Published in final edited form as:

Neurobiol Aging. 2011 December ; 32(12): 2328.e1–2328.e9. doi:10.1016/j.neurobiolaging.2011.06.016.

TOMM40 poly-T repeat lengths, age of onset and psychosis risk in Alzheimer disease

Su Hee Chu, B.S.^a, Kathryn Roeder, Ph.D.^a, Robert E. Ferrell, Ph.D.^b, Bernie Devlin, Ph.D.^c, Mary Ann A. DeMichele-Sweet, Ph.D.^c, M.Ilyas Kamboh, Ph.D.^{b,c}, Oscar L. Lopez, M.D.^{c,d}, and Robert A. Sweet, M.D.^{c,d,e}

Su Hee Chu: suheechu@gmail.com; Kathryn Roeder: kathryn.roeder@gmail.com; Robert E. Ferrell: rferrell@pitt.edu; Bernie Devlin: devlinbj@upmc.edu; Mary Ann A. DeMichele-Sweet: sweetma@upmc.edu; M.Ilyas Kamboh: kamboh@pitt.edu; Oscar L. Lopez: LopezOL@upmc.edu; Robert A. Sweet: sweetra@upmc.edu

^aDepartments of Statistics, Carnegie Mellon University, Pittsburgh, PA, 15213, USA

^bDepartment of Human Genetics, University of Pittsburgh, Pittsburgh, PA, 15213, USA

^cDepartment of Psychiatry, University of Pittsburgh, Pittsburgh, PA, 15213, USA

^dDepartment of Neurology, University of Pittsburgh, Pittsburgh, PA, 15213, USA

^eVISN 4 Mental Illness Research, Education and Clinical Center (MIRECC), VA Pittsburgh Healthcare System, Pittsburgh, PA, 15206 USA

Abstract

Apolipoprotein E (*APOE*) $\epsilon 4$ alleles increase the risk for late-onset Alzheimer disease (LOAD) and decrease the age of onset. Recently, sequencing the *APOE* region in a small sample of LOAD subjects identified a variable length poly-T repeat sequence in the nearby gene, *TOMM40*, which may affect age of onset. We genotyped the *TOMM40* poly-T repeat using a novel statistical approach to refine the identification of allele length in 892 LOAD subjects and evaluated its

For questions and correspondence please contact: Robert A. Sweet, M.D., Mail: Biomedical Science Tower, Rm W-1645, 3811 O'Hara Street, Pittsburgh, PA 15213-2593, Express Mail: Biomedical Science Tower, Rm W-1645, Lothrop and Terrace Streets, Pittsburgh, PA 15213-2593, Phone: 412-383-8548, Fax: 412-624-9910, sweetra@upmc.edu, Web: <http://www.wpic.pitt.edu/research/sweetlab/>.

SUPPLEMENTAL DATA

The supplement contains the EM algorithm.

Disclosure Statement

The authors have no conflict of interest to report. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Verification

1. Author Disclosure

- 1a** The authors have no actual or potential conflicts of interest to report.
- 1b** The authors' institutions do not have contracts relating to this research through which it or any other organization may stand to gain financially now or in the future.
- 1c** The authors or their institutions have no other agreements that have a financial interest in this research.

2. Financial Support Disclosure: Supported in part by USPHS grants AG027224, AG030653, and AG05133 from the National Institute of Aging.

3. Data Verification: Data in this manuscript have not been submitted or published elsewhere, nor will they be while under review by Neurobiology of Aging

4. IRB Approval: All data collected in this study were obtained with protocols approved by the Institutional Review Board of the University of Pittsburgh.

5. Authors' Approval: All authors have reviewed this manuscript and agree with its content and the accuracy of the data.

effects on age of onset. Because psychosis in LOAD is a heritable phenotype which has shown conflicting associations with *APOE* genotype, we also evaluated the association of poly-T repeat length with psychosis. Poly-T repeat lengths had a trimodal distribution which differed between *APOE* genotype groups. After accounting for *APOE* $\epsilon 4$ there was no association of poly-T repeat length with age of onset. Neither *APOE* $\epsilon 4$ nor poly-T repeat length was associated with psychosis. Our findings do not support the association of poly-T repeat length with age of onset in LOAD. The clinical implications of this repeat length polymorphism remain to be elucidated.

