SECTION EDITOR: DAVID E. PLEASURE, MD

An Inherited Variable Poly-T Repeat Genotype in *TOMM40* in Alzheimer Disease

Allen D. Roses, MD

I coauthored a recently published research article describing a variable length, poly-T polymorphism in the *TOMM40* gene, adjacent to apolipoprotein E *(APOE)* on chromosome 19, that accounts for the age at onset distribution for a complex disease, late-onset Alzheimer disease. These new data explain the mean age at disease onset for patients with the *APOE4/4* genotype and differentiate 2 forms of *TOMM40* poly-T polymorphisms linked to *APOE,* with each form associated with a different age at disease onset distribution. When linked to *APOE3* (encoding the ε3 isoform of *APOE*), the longer *TOMM40* poly-T repeats (19-39 nucleotides) at the rs10524523 (hereafter, 523) locus are associated with earlier age at onset and the shorter *TOMM40* 523 alleles (11-16 nucleotides) are associated with later age at onset. The data suggest that the poly-T alleles are codominant, with the age at onset phenotype determined by the 2 inherited 523 alleles, but with variable expressivity. Additional data will further refine the relationship between the length of the poly-T alleles and age at disease onset and determine if the relationship is linear.

Arch Neurol. 2010;67(5):536-541

Genetics has progressed remarkably fast and probably has affected diagnosis and treatment in neurology as significantly as in any other area of medicine except, perhaps, oncology. New drug treatments are more difficult and take longer to discover and develop in the field of neurology.[1] Much of the genetics of neurological diseases only recently has been learned, creating a long list of metabolic and structural mutations that seems to justify the point of view of "splitters" in the old wars of "lumping or splitting" similar clinical presentations. Several genetic mutations that explain most early-onset Alzheimer disease (AD) have been discovered.[2] Estimates of the genetic load for late-onset AD (LOAD) owing to *APOE4* (encoding the ε4 isoform of apolipoprotein E [GenBank AF261279]) range from 20% to 70%, with most investigators agreeing on approximately 50%.[2]

In the recent article describing the association between *TOMM40* (GenBank AF043250) poly-T polymorphisms and age at LOAD onset, a phylogenetic approach was used to examine the region around the *APOE* gene. Two related discoveries were made: age at onset for this late-onset, complex disease is determined by a traditional form of mendelian inheritance at the poly-T locus in *TOMM40*. The age at onset phenotype complexity is introduced by each polymorphic, poly-T allele rather than a myriad of small effects from many other genes. The phylogenetic method was critical for identifying the age at onset-associated polymorphism. The *TOMM40* gene is located next to the *APOE* gene in a region of strong linkage disequilibrium (LD). This means that through human evolution few recombination events have occurred between the 2 genes, and specific variations in each gene have been coinherited. However, additional mutations have accumulated within the LD region that contains *TOMM40* and *APOE*, providing genetic heterogeneity. The phylogenetic experiments and analyses leading to the discovery of the role of the *TOMM40* 523

Author Affiliations: Department of Neurobiology and Neurology and the Deane Drug Discovery Institute, Duke University, Durham, North Carolina.

Downloaded from www.archneurol.com at STANFORD Univ Med Center, on May 15, 2010
©2010 American Medical Association. All rights reserved.

