Alzheimer's & Dementia 12 (2016) 687-694

Perspective

Understanding the genetics of *APOE* and *TOMM40* and role of mitochondrial structure and function in clinical pharmacology of Alzheimer's disease

Allen Roses[a,b,*], Scott Sundseth[a,b], Ann Saunders[a,b], William Gottschalk[a,b], Dan Burns[a,b], Michael Lutz[a,b]

[a]*Department of Neurology, Duke University School of Medicine, Durham, NC, USA*
[b]*Semillon Pharmaceuticals, Inc., Chapel Hill, NC, USA*

Abstract The methodology of Genome-Wide Association Screening (GWAS) has been applied for more than a decade. Translation to clinical utility has been limited, especially in Alzheimer's Disease (AD). It has become standard practice in the analyses of more than two dozen AD GWAS studies to exclude the apolipoprotein E (*APOE*) region because of its extraordinary statistical support, unique thus far in complex human diseases. New genes associated with AD are proposed frequently based on SNPs associated with odds ratio (OR) < 1.2. Most of these SNPs are not located within the associated gene exons or introns but are located variable distances away. Often pathologic hypotheses for these genes are presented, with little or no experimental support. By eliminating the analyses of the *APOE-TOMM40* linkage disequilibrium region, the relationship and data of several genes that are co-located in that LD region have been largely ignored. Early negative interpretations limited the interest of understanding the genetic data derived from GWAS, particularly regarding the *TOMM40* gene. This commentary describes the history and problem(s) in interpretation of the genetic interrogation of the "*APOE*" region and provides insight into a metabolic mitochondrial basis for the etiology of AD using both *APOE* and *TOMM40* genetics.

© 2016 The Authors. Published by Elsevier Inc. on behalf of the Alzheimer's Association. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Genetics; *APOE*; *TOMM40*; Mitochondria; Alzheimer's disease short structural variations [SSVs]; Sequencing; Phylogenetic mapping

In a seminar for medical students on the philosophy of science regarding research methodology, the question was raised regarding "how to keep-up with massive amount of literature." The answer was that, even for an investigator working in a highly specialized field, it is extremely difficult to be current. One approach to this dilemma is the strategy for reading research articles suggested by the late G. Milton Shy, Professor and Chairman of Neurology, University of Pennsylvania School of Medicine which was to "forget the abstract, which is basically the authors' interpretations but rather focus on the data and the published interpretation. Ask the question of whether the methods and analyses support the conclusions."

Even with modern "precision" genetic data, reader interpretations may miss the relevant facts. Specifically, as an illustrative example, this perspective will explain how interpretations concerning the association of apolipoprotein E (*APOE*) ε4 with late-onset Alzheimer's Disease in 1993 have overlooked such facts in more than two dozen subsequent Genome-Wide Association Studies (GWAS)—without recognizing that the particular SNP associations could not discriminate *APOE* alleles. The two *APOE* allele-defining SNPs (for *APOE* ε2, *APOE* ε3, or *APOE* ε4) were not included on the commercial platforms used

*Corresponding author. Tel.: +1 919-660-8065.
E-mail address: allen.roses@duke.edu

http://dx.doi.org/10.1016/j.jalz.2016.03.015
1552-5260/ © 2016 The Authors. Published by Elsevier Inc. on behalf of the Alzheimer's Association. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

before 2012 (Fig. 1) [1,2]. Most of the AD GWAS literature pre-dates that 2012 and exclude further analyses of the "*APOE*" region, whereas a few newer studies now use the Illumina OmniQuad Array (Fig. 1).

The precise nature of this "interpretation problem" might be better understood in the context of the rapidly developing GWAS field between 1997 and 2012. In a 1997 Science Commentary entitled "Snipping Away at Genome Patenting", Dr. David Cox (a distinguished and popular geneticist) was quoted as saying that SNP patents could become "a nightmare" for companies and basic researchers alike, and that Francis Collins, then head of the National Genome Research Institute (NHGRI), and others would like to "head off" [3]. In 1998, the formation of The SNP Consortium provided an opportunity for industry to work with the NIH to insure that SNPs would be in the public domain to promote widespread adoption in research [4]. The SNP Consortium was formed and supported by 10 pharmaceutical companies as a partnership to publish and distribute enough SNP data into the public domain to diminish the likelihood of proprietary SNP panels [4]. GlaxoWellcome (GW) Genetics had organized the Consortium and, as SNPs were identified, preceded to test whether enabling GWAS with new high volume/low cost commercially available genotyping platforms would be feasible.