Keywords

Apolipoprotein E (*APOE*) $\epsilon 4$; late-onset Alzheimer disease (LOAD); psychosis; *TOMM40*; variable length poly-T repeat sequence

INTRODUCTION

Late-onset Alzheimer disease (LOAD) is a neurodegenerative illness with substantial heritability (Gatz et al, 1997). Similarly, and not surprisingly, as the risk of LOAD increases in a highly age-dependent manner (Ferri et al, 2005), the age of onset of LOAD is also heritable (Pedersen et al, 2001). The gene with the most strongly established relationship to LOAD risk is apolipoprotein E (*APOE*), with increased risk of LOAD found in individuals carrying one or two copies of the $\epsilon 4$ allele (Farrer et al, 1997). The primary effect of *APOE* $\epsilon 4$ alleles on LOAD risk appears to be mediated via lowering the age of onset of Alzheimer disease, with a reduction of up to 7–9 years for each $\epsilon 4$ allele (Reitz and Mayeux, 2009).

The *APOE* $\epsilon 4$ allele has also been extensively examined for association with the presence of the psychotic phenotype of LOAD (LOAD+Psychosis, LOAD+P) (Sweet et al, 2003). LOAD+P is heritable (Bacanu et al, 2005; Sweet et al, 2010), and identifies a subgroup of LOAD subjects with more severe cognitive impairment and more rapid cognitive decline (Emanuel et al, 2011; Ropacki and Jeste, 2005). Unlike age of onset, the association of *APOE* $\epsilon 4$ allele with psychosis has revealed inconsistent findings, with slightly more negative than positive studies, and some studies showing evidence for a protective effect (DeMichele-Sweet and Sweet, 2010). Such a pattern could result solely from Type I error due to small cohorts with varying approaches to clinical characterization and analysis, however, a variable pattern of association can also arise due to a causal association with genetic variation in linkage disequilibrium with the *APOE* $\epsilon 4$ allele.

APOE $\epsilon 4$ is defined by a two SNP haplotype in *APOE* exon 4. SNPs rs429358 and rs7412 each code for either arginine (C) or cysteine (T). *APOE* $\epsilon 4$ alleles are the CC haplotype (with TT and TC defining $\epsilon 2$ and $\epsilon 3$ alleles, respectively). Recent investigations fine-mapping the region within and surrounding *APOE* on chromosome 19 identified a set of SNPs within the nearby gene, *TOMM40*, in linkage disequilibrium with the *APOE* $\epsilon 4$ allele (Yu et al, 2007) and affecting *APOE* expression (Bekris et al, 2010). This finding, in part, motivated an effort to sequence the *APOE* and *TOMM40* region in subjects with AD, in an effort to identify possible causal variants within the linked region (Roses et al, 2010). Sequencing identified a variable length poly-T repeat sequence in intron 6 of *TOMM40* that was in linkage disequilibrium with *APOE* $\epsilon 4$. Individuals with *APOE* $\epsilon 3/\epsilon 4$ genotype and long poly-T repeats (defined as ≥ 27) had significantly lower age of onset of LOAD than individuals with *APOE* $\epsilon 3/\epsilon 4$ genotype and short repeats (Roses et al, 2010).

Ultimately, reconciling the independent effects of *APOE* $\epsilon 4$ and *TOMM40* repeat length polymorphism on age of onset of LOAD will require concurrent genotyping of large numbers of subjects. To address this goal, we developed an approach to high throughput

genotyping of the *TOMM40* poly-T repeat length polymorphism by starting with PCR to generate an initial estimate of allele sizes and then refining these estimates with a statistical model. We evaluated the independent and joint effects of these genetic variants in a large population of 892 Caucasian individuals with LOAD, examining both the age of onset and LOAD+P phenotypes.