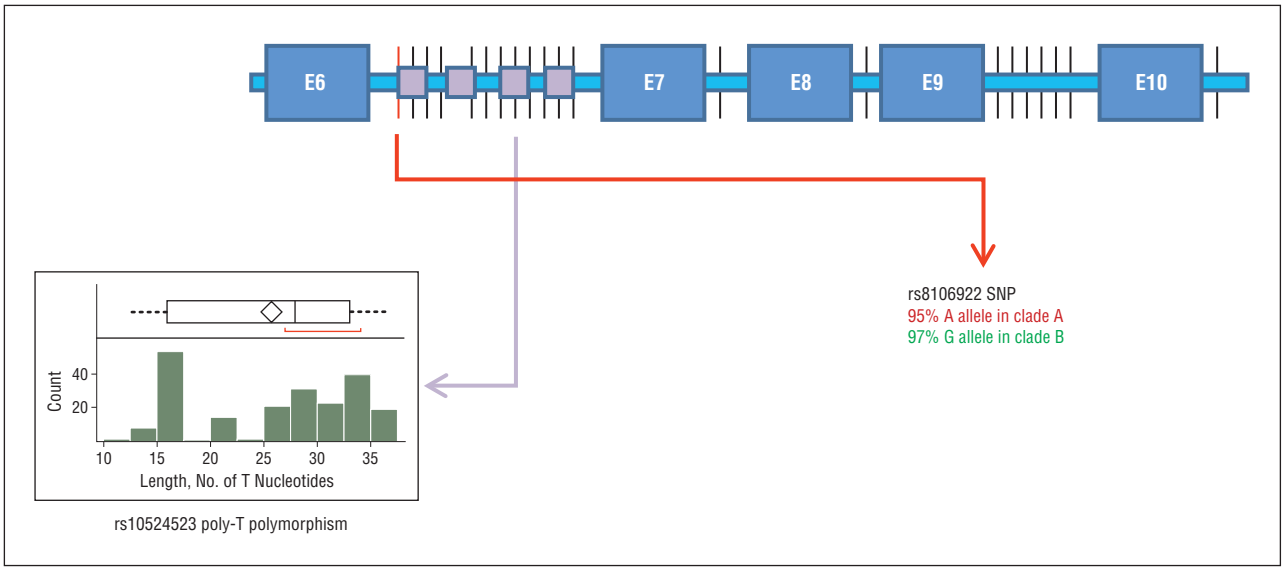

Figure 1. The 10-kilobase region of *TOMM40* included in the phylogenetic analysis. Shown are exons 6 through 10 (E6-E10) of *TOMM40* and intervening introns. Vertical lines indicate single nucleotide polymorphisms (SNPs), and small boxes indicate poly-T variants.

polymorphism are detailed elsewhere.[3] The purpose of this review is to explain the methods used, the data, and the clinical implications of the *TOMM40* 523 polymorphism for physicians and clinical scientists.

ANALYSIS OF THE SEQUENCING DATA: THE NEEDLE IN THE HAYSTACK

Figure 1 shows the location of many single nucleotide polymorphisms and at least 4 poly-T tracts in *TOMM40*. Located in intron 6 is 523, the variable poly-T repeat that is associated with age of LOAD onset. A histogram of the frequency of the different-length poly-T alleles (in numbers of T nucleotides) in a case-control population is shown. Also indicated in Figure 1 is rs8106922 (hereafter, 922), which is the single nucleotide polymorphism that, in our analyses, anchors the split between 2 groups of evolutionarily related sequences that comprise the branches designated as clades A and B (**Figure 2**). The initial search for the potentially functional location in the LD region that comprises *APOE, TOMM40,* and *APOC1* was guided by previous biological data.[4-6] The goal was to see if one of the clades was enriched for patients with LOAD. Within this LD region, polymorphisms within a specific 10-kilobase piece of *TOMM40* supported the formation of a statistically robust phylogenetic tree that enriched LOAD cases into one of the clades. There were 2 patients for every control in the study shown in Figure 2. In clade A, however, the patient-control ratio was increased to 2:7, whereas in clade B the ratio decreased to 1:7 (Figure 2A). No other region that was analyzed supported a robust phylogenetic tree structure and did not enrich for patients into 1 clade. When the distribution of *APOE* genotypes across the clades was examined, almost all *APOE4/4* resolved to clade A, whereas *APOE3/4* and *APOE3/3* were present in both clades. The data from the first case-control series analyzed in this manner were presented at the IX International Meeting on Human Genome Variation and Complex Genome Analysis (September 7, 2007). There was a strong hint that something was going on in this region, which happened to be located in the intronic sequence of *TOMM40*, not *APOE*.

Figure 2 illustrates the data from the second of the 2 independent, case-control series that were analyzed.[3] The samples for this analysis were provided by the Arizona Alzheimer Disease Research Center. There were 105 individuals, or 210 chromosomes, in the phylogenetic analysis with approximately 2 patients with LOAD for each cognitively normal, *APOE* genotype–matched, and age- and sex-matched control. The polymorphisms that were analyzed were identified by deep Sanger sequencing of the region of interest from all copies of chromosome 19 in the analysis. The data in Figure 2 are identical but with different information superimposed for explanatory purposes.