The association of *APOE* ε4 with so-called late-onset sporadic Alzheimer's disease had been previously discovered in 1992 and published in 1993, a decade before any commercial GWAS platforms became available [1,2]. In 1997, the GW Genetics Group created a relatively dense map (at that time) of newly identified human SNPs from a 2-megabase region on either side of the *APOE* gene locus. This region was used to model a GWAS and test whether a GWAS approach, using the association of *APOE* and AD, would be easily recognized. This proof of concept was based on a simple question: could future GWAS screening identify the location of the *APOE* locus associated with age of onset distributions of AD if there were no prior knowledge [5–7].

The *APOE* locus was confirmed—including SNPs from *APOE* and SNPs within *PEREC1* (an expressed gene later discovered to be *TOMM40*) and *APOC1*. Specific allele-determining *APOE* SNPs as well as SNPs from the other recognized genes and noncoding SNPs in the linkage disequilibrium (LD) region provided very strong statistical evidence of association. Eventually, the Human Genome Project clearly defined this region as having strong LD.

GW had also collected well-phenotyped clinical cohorts with banked deoxyribonucleic acid (DNA) for 17 different diseases. After a merger created GlaxoSmithKline [GSK]

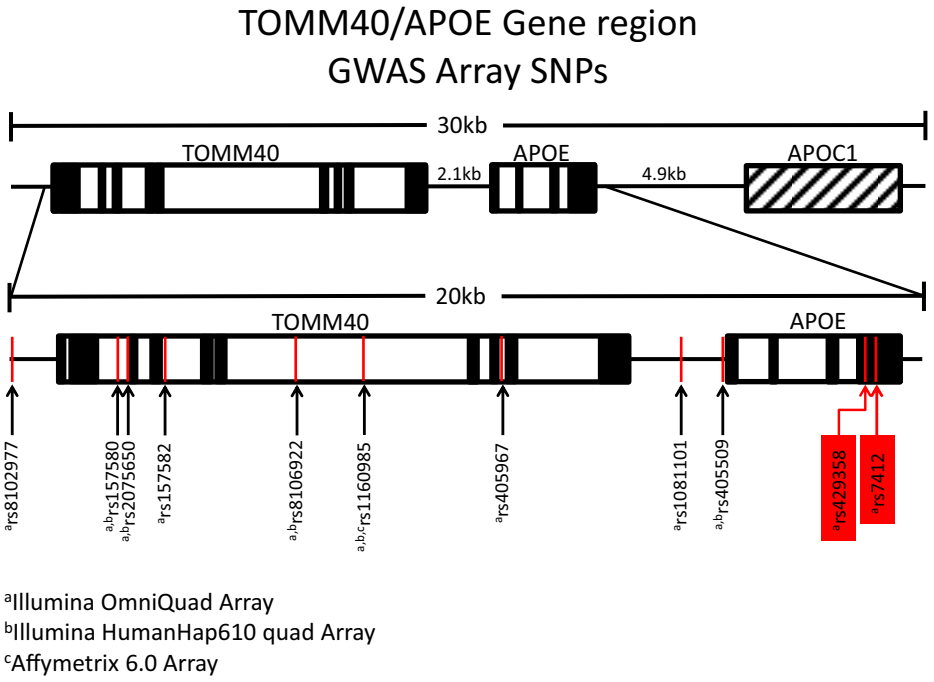

Fig. 1. The region, order, and distance between the *TOMM40*, *APOE* and *APOC1* genes are illustrated on the upper 30kp map. The lower map enhances the 20 KB region containing both the TOMM40 (12.4kb) and APOE (3.6kb) genes shown in a 20kb region below (black=exons; white=introns). The arrows below indicate the position of SNPs from three SNP arrays. SNPs labeled "a" represent the newest array [Illumina OmniQuad] coming into limited use post-2012. The SNPs labeled "b" are those on the commonly used pre-2012 GWAS studies in AD. The SNP labeled "c" is the only representative in this LD region [Affymetric 6.0] and was occasionally used pre-2012 in AD GWAS studies. The SNPs marked in the red boxes are the APOE allelic determining SNPs, and are present only on the Illumina OmniQuad arrays. Prior to 2012 no AD GWAS reported using the Illumina OmniQuad arrays containing these two SNPs directly. The Illumina HumanHap610.quad array that was widely used for GWAS studies contained SNPs located within the introns of TOMM40 (rs157580, rs2075650, rs8106922, rs1160895) or the intragenic region (rs405509). All Affymetrix GWAS arrays contained only the single rs1160985 SNP located in intron 5 of TOMM40.

in 2000, newly available commercial SNP platforms were purchased from several technology companies to perform parallel beta pilot testing. SNP genotyping platforms were directly compared using the same DNA samples to determine functional characteristics of each platform.

By 2006, GSK had completed GWAS studies from seven different disease cohorts, using several different platforms. The results clearly demonstrated that SNPs from immediate neighbors within the "*APOE*" locus had a unique association with AD. A similarly robust association was not observed at any measured locus in the other six diseases. It was clear that this strategy would not be generalized. These negative GWAS studies were not increased in size to detect small signals and were concluded in 2007. Two of the commercial platforms (without the two *APOE* allele-determining SNPs) became widely used as the winning platforms. GWAS became the hot area of research throughout medicine fueled by commercial SNP panels. The results from the "*APOE*" region were (and still are) unique. Instead of simply adding more patients in more diseases, at an extraordinary cost, GSK genetics initiated a post-GWAS strategy using phylogenetic mapping to dissect the "*APOE*" linkage disequilibrium (LD) region.