METHODS

Subjects

A total of 892 Caucasian, non-Latino subjects with a final diagnosis of possible or probable AD (McKhann et al, 1984), all evaluated at the University of Pittsburgh Alzheimer Disease Research Center (ADRC), were included. All subjects were assessed as described previously (DeMichele-Sweet et al, 2011). All data collected in this study were obtained with protocols approved by the Institutional Review Board of the University of Pittsburgh.

Assessment of Psychosis

Psychosis was evaluated with the CERAD behavioral rating scale (Tariot et al, 1995), as described previously (DeMichele-Sweet et al, 2011). Subjects were characterized as having no psychotic symptoms, a single psychotic symptom at only one time point, or multiple/recurrent psychotic symptoms, reflecting the increasing genetic loading associated with this hierarchy (Bacanu et al, 2005; Sweet et al, 2010). Finally, because the occurrence of psychosis is less frequent in the early stages of AD, subjects were required to have a Mini Mental State Exam (Folstein et al, 1975) score ≤ 20 in order to be classified as having LOAD without psychosis.

Genotyping

Assay—The *TOMM40* polymorphic repeat was genotyped by PCR amplification with forward primer 5'-VIC-GAGATGGGGTCTCACTATG-3' and reverse primer 5'-GTACAGGCCACAATGTG-3', with an initial 3 minute denaturation at 95°C, followed by 35 cycles of denaturation at 95°C for 30 sec., annealing at 56°C for 30 sec, and extension at 72°C for 30 sec. PCR was carried out in a final volume of 10 ul containing 10pM primers, 200uM dNTPs, 2 mM MgCl and 1U of Taq polymerase (Invitrogen). Fragments were resolved on an ABI 3730 automatic fragment analyzer, with a LIZ500 size standard, and fragment sizes were initially estimated using the output from GeneMapper v4.0 software (Applied Biosystems Inc.).

Preliminary allele calls—The *TOMM40* DNA sequence has an intronic poly-T (multiple thymine base pairs) that is highly variable at a population level. Our first statistical goal was to estimate the counts of T for each of the pair of alleles carried by subjects using the intensity signals obtained from the GeneMapper readout, where intensity (*I*) is some function of the number of times a particular length was replicated in the PCR process. The expected pattern for our method of measuring alleles would be to observe, for each allele, a maximum intensity of signal near the true count of T ($N[T]$), but distributed continuously with error around the true value, and PCR-based stutter around that peak that decays in intensity with increasing (integer) distance from $N[T]$. Thus, for an individual with two distinctly different poly-T alleles, a reasonable preliminary estimate can be obtained as the length associated with the maximum intensity of the two distinct peaks in an individual's GeneMapper readout (Figure 1). Note in Figure 1 the PCR stutter. Measurement error is illustrated by Figure 2. If an individual is homozygous for poly-T alleles, then the global maximum is a good estimator for both alleles.

Because of the PCR stutter and continuous measurements, we smoothed the sizing trace from each individual as follows. (1) Lay down a series of overlapping bins of size 1 on the size range 337–364 (min-max observed), which when accounting for the 320 bp flanking region maps onto 17 to 44 T repeats. Bins were separated by 0.1, such that the first bin was 337–338, the second bin was 337.1–338.1, and so on. (2) Find the set of bins, starting points separated by count 0.1, which contain the maximum intensity signal in those bins. This establishes a grid of bins. (3) If the alleles are well separated ($\geq 2T$) find the pair of bins with maximum intensity mass. A maximum is defined in the context of surrounding bins so that to the left of the maximum the bin mass is increasing and to the right it is decreasing (Figure 1, in which the dots represent the maxima I for each bin). If there were more than 2 maxima, the largest two were chosen as the alleles and the rest assumed to be stutter artifact (Figure 1). If there were only 2 maxima, and the average intensity of the smaller maxima $I < 300$, then this maximum was again assumed to be a stutter artifact and the genotype assumed to be homozygous. Likewise, if there were only one maximum, again the genotype was assumed to be homozygous. (4) Round the length or lengths to integer value.