In Figure 2, beginning with the single nucleotide polymorphism 922, the vertical lines represent successive mutation events in the parent sequence that create branch points in the phylogenetic tree. The 2 alleles of the single nucleotide polymorphism, 922, most effectively distinguish the 2 main branches of the phylogenetic tree; the A allele is shared by all sequences in clade A and the G allele is shared by all sequences in clade B. This is an unrooted phylogenetic tree; it has not been rooted by including an obvious ancestral sequence that would help to determine the absolute chronology of mutational events. Therefore, in this article, use of terms that imply timing refer to the apparent order of events according to this tree structure and not the absolute timing of mutational events. However, even in the absence of absolute chronology, comparison of divergent clades that enrich for a phenotypic variable, such as age at LOAD onset, provides an opportunity to screen the clades for mutations related to that disease. Some of the branch-point mutations of the phylogeny are unique to the subsequent branch (clade) and can be used to differentiate one clade from the other, whereas other mutations appear in more than 1 branch of the tree. Each horizontal line in Figure 2 indicates a sequence and a stack of horizontal lines indicates a sequence that is shared by multiple individuals. Figure 2 illustrates the point that mutations occur on chromo-

Downloaded from www.archneurol.com at STANFORD Univ Med Center, on May 15, 2010
©2010 American Medical Association. All rights reserved.

