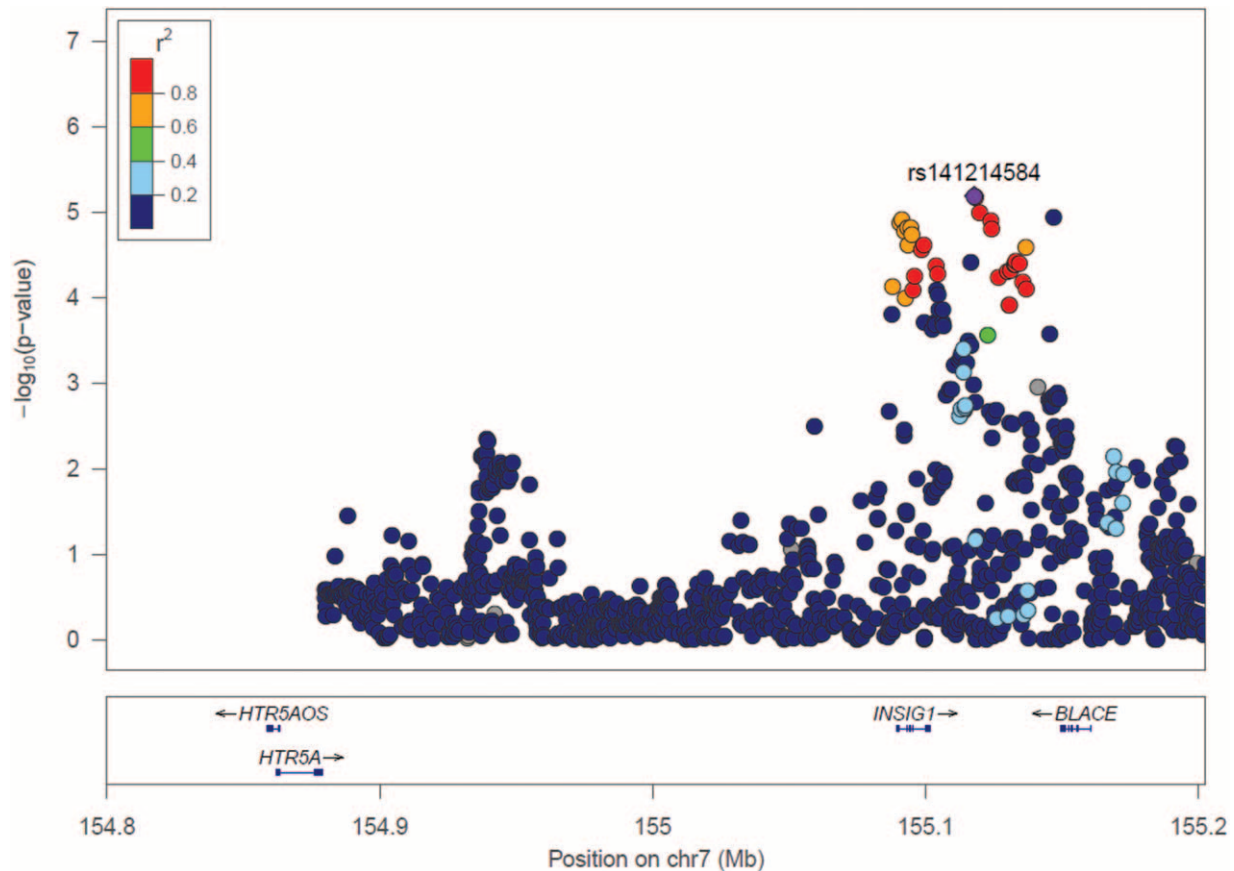
**FIG. 2.** Regional association plot on chromosome 6. The  $x$ -axis represents the genomic positions on chromosome 6 from  $\sim 75.7$  to  $79.6$  Mb, the  $y$ -axis on the left represents  $-\log_{10}$  of the  $P$  values from the meta-analysis, and the  $y$ -axis on the right represent the recombination rate. The colors show the strength of the linkage disequilibrium between the most strongly associated single nucleotide polymorphism (SNP) (rs11754089) and the SNPs studied.