Measurement Error: To assess measurement error, four samples showing different genotypes were measured five times. For each measurement we determined maxima as described above and determined the standard deviation of the maxima for each allele (Figure 2). We then evaluated whether the standard error was a function of repeat lengths, and could find no significant relationship. Counter to intuition, the error decreased slightly but non-significantly with increasing count of T. Thus we assumed the standard deviation of measurement is constant and equals the average over all repeated measured alleles, specifically 0.24. Note that roughly 95% of all maxima should be within rounding error of their true allele size.

Refining Preliminary Allele Calls: To further refine allele calls, we next applied least squares methods to the data. This approach can be broken down into four steps in which two key assumptions are made. Our first assumption is that the intensity obtained from the GeneMapper readout reflects the probability of the pair of alleles true sizes (Figure 1). Using the grid of bins established previously, we assume the relative intensity mass (relative to total intensity) in each bin is equivalent to the probability of an allele's true size. Our second assumption is that the distribution of probabilities of obtaining any lengths given a selected "true mean" is approximately normal. For the standard deviation of these normal distributions we use the standard error obtained from our measurement error analyses.

The steps in the least squares method follow: (1) Under the first assumption, create a probability mass function from the GeneMapper readout that yields discrete probabilities for each allele length which, when summed, is equal to 1. We will call this the stutter probability mass function (PMF). (2) Under the second assumption, find the probability mass function for each proposed possible true genotype, which also sums to 1. We will call these the normal PMFs. (3) Find the proposed genotype that minimizes the residual sum of squares of the two PMFs.

In the first step, in order to obtain the stutter PMF, we first found the sum of all the averaged binned intensities, and divided each binned intensity average by the total. This yielded the probabilities for each discrete integer value for length, and forms the stutter PMF (Figure 1).

In the second step, we first found all plausible pairs of alleles that could be the true genotype of the individual. Based on measurement error, these correspond to one repeat length to the left and right of each allele in the proposed genotypes defined by the identified maxima. All possible pairwise combinations of these alleles – following the rule of gametes – comprise the set of proposed genotypes. For each proposed genotype, the normal distribution for each

allele follows from its location (mean), and the standard deviation of measurement error. The height of the normal curve at each discrete length was treated as reflective of the probability of the allele's true size. We then found the sum of all the heights of the discrete integer lengths from 0 to 45, and divided each height by this sum. This yielded probabilities for each discrete integer value for length and forms the normal PMFs.

In our third step, we compare the two PMFs. Denote the stutter PMF as $f(z)$, and the normal PMF as $f(z^*)$. Find the genotype that minimizes the residual sums of squares (RSS).

$$RSS = \sum_{i=0}^{i=45} (f(z_i) - f(z_i^*))^2$$

Final Allele Calls: We used the expectation-maximization algorithm (Supplement) to arrive at the final genotype calls.

Associating allele calls to risk for AD and comorbid psychosis: In our modeling we had two response variables, age of onset of LOAD measured in years and LOAD+P (multiple/recurrent psychotic events, N=295 versus no psychotic events, N=324). Individuals with just one psychotic event were not included in the analyses. Predictor variables were the count of *APOE* $\epsilon 4$ alleles summary variables derived from the complex poly-T allele distribution. Specifically, to encode the poly-T genotype we followed *Roses et al.* (Roses et al, 2010) by creating a binary variable that had a division between small and long repeats (long > 27), referred to as the Roses encoding. For a richer predictor we defined an ordinal factor based on the poly-T allele distribution (Table 2). We analyzed each response variable by first constructing marginal models with each single predictor variable, and then constructing joint models using combinations of predictor variables that always included the *APOE* $\epsilon 4$ allele count. For age of onset, we used linear regression models and, for psychosis risk, we used conditional logistic regression models. All analyses were implemented in R.