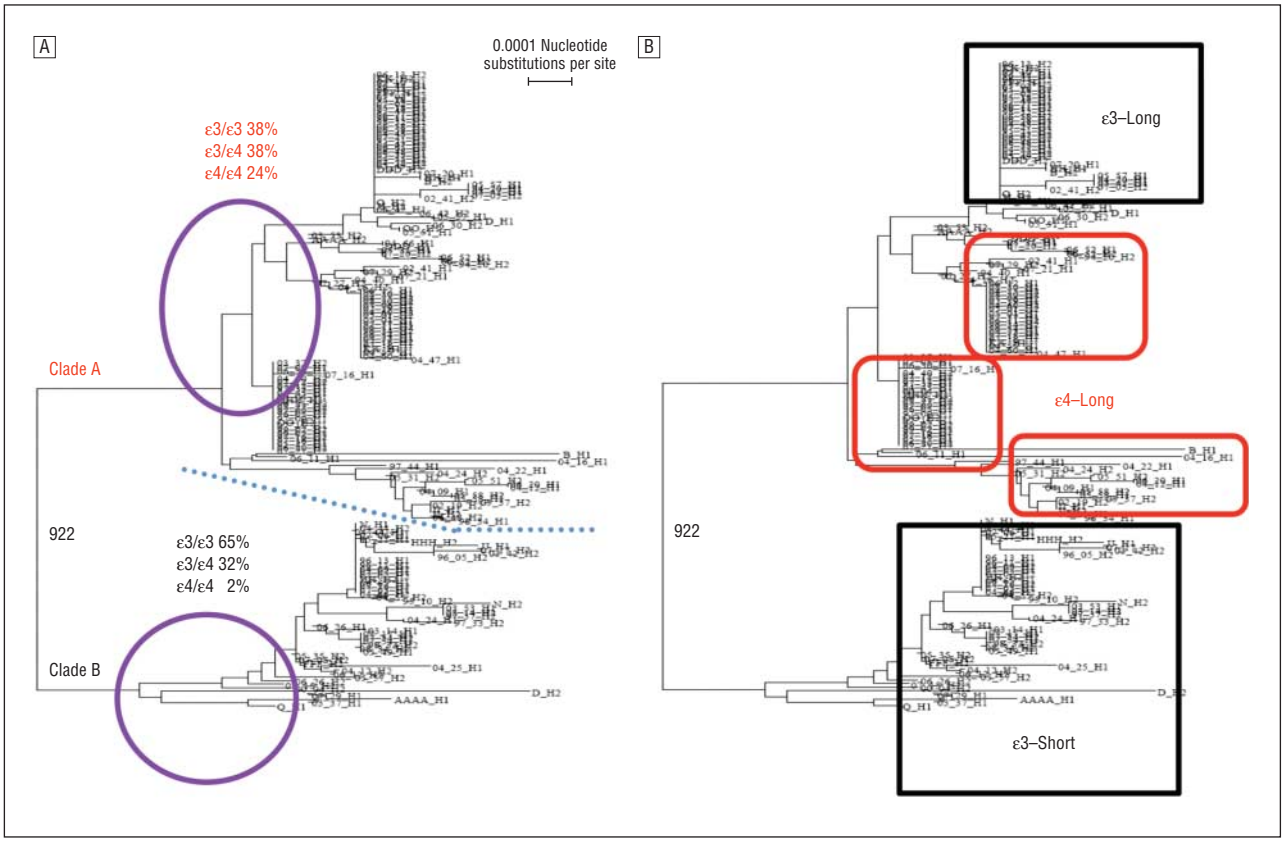

Figure 2. Phylogenetic tree constructed for the region of interest in *TOMM40*. A, The first major branch point in this tree, at rs8106922 (ie, 922) is indicated. The ovals indicate subsequent branch points (mutation events). Frequencies of apolipoprotein (*APOE*) genotypes in clades A and B are also indicated. The dotted line highlights the separation between the 2 major clades. B, The clades that are differentiated by length of the poly-T allele (ie, 523). *APOE3* can be linked to long or short 523 poly-T alleles (black boxes). *APOE4* is connected to long poly-T alleles (red boxes).

somes in the context of earlier mutations, giving rise to sequence diversity even within regions of strong LD. The ovals in Figure 2A enclose mutations that introduce sequence diversity and divide the population into related, but distinct, clades. Terminal clades are enclosed by boxes in Figure 2B. Each terminal clade has a differential mutational history. Identifying a collection of phased mutations or polymorphisms within a region of strong LD, as represented in Figure 2, typically exceeds the capacity of genome-wide association studies. In addition, because many of these multiple mutations are uncommon, they are not adequately assayed in genome-wide association studies.

As described previously, clade A on the phylogenetic tree constructed from this region of *TOMM40* was enriched for patients with LOAD. Further analysis showed that this enrichment was at least in part owing to an unequal distribution of individuals with the *APOE4/4* genotype into clade A; 24.3% of the genotypes in clade A were *APOE4/4*, whereas *APOE4/4* represented only 3.3% of clade B. The *APOE3/4* genotype was evenly distributed between the 2 clades, and *APOE3/3* was more prevalent in clade B.

An attempt was made to develop a complex algorithm, based on SNPs that differentiated clades A and B, which would estimate risk of LOAD. This proved difficult because some mutations were shared by multiple terminal clades, although on apparently different sequence backgrounds. The task became much simpler when the polymorphism that differentiated all the stacks of common haplotypes, enclosed by the boxes in Figure 2B, was identified. It was surprising to discover that each of these boxes was distinguished by the length of a polymorphic, poly-T locus, 523. Although the poly-T tract was approximately the same length for all haplotypes within each box, the average length of the poly-T tract was significantly different among the boxes. When the *APOE* background for each poly-T length allele was investigated, it became obvious why *APOE4* had been identified as the risk gene for LOAD: *APOE4* was virtually always connected to a long poly-T variant. Conversely, short poly-T variants were virtually always on *APOE3* strands. However, some *APOE3* alleles were connected to long poly-T variants.

Figure 3 shows histograms of the lengths of the 523 mutations on each *APOE* backbone for all individuals (patients with AD and controls) in this experiment. Looking first at the *APOE* backbones, all the poly-T lengths were greater than 19 T nucleotides and were typically 22 to 29 T nucleotides long. Looking next at the *APOE3* backbones, there were clearly 2 separate poly-T length groups in relatively similar proportions. The short poly-T lengths were 11 to 16 T nucleotides, and the longer poly-T lengths were 29 to 39 T nucleotides. Thus, *APOE3* alleles can now be characterized on the basis of being connected to either a long or short poly-T.

Further analyses demonstrated that all of the 523 short repeats always separated into clade B, whereas the long

Downloaded from www.archneurol.com at STANFORD Univ Med Center, on May 15, 2010
©2010 American Medical Association. All rights reserved.