AOM in Day et al.<sup>11</sup> Contrary to previously published results that focused on AOM in the general population, our analysis looked at AOM in a population enriched for longevity. Therefore, some novel associations may be specifically linked to healthy aging and longer survival. With this rationale, we were intrigued to see a cluster of association in proximity with the insulin-induced gene *INSIG1* (Fig. 3). This is an interesting gene because it is linked to insulin regulation, cholesterol metabolism, and lipogenesis all of which have been implicated in aging.<sup>45</sup> Variation in BMI and obesity are associated with AOM,<sup>17,46</sup> and life expectancy and the association found here, if replicated, could suggest that a potential relation between AOM and longevity is related to fat metabolism.

The study has some limitations. The overall size of the aggregated studies is small. We recognize that our study was somewhat underpowered as there were just two cohorts in our discovery set. In contrast to previous GWAS of AOM, the goal of the present study was, however, to discover additional variants associated with AOM in participants enriched for longevity, and this criterion limited the number of cohorts that were included in the present study. AOM is retrospectively determined in both the discovery and replication studies and, ideally, the onset of menopause should be determined

prospectively in the setting of a longitudinal study that follows people to extreme older age. We, however, do not see a reason for why there would be a consistent bias toward estimating AOM as premature or delayed in this study. As noted in Stolk et al.,<sup>8</sup> most studies determine AOM retrospectively. When the authors in the cited paper tested for differences in the effect sizes of the 20 most significant SNPs using retrospectively versus prospectively collected AOM in the FHS, they did not find any significant differences, suggesting that self-reported AOM would not introduce any systematic bias. Moreover, in a large-scale meta-analysis of 33 studies,<sup>11</sup> the self-reported AOM (40-60 y of age) was used.

Although we focused on AOM, some would argue that the total length of reproductive age would be a more informative phenotype to study.<sup>47</sup> Obtaining an accurate age of menarche from women who experienced this many decades ago is fraught with inaccuracy. Our previous work and that of others have not noted an association between healthy aging or prolonged survival and earlier menarche (which would increase the total number of reproductive capable years). Furthermore, as we hypothesize that later menopause may be associated with slower aging and decreased risk for aging-related diseases, we have focused on AOM.



**FIG. 3.** Regional association plot on chromosome 7. The  $x$ -axis represents the genomic positions on chromosome 7 from ~154.8 to 155.2 Mb, the  $y$ -axis on the left represents  $-\log_{10}$  of the  $P$  values from the meta-analysis, and the  $y$ -axis on the right represent the recombination rate. The colors show the strength of the linkage disequilibrium between the most strongly associated single nucleotide polymorphism (SNP) (rs141214584) and the SNPs studied.

We noted a SNP associated with AOM that has previously been associated with body mass index and we would have liked to include BMI data in this study; however, BMI data were not available for when study participants experienced menopause; another reason why a prospective study that includes when participants experienced menopause would be preferable. Another limitation is the ages achieved by the participants; most of the oldest generation in the LLFS were in their 90s with a few achieving ages 100+ years old. Perhaps if the sample was more enriched for even older survival, stronger and additional associations might have been noted. Lastly, virtually all of the significant SNPs identified in the present study are located in the noncoding regions, making it difficult to understand the casual mechanism by which these variants affect AOM. Although this study provides new interesting findings, much more work needs to be done to infer functional effects of these genetic variants.

### CONCLUSIONS

This work provides additional replication of associations between genetic variants in *MCM8*, *UIMC1*, *ZNF346*, *BRSK1*, and *TMEM224* and AOM, and suggestive associations of new variants with AOM. The findings provide further evidence for genetic basis of AOM. In addition, the discovery of new

variants in a study enriched of long-lived individuals suggests that there may be genetic mechanisms of AOM that are linked to human longevity. Replication of these findings in additional cohorts of individuals selected for longevity is needed.