RESULTS

Demographics and clinical characteristics for the 892 participants with *APOE* genotypes are shown in Table 1. All subjects had late-onset AD (age of onset ≥ 60); the vast majority (91.3%) was diagnosed with Probable AD.

The frequency distribution of repeat lengths over all the subjects as a whole (Figure 3) has several peaks in the distribution, consistent with published estimates (Roses et al, 2010). These peaks occur at approximately 17, 31, and 37, with diminishing frequencies of counts on either side of the peaks. The distribution is easily categorized to small, medium and long lengths of repeats. Based on these natural separations, we defined our poly-T variable with small repeats as being those of length ≥ 17 but < 25 , medium repeats as those ≥ 25 but < 34 repeats, and long repeats as those repeats ≥ 34 .

The distribution of repeat lengths of the $\epsilon 3/\epsilon 3$ carriers peaks at two locations, the most common allele length of 17, the other at allele length of 37, with respective relative frequencies being approximately 0.59 and 0.20. The $\epsilon 3/\epsilon 4$ carrier distribution of repeat lengths peaks in three places; 17, 31, and 37, with respective relative frequencies of 0.35, 0.26, and 0.13. For our $\epsilon 4/\epsilon 4$ carrier distribution a peak count of Ts falls at repeats 30 and 31, with approximate relative frequencies 0.53 and 0.30, respectively (Figure 4).

For age of onset, we fitted models with one predictor variable at a time and found that the count of *APOE* $\epsilon 4$ alleles in genotype, the Roses encoding of poly-T repeats, and the count of medium lengths in our ordinal factor variable were significant (Table 2). We then created joint models for age of onset in which the count of *APOE* $\epsilon 4$ alleles was always in the models. After accounting for the predictive value of *APOE* $\epsilon 4$ allele counts, neither of the poly-T predictors was significant (Table 3).

To predict LOAD+P, we again constructed marginal models with each predictor variable. Neither the Roses encoding nor the count of medium repeats in the ordinal factor variable were significant (Table 2). No association was noted for *APOE* $\epsilon 4$ (Table 2). The joint logistic models supported the findings of no association to psychosis risk (Table 3).

Finally, to mimic the analysis in Roses et al (Roses et al, 2010), we analyzed data restricted to carriers of the *APOE* $\epsilon 3/\epsilon 4$ genotype. We fit models to predict age of onset and LOAD+P with each independent variable, but none of the results were significant (Table 4). Therefore, we conclude that only the *APOE* polymorphism has an impact on age of onset and there is no association between *APOE* regional variation and LOAD+P.

DISCUSSION

We did not find that the *TOMM40* poly-T repeat length polymorphism is associated with age of onset in LOAD independently of the effect of *APOE* $\epsilon 4$. The lack of independent association was consistent whether categorizing the *TOMM40* poly-T repeat length polymorphism along the lines of its apparent trimodal distribution, or categorizing it as originally defined (Roses et al, 2010). Similarly, we found no association when confining our analysis to the additional impact of long *TOMM40* poly-Ts in subjects with the *APOE* $\epsilon 3/\epsilon 4$ genotype. The reason for the discrepancy between the current findings and those reported by Roses et al (Roses et al, 2010) is not known, but could include differences within the two subject pools in the nature of linkage disequilibrium between *APOE* and the poly-T repeat length polymorphism. Nevertheless the large size of our sample, including 405 *APOE* $\epsilon 3/\epsilon 4$ subjects in contrast to 34 *APOE* $\epsilon 3/\epsilon 4$ subjects in the original report, enhances confidence in the current findings.