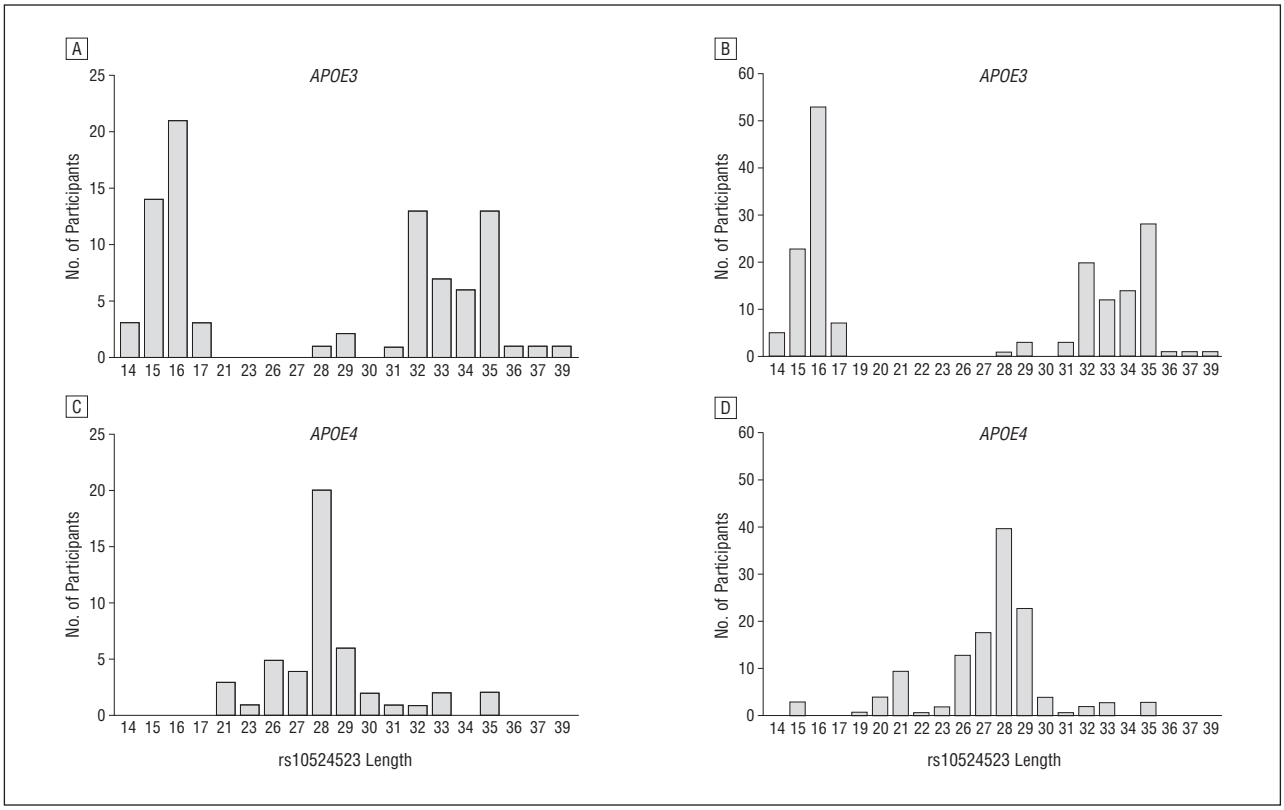

Figure 3. Frequency distributions of lengths of the poly-T tracts at rs10524523 (ie, 523) length (No. of T residues). Data are from 2 independent late-onset Alzheimer disease case-control populations (population 1, panels A and C; population 2, panels B and D). The frequency with which each 523 allele was connected to either *APOE3* (A and B) or *APOE4* (C and D) is shown. Two (rare) instances of *APOE4* linked to a short 523 allele (length=15 T residues) (D) are also shown. E3 and E4 indicate *APOE3* and *APOE4*, respectively.

poly-T length repeats, regardless of whether they were linked to *APOE3* or *APOE4*, segregated to clade A. *APOE4* occurred in clade A 97.8% of the time. Thus, looking at the precise distribution of poly-T lengths in geographically determined population studies of thousands of individuals will be much more informative than examining *APOE* status.

After confirming the phylogenetic structure of this region of *TOMM40* and the distribution of the *APOE* and 523 genotypes on the tree, the association between 523 and the central endophenotype for LOAD, age at disease onset, was examined. This question was addressed using a separate population of patients with LOAD who had the *APOE3/4* genotype and asking whether there was a difference in mean age at onset for carriers of the *APOE3* long 523 vs *APOE3* short 523 haplotype. The DNA and age at onset data for the well-characterized patients were obtained from the Duke Bryan Alzheimer Disease Research Center. All patients carried 1 *APOE4* long 523 haplotype. Those patients who also carried a short 523 allele connected to *APOE3* (n=5) had a mean age at onset of 78 years (**Figure 4**). Those patients who carried a second long 523 allele (n=29) connected to *APOE3* had a mean age at onset of 70 years. This is similar to the age, first determined by Corder et al[7] and verified during the last 17 years, at which 50% of *APOE4/4* individuals have diagnosed disease. Analysis of the age at onset for a limited number of *APOE3/3* patients suggests that there is earlier onset of cognitive impairment or LOAD for patients carrying long 523 alleles, consistent with observations for the *APOE3/4* patients. The data discussed herein are from case-control cohorts rather than an epidemiologic study. Conclusions about population risk will have to wait for analyses of prospective, observational series of aged individuals. Several prospective studies of LOAD have been ongoing for 4 to 18 years and the plan