GWAS has provided massive information to the public databases: yet, it should be considered whether the data are representative of a gene or actually an LD region. Identification of a "gene" located near a particular SNP on the platform gets reported frequently from GWAS from many diseases, including associated press releases. Historically, microsatellite markers for genetic mapping had been anchored to genes. However, with the success of the Human Genome Project, genetic markers can be accurately mapped to primary DNA. Currently, GWAS is transitioning to next-generation sequencing (NGS), but it has unfortunately become second nature to refer to specific SNPs with low odds ratio association data as "disease genes," even if the SNP is at some distance from the nominated gene. As will be discussed below, even common complete NGS technologies in 2016 do not accurately measure the length of short tandem repeat (STR) sequences, which may indeed be more important to define common complex diseases from more than a million variable structural variations in the human genome. Recently, there has been some progress in measuring repeat length sequences accurately, but expectations have been high for this technology for the past several years. One of the experimental data sets being investigated is the *TOMM40′*523 locus. When Dr. David Botstein joined the Board of Pac-Bio in 2012, he stated in the press release [8]

> "We are moving toward an exciting era where the cost of producing sequencing information is going down, and now the emergence of the high-quality, multi-kilobase sequence will be a double game changer. Even with today's advanced sequencers, a lot of good biology is being left on the table because we cannot fully assemble genomes for the many organisms for which there is no reference, or because we are missing difficult to sequence regions of critical variation. Pacific Biosciences' long-read technology offers the field something that currently does not otherwise exist—the ability to create high quality de novo assemblies, improve re-assemblies for complex organisms, and understand DNA modifications within sequences."

1. *TOMM40* and AD

In 2005, GSK was unable to fully sequence the "*APOE*" LD region for phylogenetic mapping purposes. A variable length of region in intron 6 of the *TOMM40* gene could neither be sequenced in two GSK sequencing laboratories nor in 5 of 6 commercial sequencing companies. However, one out-sourced laboratory accomplished the sequencing successfully (Polymorphic Technologies, Inc.; Alameda, CA). Using their specialized methods (now validated), all the SNPs and multiple repeat sequences in the LD region were sequenced. Sanger sequencing protocols were validated to map the DNA strands and to begin defining variable nucleotide length repeats on the same chromosomal strand (*cis*). GSK used phylogenetic mapping to place evolutionary variants within the *TOMM40-APOE-APOC1* LD region in well-defined clades. Fig. 2A depicts the phylogenetic map, as published, for an Arizona Alzheimer's Disease Research Center (ADRC) AD cohort [9]. Similar lengths of polyT repeats at the *TOMM40′*523 locus were characteristic of several major phylogenetic clades in another independent Caucasian cohort (Canadian). The clusters of polyT length polymorphisms were subsequently associated with the age of onset distributions, using a prospectively ascertained clinical AD cohort followed at the Bryan ADRC. Fig. 2B illustrates the phylogenetic clade intersection which separates the ancestral Long *TOMM40′*523 alleles from the separate VL and S clades [9–11].