We had hypothesized that conflicting findings regarding the association of *APOE* with LOAD+P could result from linkage disequilibrium with the intron 6 *TOMM40* poly-T repeat sequence, with the latter providing the true association with LOAD+P. The current findings do not support this hypothesis. Our findings do not preclude the possibility that other nearby genetic variation may contribute to the inconsistent associations between LOAD+P and *APOE*. Alternatively, the conflicting prior reports may have resulted due to differences in subject populations, sample sizes, definitions of LOAD+P, analytic approaches, or the observation that small, but real effects often show inconsistent results in studies with small sample sizes. In that regard, the current findings in a homogeneously assessed large cohort provide substantial evidence against any true association of *APOE* $\epsilon 4$ with LOAD+P. Similarly, we recently evaluated the association of *APOE* with LOAD+P in the Uniform Data Set collected by the National Alzheimer's Coordinating Center, comprising 2317 individuals with LOAD, 802 (34.6%) with psychosis, the largest cohort examined to date (DeMichele-Sweet et al, 2011). Once again *APOE* $\epsilon 4$ was not associated with LOAD+P.

Accurate assessment of the number of Ts in a polyT repeat of this size is challenging using standard molecular technology. Nonetheless we obtained very good precision and good accuracy by PCR amplification followed by size estimation from an ABI 3730 automatic fragment analyzer. Based on results from repeated assessment of the same samples we showed that 90–95% of these length estimates were within rounding error of their true size.

To make these estimates even more accurate we built a statistical model that essentially shrank the measurement error. In this way we developed a cost-effective and accurate means of determining poly-T genotypes at the *TOMM40* locus. Though our findings cast doubt on the importance of assessing the intron 6 poly-T genotype in *TOMM40* for LOAD genetics, nevertheless our method should have straightforward application to poly-T repeat sequences at other loci.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by National Institute on Aging (NIA) [Grants AG 027224, AG030653, and AG 05133].

Reference List

- Bacanu SA, Devlin B, Chowdari KV, DeKosky ST, Nimgaonkar VL, Sweet RA. Heritability of psychosis in Alzheimer disease. *Am J Geriatr Psychiatry*. 2005; 13:624–627. [PubMed: 16009739]
- Bekris LM, Galloway NM, Montine TJ, Schellenberg GD, Yu CE. APOE mRNA and protein expression in postmortem brain are modulated by an extended haplotype structure. *Am J Med Genet B Neuropsychiatr Genet*. 2010; 153B:409–417. [PubMed: 19554612]
- DeMichele-Sweet MA, Klei L, Devlin B, Ferrell RE, Weamer EA, Emanuel JE, Lopez OL, Sweet RA. No association of psychosis in Alzheimer disease with neurodegenerative pathway genes. *Neurobiol Aging*. 2011; 32:555.e9–11. [PubMed: 21093110]
- DeMichele-Sweet MA, Lopez OL, Sweet RA. Psychosis in Alzheimer disease in the National Alzheimer Disease Coordinating Center Uniform Data Set: Clinical correlates and association with apolipoprotein E. *International Journal of Alzheimer's Disease*. 2011 in press.
- DeMichele-Sweet MA, Sweet RA. Genetics of psychosis in Alzheimer's disease: a review. *J Alzheimers Dis*. 2010; 19:761–780. [PubMed: 20157235]
- Emanuel JE, Lopez OL, Houck PR, Becker JT, Weamer EA, Michele-Sweet MA, Kuller L, Sweet RA. Trajectory of cognitive decline as a predictor of psychosis in early Alzheimer disease in the cardiovascular health study. *Am J Geriatr Psychiatry*. 2011; 19:160–168. [PubMed: 20808116]
- Farrer LA, Cupples LA, Haines JL, Hyman B, Kukull WA, Mayeux R, Myers RH, Pericak-Vance MA, Risch N, van Duijn CM. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis APOE and Alzheimer Disease Meta Analysis Consortium. *JAMA*. 1997; 278:1349–1356. [PubMed: 9343467]
- Ferri CP, Prince M, Brayne C, Brodaty H, Fratiglioni L, Ganguli M, Hall K, Hasegawa K, Hendrie H, Huang Y, Jorm A, Mathers C, Menezes PR, Rimmer E, Sczufca M. Global prevalence of dementia: a Delphi consensus study. *Lancet*. 2005; 366:2112–2117. [PubMed: 16360788]
- Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975; 12:189–198. [PubMed: 1202204]
- Gatz M, Pedersen NL, Berg S, Johansson B, Johansson K, Mortimer JA, Posner SF, Viitanen M, Winblad B, Ahlbom A. Heritability for Alzheimer's disease: the study of dementia in Swedish twins. *J Gerontol A Biol Sci Med Sci*. 1997; 52:M117–M125. [PubMed: 9060980]
- 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; 34:939–944. [PubMed: 6610841]
- Pedersen NL, Posner SF, Gatz M. Multiple-threshold models for genetic influences on age of onset for Alzheimer disease: findings in Swedish twins. *Am J Med Genet*. 2001; 105:724–728. [PubMed: 11803520]
- Reitz C, Mayeux R. Endophenotypes in normal brain morphology and Alzheimer's disease: a review. *Neuroscience*. 2009; 164:174–190. [PubMed: 19362127]