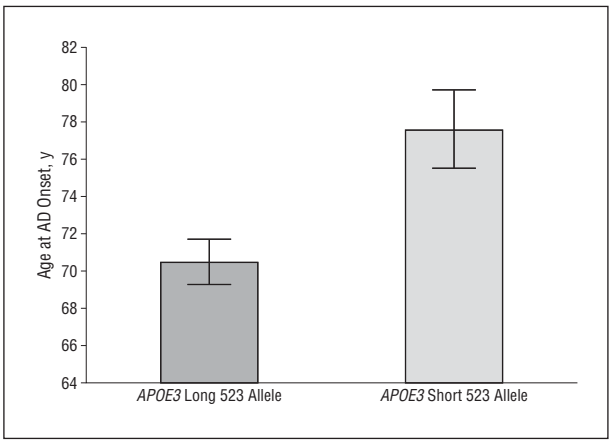

Figure 4. The mean (SD) age at onset of late-onset Alzheimer disease (AD) for *APOE3/4* (*APOE3/4*) patients carrying different 523 length polymorphisms connected to *APOE3*. All *APOE4* alleles are connected to long 523 alleles; the patients in this cohort differ by the length of the 523 allele connected to *APOE3*. The mean age at disease onset was significantly younger for individuals carrying *APOE3* long 523 alleles than for patients carrying *APOE3* short 523 alleles (approximately 70 years vs approximately 78 years; *P*=.03). These data are from a cohort of 34 white patients with *APOE3/4* and late-onset AD for whom age at disease onset was accurately documented.

Downloaded from www.archneurol.com at STANFORD Univ Med Center, on May 15, 2010
©2010 American Medical Association. All rights reserved.

is to genotype the 523 and *APOE* loci for patients who develop LOAD during the course of the studies and for those individuals who remain free of disease according to accepted neuropsychological criteria.

COMMENT AND CLINICAL IMPLICATIONS

The phylogenetic analysis of the *APOE* LD region indicates that each strand of the pair in every person has the capacity to add great phenotypic heterogeneity even within a relatively homogeneous group such as whites. In the case of LOAD, phenotypic complexity is measured as age at onset, obscuring the mendelian inheritance of the 523 locus and presenting as a late-onset, complex disease. The variable repeat that is inherited from each parent provides the phenotypic complexity measured by age at onset.

These new data inspire a new question: Is *APOE* the risk gene or is the disease-causing locus the 523 poly-T in *TOMM40*? Genetic parsimony and Occam's razor would rest with 523 as the risk allele and *APOE4* being identified because it is almost always connected to a long 523 poly-T allele. The genetic inheritance of the disease would now seem to be clearer than the biology.

For a drug discovery program, more than gene identification is relevant; critical for LOAD is to understand its pathogenic mechanisms and triggering events. It is known that apoE protein interacts with the mitochondrial outer membrane and that different apoE isoforms interact differently with and cause differential effects on the dynamics and function of neuronal mitochondria, leading to differences in neurite outgrowth.[8-11] There is, therefore, likely also a role for *APOE* in disease pathogenesis. The proteins apoE and Tom40 are implicated in mitochondrial toxicity early in the proposed pathogenic process of AD, and perturbation of the normal activities of these proteins can explain the formation of amyloid plaques and other toxic protein aggregates. The so-called amyloid cascade exists and may exacerbate pathogenesis once initiated. However, early toxicity is likely initiated by apoE-Tom40 protein interactions, causing the release of cytochrome C from damaged mitochondria with subsequent apoptosis and initiation of the amyloid cascade.[12,13] Targeting the long-term poisoning of mitochondria and increased apoptosis would seem to be a good way to delay the age at onset and interrupt the course of AD: the best place to turn off the water in a cascade is to block the source.

The variable-length, poly-T polymorphism appears to be stably inherited through human evolution rather than representing a site for sporadic, recurrent, contemporary mutational events. The phylogenetic approach identifies stable length mutations that occur on different evolutionary mutational backgrounds. The situation is, therefore, unlike the dominant triplet repeat diseases that have variable expressivity, such as myotonic muscular dystrophy, where the number of repeats is unstable through successive generations. Thus, it would seem that the 523 polymorphism could be used in a LOAD age at onset predictive test. On the basis of the age of the individual (between the ages of 60 and 87 years) and his or her 523 genotype, it is proposed that a risk estimate (ie, high or low) can be made for the onset of mild cognitive impairment symptoms and conversion to AD during the next 5 to 7 years (ie, not lifetime risk).