The frequency of major clusters of polyT length in major clades differs in multiple ethnic groups [10,11]. The accuracy of the polyT lengths using validated methods (Polymorphic DNA Inc. for research use, and Quest Diagnostics in more than 22,000 clinical trial tests under Food and Drug Administration standards) allowed the observation in African-Americans of new cis-haplotypes that had not been observed in multiple Caucasian populations studied at more than 50 centers in the United States, United Kingdom, Germany, Switzerland, Australia, and Russia (Siberia) participating in the TOMMORROW clinical trial testing an age-dependent risk algorithm using *TOMM40′*523, *APOE*, and age of onset [12,13].

Within Caucasian cohorts from around the world, the pattern of phylogenetic maps was shown to be highly reproducible [9,13]. More than 28 different polyT lengths were mapped to several large evolutionary clades. The large clades were grouped into three polyT length clusters which were labeled short (S = <20 T's), long (L = 20–30), and very long (VL = 30) polyT

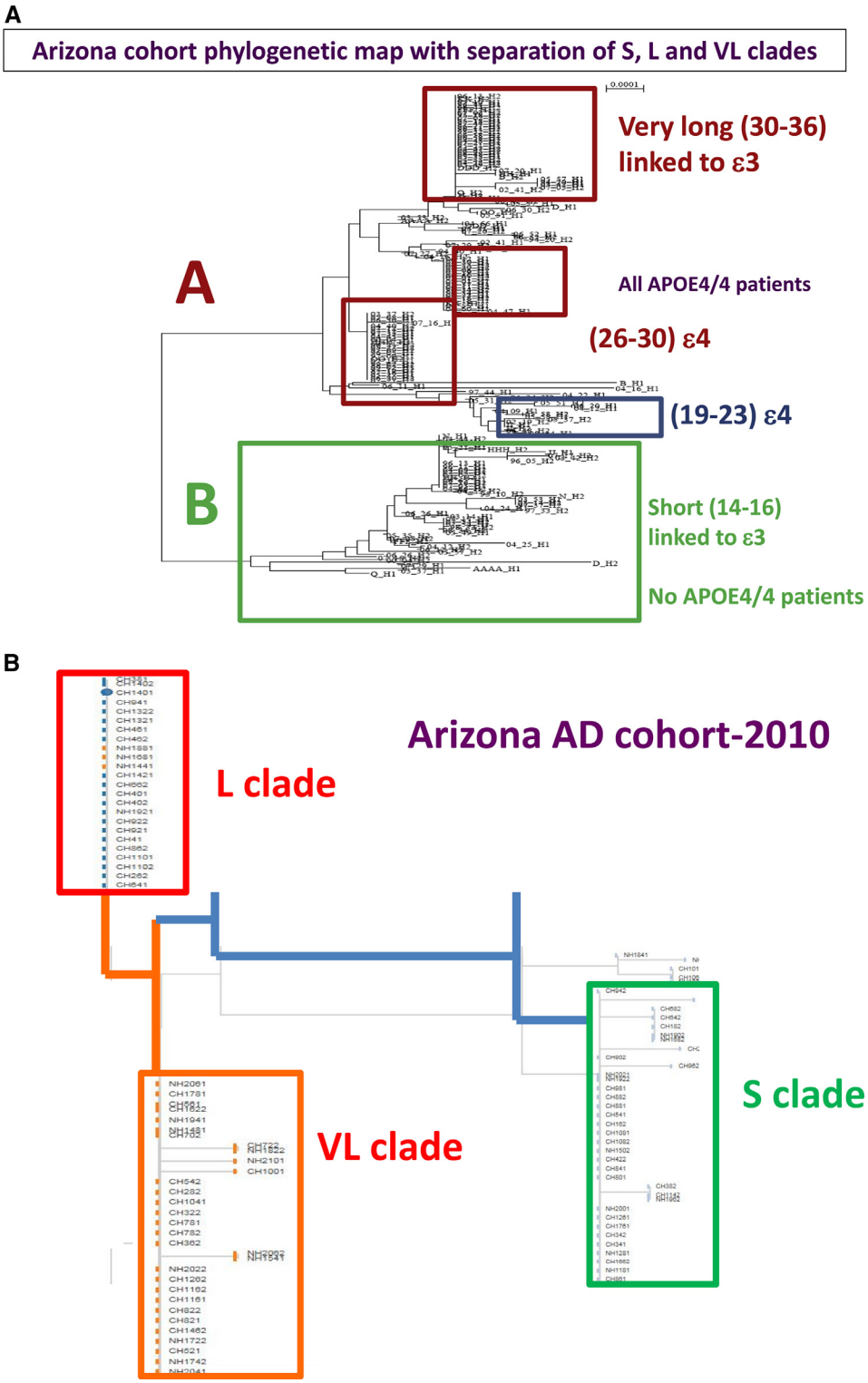

Fig. 2. (A) *TOMM40'*523 poly-T length polymorphisms evolved independently over time and on different genetic backgrounds at a single highly variable locus. Each major clad is characterized by new mutations based on the repeat sequence common in that clade [9]. (B) Another view of the Arizona cohort mapped on new recent TreeLink Software. [22] The use of TreeLink Software allows the major clades containing VL and S "clades" to be visualized as separately branching from the main evolutionary L clade at two distinct times. Horizontal distances are directly from the software and are commensurate with evolutionary time scale. This illustrates that the major clades differ in the evolutionary timing of formation and contains subsequent chromosomal SNPs [sometimes referred to as "mutations"] according to the S, L and VL polyT lengths.