- Ropacki SA, Jeste DV. Epidemiology of and risk factors for psychosis of Alzheimer's disease: a review of 55 studies published from 1990 to 2003. *Am J Psychiatry*. 2005; 162:2022–2030. [PubMed: 16263838]
- Roses AD, Lutz MW, mrine-Madsen H, Saunders AM, Crenshaw DG, Sundseth SS, Huentelman MJ, Welsh-Bohmer KA, Reiman EM. A TOMM40 variable-length polymorphism predicts the age of late-onset Alzheimer's disease. *Pharmacogenomics J*. 2010; 10:375–384. [PubMed: 20029386]
- Sweet RA, Bennett DA, Graff-Radford NR, Mayeux R. Assessment and familial aggregation of psychosis in Alzheimer's disease from the National Institute on Aging Late Onset Alzheimer's Disease Family Study. *Brain*. 2010; 133:1155–1162. [PubMed: 20147454]
- Sweet RA, Nimgaonkar VL, Devlin B, Jeste DV. Psychotic symptoms in Alzheimer disease: evidence for a distinct phenotype. *Mol Psychiatry*. 2003; 8:383–392. [PubMed: 12740595]
- Tariot PN, Mack JL, Patterson MB, Edland SD, Weiner MF, Fillenbaum G, Blazina L, Teri L, Rubin E, Mortimer JA, Stern Y. Behavioral Pathology Committee of the Consortium to Establish a Registry for Alzheimer's Disease. The behavior rating scale for dementia of the Consortium to Establish a Registry for Alzheimer's Disease. *Am J Psychiatry*. 1995; 152:1349–1357. [PubMed: 7653692]
- Yu CE, Seltman H, Peskind ER, Galloway N, Zhou PX, Rosenthal E, Wijsman EM, Tsuang DW, Devlin B, Schellenberg GD. Comprehensive analysis of APOE and selected proximate markers for late-onset Alzheimer's disease: patterns of linkage disequilibrium and disease/marker association. *Genomics*. 2007; 89:655–665. [PubMed: 17434289]

Figure 1. Count of PCR fragment sizes (intensity) versus size of fragment (top) and probability mass per integer bin versus size of fragments (bottom) for well-separated (left) and coalesced (right) distributions

Distributions of the measured intensity scores from the GeneMapper readout (top panels) and probability of an allele being a particular integer length of repeats (bottom panels). Note the dots in the bottom panels are separated by 1 unit, corresponding to 1 T, and represent the maximum intensity falling into bins of length 1 defined implicitly by the location of the dots.

Figure 2. Count of PCR fragment sizes (intensity) versus size of fragment
 GeneMapper readout for a sample that was measured five times (2A) and close-ups of the sample's first allele readouts (2B). Each read is slightly different due to measurement error, but the error is very small. This is further illustrated in 2B, where the dashed line occurs at the location of maximum intensity for the smaller allele and the first measurement. Note that the differences in readout are very small across all the measurements.