Replication and validation of the relationship between 523 and age at LOAD onset are necessary before the marker is used in clinical practice. For that reason, a prospective clinical study has been designed that combines validation of the genetic marker with assessment of the efficacy of a safe drug for delay of disease onset.[14] A number of considerations regarding this clinical study have been discussed with scientists at the Food and Drug Administration via the Voluntary Exploratory Data Submission mechanism. Discussions with pharmaceutical companies have already been initiated by Zinfandel Pharmaceuticals Inc (Durham, North Carolina), a "virtual" drug development company organized to design, accelerate, and manage the clinical trial. During this prospective study, the poly-T assay will be developed into a Food and Drug Administration–qualified test. Even with the improved estimate of age at onset risk provided by the genetic marker, a study of this nature will take 5 or more years to complete.

Although the choice was made not to commercialize this test for clinical use until after validation, the test will be available for sponsored academic research. In parallel, licenses for commercial studies will be negotiated because the marker will have application in stratification of results from ongoing clinical trials of AD interventions and prospective stratification of studies. Given the range of *APOE* allele frequencies across ethnic groups, it is likely that different ethnic groups will have different allele frequencies for the poly-T variant as well. Ethnic diversity is not a problem if anticipated, and primary phylogenetic studies will be necessary to describe any differences in evolution of the LD region and to calculate the allele frequencies.

I anticipate, especially based on my early 1990s experience with the spirited debate around the diagnostic use of *APOE* genotyping in AD, that there will be differences of opinions on various ethical, legal, and social issues. I have therefore already spent more than a year in consultation with a panel of external, worldwide ethical, legal, and social issues experts who helped me develop the clinical plan.

There are many unanswered questions that will be addressed in the coming years, and one of them refers to the genetics of other so-called complex diseases. It is reasonable to consider that other small structural polymorphisms not detected by genome-wide association studies could be responsible for the variable expression of other complex diseases. These loci could be identified using targeted phylogenetic strategies. The work described in this review demonstrates the complexity of the genotype-phenotype relationship and is a reminder that we researchers are just beginning our journey to understand the human genome.

Accepted for Publication: January 21, 2010.
Correspondence: Allen D. Roses, MD, Deane Drug Discovery Institute, 1 Science Dr, Ste 342, Campus Box 90344, Durham, NC 27708 (allen.roses@duke.edu).

Downloaded from www.archneurol.com at STANFORD Univ Med Center, on May 15, 2010
©2010 American Medical Association. All rights reserved.

Financial Disclosure: Dr Roses is the president of 3 companies filed as S Corporations in North Carolina: Cabernet Pharmaceuticals is a pipeline pharmacogenetic consultation and project management company that has other pharmaceutical companies as clients; Shiraz Pharmaceuticals Inc is focused on the commercialization of diagnostics, including companion diagnostics, for universities, pharmaceutical companies, and biotechnology companies; and Zinfandel Pharmaceuticals Inc is the sponsor of Opportunity to Prevent Alzheimer's Disease, which is a combined clinical validation of a diagnostic and a pharmacogenetics-assisted delay-of-onset clinical trial. These companies are independent of Duke University, but the diagnostic intellectual property generated by Dr Roses or his team is intended to be treated as Deane Drug Discovery Institute property once there is an established commercial value.

Funding/Support: The work described in this article would not have been possible without the generous contribution of DNA samples from the Netherlands Brain Bank (under the direction of Rivka Ravid, PhD), the Banner Sun Health Research Institute (under the direction of Thomas Beach, MD, PhD), the Arizona Alzheimer's Disease Core Center, and the Joseph and Kathleen Bryan Alzheimer's Disease Research Center. Work at the Arizona Alzheimer's Disease Care Center was supported in part by grants P30 AG019610 and R01 AG031581 from the National Institute on Aging (Eric Reiman, MD); grant R01 NS059873 from the National Institute of Neurological Disorders and Stroke and Science Foundation Arizona (Matthew Huentelman, PhD); the Arizona Alzheimer's Consortium; and the state of Arizona. The work at the Joseph and Kathleen Bryan Alzheimer's Disease Research Center was supported in part by grant P30 AG028377 from the National Institute on Aging (Kathleen Welsh-Bohmer, PhD). Dr Roses is supported in part by grant RC1 AG035635-01 from the National Institute on Aging.