lengths. Studies of several Caucasian cohorts demonstrated that: S lengths are in phase *cis* to *APOE* ε3 and *APOE* ε2 alleles; the L alleles segregate *cis* to *APOE* ε4 chromosomes. Nearby SNPs that are also mapped within each of these large clades also may reflect association to age of onset distributions. The *TOMM40′*523 polyT lengths connected to *APOE* alleles could now allow *TOMM40′*523-*APOE* cis-haplotypes to define age of onset distributions for >97% of Caucasians, connecting the clinical phenotype of AD to *APOE* genotypes, as was reported in 1993 [2,12]. The major distinction of this work is not simply the association of the region with AD but the informative estimates of risk for >97% of individuals. It is important to note that there are three additional distributions of *APOE* ε3/3 carriers (S-S, S-VL, and VL-VL) that are now providing differential age of onset information using *TOMM40′*523 that were not available using *APOE* ε4 alone (4/4, 3/4, and 2/4). Both *APOE* ε2/2 (1%–2%) and *APOE* ε2/4 (1%–2%) are uncommon genotypes in Caucasian AD patients, so there are not yet enough age of onset data points to provide a distribution in clinical cohorts, but such data will be available and shared publically when the TOMMORROW clinical delay of mild cognitive impairment onset trial is completed [12]. The *TOMM40′*523/*APOE* cis-haplotypes are even more informative, including data from several ongoing studies of other ethnicities, including African-Americans, that recognize cis-haplotypes rarely encountered in Caucasians [10–12,15].