Figure 3. Estimated number of poly-T repeats in the full sample
The most common allele calls in these samples occur at lengths 17, 31, and 37.

Figure 4. Frequency and relative frequency of repeat lengths by APOE allele combinations
The frequency distributions (left panels) show the absolute counts of poly-T repeats per bin for each *APOE* genotype, and the relative frequency distributions (right panels) show the proportions (the counts of alleles divided by the total count of alleles, within each *APOE* genotype).

Table 1

Demographic and Clinical Characteristics of 892 Non-Latino Caucasian Subjects

Variable	N(%) or Mean (SD)
Age, years	84 (6.7)
Age of Onset, years	73 (6.3)
range	60–91
Sex	
male	296 (33.2)
female	596 (66.8)
Diagnosis	
probable Alzheimer Disease	814 (91.3)
possible Alzheimer Disease	78 (8.7)
MMSE Score and Severity Range	13.26 (5.9)
mild impairment (20–30)	113 (12.7)
moderate impairment (10–19)	544 (61)
severe impairment (0–9)	235 (26.3)
APOE ϵ4 genotypes	
2/2	1 (0.1)
2/3	27 (3.0)
2/4	25 (2.8)
3/3	355 (39.8)
3/4	405 (45.4)
4/4	79 (8.9)

MMSE: Mini Mental State Exam; *APOE*: Apolipoprotein E

Table 2

Results from models predicting age of onset and psychosis risk

Predictors	Levels	β	se(β)	p
<i>Age of Onset</i>				
number long ("Roses encoding") poly-Ts		-0.767	0.237	0.001
count <i>APOE</i> ϵ 4		-2.720	0.322	< 2e-16
count of repeats	medium	1.542	0.297	2.52e-07
	long	0.001	0.296	0.997
<i>LOAD+P</i>				
number long ("Roses encoding") poly-Ts		0.052	0.090	0.568
count <i>APOE</i> ϵ 4		0.060	0.125	0.628
count of repeats	medium	0.018	0.112	0.871
	long	0.125	0.114	0.274

Results were obtained by testing the predictive value of the independent variables one at a time.

APOE: Apolipoprotein E; *LOAD+P*: late-onset Alzheimer disease plus psychosis

Table 3

Results from joint models predicting age of onset and psychosis risk

Predictors	Levels	β	se(β)	P
<i>Age of Onset</i>				
count <i>APOE</i> ϵ 4		-2.660	0.343	<i>2.23e-14</i>
number long ("Roses encoding") poly-Ts		-0.126	0.244	0.605
count <i>APOE</i> ϵ 4		-2.833	0.439	<i>1.78e-10</i>
count of repeats	medium	-0.059	0.382	0.878
	long	-0.226	0.291	0.438
<i>LOAD+P</i>				
count <i>APOE</i> ϵ 4		0.039	0.133	0.766
number long ("Roses encoding") poly-Ts		0.041	0.097	0.672
count <i>APOE</i> ϵ 4		0.218	0.174	0.210
count of repeats	medium	0.145	0.151	0.339
	Long	0.141	0.115	0.221

Results were obtained by testing the predictive value of the independent variables after accounting for the effect of ϵ 4 count.

APOE: Apolipoprotein E; *LOAD+P*: late-onset Alzheimer disease plus psychosis

Table 4

Results from models predicting age of onset and psychosis risk analyzing only the data from subjects with the $\epsilon 3/\epsilon 4$ genotype

Predictors	Levels	β	se(β)	P
<i>Age of Onset</i>				
number long ("Roses encoding") poly-Ts		0.152	0.345	0.659
count of repeats	medium	-0.229	0.444	0.607
	long	0.059	0.484	0.902
<i>LOAD+P</i>				
number long ("Roses encoding") poly-Ts		0.042	0.141	0.765
count of repeats	medium	0.074	0.181	0.681
	long	0.179	0.194	0.354

Results were obtained by testing the predictive value of the independent variables one at a time.

LOAD+P: late-onset Alzheimer disease plus psychosis