Previous Presentation: This study was presented in part at the IX International Meeting on Human Genome Variation and Complex Genome Analysis; September 7, 2007; Barcelona, Spain.

REFERENCES

1. Eastman P. Safe ways to speed CNS drugs to patients, including provisional approval. *Neurology Today*. 2009;8(7):6. doi:10.1038/tpj.2009.69.
2. Avramopoulos D. Genetics of Alzheimer's disease: recent advances. *Genome Med*. 2009;1(3):34. doi:10.1186/gm34.
3. Roses AD, Lutz MW, Amrine-Madsen H, et al. A TOMM40 variable length polymorphism predicts the age of late-onset Alzheimer's disease. *Pharmacogenomics J*. http://www.nature.com/tpj/journal/vaop/ncurrent/abs/tpj200969a.html. In press.
4. Anandatheerthavarada HK, Biswas G, Robin M-A, Avadhani NG. Mitochondrial targeting and a novel transmembrane arrest of Alzheimer's amyloid precursor protein impairs mitochondrial function in neuronal cells. *J Cell Biol*. 2003;161(1):41-54.
5. Devi L, Prabhu BM, Galati DF, Avadhani NG, Anandatheerthavarada HK. Accumulation of amyloid precursor protein in the mitochondrial import channels of human Alzheimer's disease brain is associated with mitochondrial dysfunction. *J Neurosci*. 2006;26(35):9057-9068.
6. Liang WS, Reiman EM, Valla J, et al. Alzheimer's disease is associated with reduced expression of energy metabolism genes in posterior cingulate neurons. *Proc Natl Acad Sci U S A*. 2008;105(11):4441-4446.
7. Corder EH, Saunders AM, Strittmatter WJ, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science*. 1993;261(5123):921-923.
8. Brodbeck J, Balestra M, Freedman S, Weisgraber KH, Mahley RW, Huang Y. Apolipoprotein E4 and its fragment impair mitochondrial dynamics in neuronal cultures. *Alzheimers Dement*. 2008;4(4)(suppl 1):T163.
9. Brodbeck J, Balestra ME, Saunders AM, Roses AD, Mahley RW, Huang Y. Rosiglitazone increases dendritic spine density and rescues spine loss caused by apolipoprotein E4 in primary cortical neurons. *Proc Natl Acad Sci U S A*. 2008;105(4):1343-1346.
10. Brodbeck J, Miranda RD, Freedman SB, Weisgraber KH, Mahley RW, Huang Y. Apolipoprotein E4 impairs mitochondrial function and increases reactive oxygen species in neuronal cells. *Alzheimers Dement*. 2009;5(4)(suppl 1):307. doi:10.1016/j.jalz.2009.04.469.
11. Chang S, ran Ma T, Miranda RD, Balestra ME, Mahley RW, Huang Y. Lipid- and receptor-binding regions of apolipoprotein E4 fragments act in concert to cause mitochondrial dysfunction and neurotoxicity. *Proc Natl Acad Sci U S A*. 2005;102(51):18694-18699.
12. Goedert M, Spillantini MG. A century of Alzheimer's disease. *Science*. 2006;314(5800):777-781.
13. Roses AD, Saunders AM. Perspective on a pathogenesis and treatment of Alzheimer's disease. *Alzheimers Dement*. 2006;2(2):59-70.
14. Zinfandel Pharmaceuticals Inc. Opportunity for the Prevention of Alzheimer's Web site. http://www.opalstudy.org. Accessed February 11, 2010.

Announcement

Calendar of Events: A New Web Feature. On the new Calendar of Events site, available at http://pubs.ama-assn.org/cgi/calendarcontent and linked off the home page of the *Archives of Neurology*, individuals can now submit meetings to be listed. Just go to http://pubs.ama-assn.org/cgi/cal-submit/ (also linked off the Calendar of Events home page). The meetings are reviewed internally for suitability prior to posting. This feature also includes a search function that allows searching by journal as well as by date and/or location. Meetings that have already taken place are removed automatically.

Downloaded from www.archneurol.com at STANFORD Univ Med Center, on May 15, 2010
©2010 American Medical Association. All rights reserved.