2. Precision versus accuracy

Between 1993 and 1995, most AD geneticists were skeptical of the role of *APOE* ε4 in sporadic AD [14]. However, as *APOE* isotyping (using gel electrophoreses to distinguish the apoE proteins) was available within several hospitals (primarily European), the *APOE* ε4 association with AD was confirmed rapidly [15]. The association has held up remarkably as a confirmed risk factor over the years [15–18]. However, it was recommended in 1997 by an NIH Conference Panel that *APOE* ε4 should not be used as a diagnostic in clinical practice [19]. Because Duke University owned the issued patent on the association of *APOE* ε4 and AD, Duke made the decision to exclude the two SNPs that determine the *APOE* alleles from the developing GWAS platforms so that the two SNPs would not be used inappropriately as a clinical genetic test for a disease with inadequate therapy but would be available for use in research and as a clinical pharmacogenetic test [20]. This did discourage the use *APOE* genotyping as a commercial medical diagnostic product. However, when the early SNP platforms were being beta-tested, and until 2012, commercial GWAS platforms used for AD studies contained neither coding SNPs nor any SNPs from introns of *APOE* (Fig. 1). The use of other SNPs in the so-called "*APOE*" LD region allowed precise location to the LD area but were not accurate for expanding to *APOE* genetics [20]. Four of the 5 SNPs on the most used platforms before 2012 used noncoding SNPs outside the APOE locus to impute the LD region and could not make an accurate call of specific *APOE* alleles, as measured by the two SNPs that define the gene variance.

The first wave of GWAS studies in AD, before 2012, represents other alleles that may be linked to APOE but are not direct assessments of *APOE* ε4 alleles [20,21]. The "*APOE* region" is the common slang for relating association data that actually represent the LD region and not *APOE* alleles for any individual. When genetic rigor is applied, the data represent *TOMM40* SNPs more accurately than *APOE*. Statements made in 2012 about separating an "*APOE*" effect from a *TOMM40* effect are not possible, because neither the *TOMM40′*523 alleles or the *APOE* alleles were specifically studied [22]. The GWAS data are fine for association within the LD region but do not test or recognize specific cis-haplotypes available from phylogenetic maps. A clinically important difference is that *TOMM40′*523 genotyping is more informative in >97% of the Caucasian population, then only 29% of *APOE* ε4 carriers and, for the first time, *APOE* ε3/3 patients can be divided with two specific *APOE* ε3-*TOMM40′*523 cis-haplotypes into three groups (S/S, S/VL, and VL/VL) with distinct age of onset curves [9,12].

This does not mean that apoE isoforms are neither involved in the pathogenesis of AD nor does it mean that *TOMM40* is the gene. It does suggest relevance of two adjacent genes to the rate of clinical expression of AD. The evidence for a known functional interactions within human metabolism is emerging, beginning with the differential effects of *APOE* isoforms on mitochondrial dynamics, differential effects on mitochondrial metabolism mitochondria, and decreased mitochondrial dynamic functions [23]. Using later SNP platforms with *APOE* allele-specific SNPs is still less robust than phased phylogenetic mapping. The genetic association of *TOMM40′*523/*APOE* cis-haplotypes is far more accurate to translate to individuals in the population than measuring a few SNPs from *TOMM40* and calling it "*APOE*" data. For the first time, three different ages of onset distribution can now be attributed to the largest group of AD patients, *APOE* ε3/3 (60% of Caucasians), by delineating three *TOMM40′*523 sub-classes (S-S, S-VL, and VL-VL), raising the informative data to >97% of individuals, compared to 29% with APOE4 alone [12].

*TOMM40′*523 association with AD age of onset has received the same skeptical welcome that had been the case for *APOE* in the early 1990s [3,16]. In one early "Online First" Editorial published in August 2012 it was stated: "Not only did genetic variants within *APOE* show association with AD but single-nucleotide polymorphisms (SNPs) around the locus also presented strong associations with the disease" [22]. The operative word is "also," because the two dozen GWAS studies that had been reported by that time did not measure the two SNPs that define *APOE* alleles. The authors also stated that

"it is very difficult to attribute a genetic and APOE-independent role of TOMM40." This is especially true when *APOE* alleles and *TOMM40′*523 are not specifically measured. Because the *TOMM40′*523 region can now be phased on the same DNA strand, the cis-haplotypes data can be far more accurate and informative than genotypes or *SNP* haplotypes. Even newer SNP platforms that contain the two specific *APOE* SNP alleles cannot provide phased chromosomal relationships.