### REFERENCES

1. Pelosi E, Simonsick E, Forabosco A, Garcia-Ortiz JE, Schlessinger D. Dynamics of the ovarian reserve and impact of genetic and epidemiological factors on age of menopause. *Biol Reprod* 2015;92:130.
2. Kato I, Toniolo P, Akhmedkhanov A, Koenig KL, Shore R, Zeleniuch-Jacquotte A. Prospective study of factors influencing the onset of natural menopause. *J Clin Epidemiol* 1998;51:1271-1276.
3. Perry JRB, Corre T, Esko T, et al. A genome-wide association study of early menopause and the combined impact of identified variants. *Hum Mol Genet* 2013;22:1465-1472.
4. Qin YY, Sun M, You L, et al. ESR1, HK3 and BRSK1 gene variants are associated with both age at natural menopause and premature ovarian failure. *Orphanet J Rare Dis* 2012;7:5.
5. Pyun JA, Kim S, Cho NH, et al. Genome-wide association studies and epistasis analyses of candidate genes related to age at menarche and age at natural menopause in a Korean population. *Menopause* 2014;21:522-529.
6. Stolk L, Zhai G, van Meurs JBJ, et al. Loci at chromosomes 13, 19 and 20 influence age at natural menopause. *Nat Genet* 2009;41:645-647.
7. He CY, Kraft P, Chen C, et al. Genome-wide association studies identify loci associated with age at menarche and age at natural menopause. *Nat Genet* 2009;41:724-728.
8. Stolk L, Perry JRB, Chasman DI, et al. Meta-analyses identify 13 loci associated with age at menopause and highlight DNA repair and immune pathways. *Nat Genet* 2012;44:260-268.

9. He CY, Kraft P, Chasman DI, et al. A large-scale candidate gene association study of age at menarche and age at natural menopause. *Hum Genet* 2010;128:515-527.
10. Wood MA, Rajkovic A. Genomic markers of ovarian reserve. *Semin Reprod Med* 2013;31:399-415.
11. Day FR, Ruth KS, Thompson DJ, et al. Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nat Genet* 2015;47:1294-1303.
12. Kinoshita Y, Johnson EM, Gordon RE, et al. Colocalization of MCM8 and MCM7 with proteins involved in distinct aspects of DNA replication. *Microsc Res Tech* 2008;71:288-297.
13. Willett W, Stampfer MJ, Bain C, et al. Cigarette smoking, relative weight, and menopause. *Am J Epidemiol* 1983;117:651-658.
14. Mckinlay SM, Bifano NL, Mckinlay JB. Smoking and age at menopause in women. *Ann Intern Med* 1985;103:350-356.
15. Sun L, Tan L, Yang F, et al. Meta-analysis suggests that smoking is associated with an increased risk of early natural menopause. *Menopause* 2012;19:126-132.
16. Luborsky JL, Meyer P, Sowers MF, Gold EB, Santoro N. Premature menopause in a multi-ethnic population study of the menopause transition. *Hum Reprod* 2003;18:199-206.
17. Hardy R, Mishra GD, Kuh D. Body mass index trajectories and age at menopause in a British birth cohort. *Maturitas* 2008;59:304-314.
18. Croft DP, Brent LJ, Franks DW, Cant MA. The evolution of prolonged life after reproduction. *Trends Ecol Evol* 2015;30:407-416.
19. Hawkes K. Human longevity: the grandmother effect. *Nature* 2004;428:128-129.
20. Hawkes K, O'Connell JF, Jones NG, Alvarez H, Charnov EL. Grandmothering, menopause, and the evolution of human life histories. *Proc Natl Acad Sci USA* 1998;95:1336-1339.
21. Hawkes K. Grandmothers and the evolution of human longevity. *Am J Hum Biol* 2003;15:380-400.
22. Perls TT, Alpert L, Fretts RC. Middle-aged mothers live longer. *Nature [Letter]* 1997;389:133.
23. Sun FG, Sebastiani P, Schupf N, et al. Extended maternal age at birth of last child and women's longevity in the Long Life Family Study. *Menopause* 2015;22:26-31.
24. Sebastiani P, Hadley EC, Province M, et al. A family longevity selection score: ranking sibships by their longevity, size, and availability for study. *Am J Epidemiol* 2009;170:1555-1562.
25. Newman AB, Glynn NW, Taylor CA, et al. Health and function of participants in the Long Life Family Study: a comparison with other cohorts. *Aging* 2011;3:63-76.
26. Sonnega A, Faul JD, Ofstedal MB, Langa KM, Phillips JW, Weir DR. Cohort profile: the Health and Retirement Study (HRS). *Int J Epidemiol* 2014;43:576-585.
27. Dawber TR, Kannel WB, Lyell LP. An approach to longitudinal studies in a community: the Framingham Study. *Ann N Y Acad Sci* 1963;107:539-556.
28. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families. The Framingham offspring study. *Am J Epidemiol* 1979;110:281-290.
29. Splansky GL, Corey D, Yang Q, et al. The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. *Am J Epidemiol* 2007;165:1328-1335.
30. Bae HT, Sebastiani P, Sun JX, et al. Genome-wide association study of personality traits in the long life family study. *Front Genet* 2013;4:65.
31. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009;5:e1000529.
32. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nat Methods* 2012;9:179-181.
33. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38:904-909.
34. Sebastiani P, Gurinovich A, Bae H, et al. Four genome-wide association studies identify new extreme longevity variants. *J Gerontol A Biol Sci Med Sci* 2017;72:1453-1464.
35. Willer C J, Li Y, Abecasis G R. METAL: fast and efficient meta-analysis of genome-wide association scans. *Bioinformatics* 2010;26:2190-2191.
36. Consortium G. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 2015;348:648-660.
37. Yu P, Zhu Q, Chen CJ, et al. Association between major histocompatibility complex class I chain-related gene polymorphisms and susceptibility of systemic lupus erythematosus. *Am J Med Sci* 2017;354:430-435.
38. Dang TN, Naka I, Sa-Ngasang A, et al. A replication study confirms the association of GWAS-identified SNPs at MICB and PLCE1 in Thai patients with dengue shock syndrome. *BMC Med Genet* 2014;15:58.
39. Zhao Y, Chen N, Yu Y, et al. Prognostic value of MICA/B in cancers: a systematic review and meta-analysis. *Oncotarget* 2017;8:96384-96395.
40. Ren J, Nie Y, Lv M, et al. Estrogen upregulates MICA/B expression in human non-small cell lung cancer through the regulation of ADAM17. *Cell Mol Immunol* 2015;12:768-776.
41. Ghosh M, Rodriguez-Garcia M, Wira CR. The immune system in menopause: pros and cons of hormone therapy. *J Steroid Biochem Mol Biol* 2014;142:171-175.
42. Aiello AE, Chiu YL, Frasca D. How does cytomegalovirus factor into diseases of aging and vaccine responses, and by what mechanisms? *GeroScience* 2017;39:261-271.
43. Wang KS, Liu Y, Xu C, Liu X, Luo X. Family-based association analysis of NAV2 gene with the risk and age at onset of Alzheimer's disease. *J Neuroimmunol* 2017;310:60-65.
44. Kosa JP, Balla B, Speer G, et al. Effect of menopause on gene expression pattern in bone tissue of nonosteoporotic women. *Menopause* 2009;16:367-377.
45. Dong XY, Tang SQ. Insulin-induced gene: a new regulator in lipid metabolism. *Peptides* 2010;31:2145-2150.
46. Akahoshi M, Soda M, Nakashima E, et al. The effects of body mass index on age at menopause. *Int J Obes Relat Metab Disord* 2002;26:961-968.
47. Bjelland EK, Hofvind S, Byberg L, Eskild A. The relation of age at menarche with age at natural menopause: a population study of 336 788 women in Norway. *Hum Reprod* 2018;33:1149-1157.