AD GWAS articles had discussed only the "*APOE*" region, ignoring the *TOMM40′*523 data although four of the five SNPs studied were located within TOMM40. Clearly, much of the field did not realize what they were actually measuring. To make life easier in finding new genes, most GWAS studies excluded the entire "*APOE*" LD region in their analyses. [21] However, the AD GWAS published by our group at GSK discussed the possibility of another gene, later defined as *TOMM40* but present in our 1998 article as PEREC-1, before it was known that this was the locus for the translocase of the external mitochondrial membrane channel [5,20].

It is inaccurate to attribute all late-onset AD genetics to *APOE* (or *TOMM40*). However, it must be realized that it is also inaccurate to depend solely on small statistical effects to nominate a gene. GWAS uses a common correction for a million SNP tests to generate statistical associations. A designated gene that is near to a specific SNP or several SNPs may be precise localization, perhaps even within LD. Accuracy and precision are not one in the same. Although the precise molecular pathogenesis affecting mitochondria leading the increased risk of AD remains undetermined, the "precision" of localization to the "*APOE* LD region" has diverted consideration of fine mapping of the actual variants within the LD region over the past 5 years. *TOMM40′*523 is a structural variation locus that is coded at a precise genetic locus with more than two dozen length variations. There is sufficient variation to be associated with individual age of onset distributions—not just association to large groups of patients labeled as AD without taking into account the ∼10% error in AD diagnosis in living patients. The important point is that *APOE* became "the" referenced gene rather than simply an important gene in a broader view of the LD field. This has diverted consideration of the LD genetic findings and hampered efforts to develop new therapeutic options to ease the burden of AD. Being precise in locating the "field," but inaccurate and off-target is of little benefit.

3. Considerations and speculations

As a broader view important to GWAS interpretations in other diseases, imputation of a nearby SNP from GWAS and assigning a "gene name" to it needs to be carefully noted [21,24–26]. Naming such genes misses the point that other SNPs and structural variants, like the polyT variants of *TOMM40′*523, can provide a clearer view of the nearby field and other more precise data can be observed within the field [9,12,26–28].

Perhaps, the most important observations with respect to *TOMM40* and the association of age of onset of AD is that there are variable length polyT repeats at a single locus within intron 6 of *TOMM40* which carry the critical phenotype of age of onset. The distribution of this single nucleotide repeat is directly related to the AD phenotype. The pathophysiology field for AD is broader than simply amyloid or tau deposition. *APOE* ε4 is commonly imputed as the important gene and is certainly clearly involved as a significant biological contributor. Similarly, the *TOMM40′*523 variants may be involved with alternative splicing or mRNA expression [27,30].

The future may have new preventative and treatment drugs with different efficacy pharmacogenetic profiles, dependent on the cis-haplotypes of *TOMM40′*523 and *APOE* [12]. *TOMM40* is the main channel through which proteins and peptides must traverse to support the dynamic functions of mitochondria, identifying pathogenic factors than may result through changes in mRNA expression and/or alternative splicing may be critical in the future [27–30]. We know that many aggregating proteins bind and traverse through Tom40 protein channel and aggregate within the channel. Besides apoE4 and apoE3, there are more than 30 proteins associated pathologically with brain plaques in AD. APP and amyloid may exert catalyzing effects on the rate of disease expression based on heterogeneous interactions with the mitochondrial channel through which it is meant to travel [30]. Studying the amyloid protein as a component of the pathogenesis is critical, but we need to know the right drug target for drug discovery [31]. None of the previously mentioned GWAS studies ever demonstrated statistical association attributed to the APP gene [31–34].

Tom40 dysfunction has been noted in Parkinson's disease and may also relate to Huntington's disease [33,34]. Unfortunately, there are many degenerative neurological diseases in which Tom40 remains unstudied. Mitochondrial dynamics are affected by apoE4 differently than by apoE3 [35–40]. The evidence for mitochondrial effects in the brain, particularly in glucose and oxygen utilization, is quite clear and evident from many positron imaging experiments [41]. Oxygen and glucose are primarily processed by mitochondria, especially in the brain. ApoE isoforms and Tom40 variants may affect that metabolism probably from birth. The course of AD may be accelerated by the inheritance of *APOE* ε4, with plaques and tangles representing visible damage. Finally, *TOMM40′*523 is related to AD pathogenesis based on the >97% informative age of onset distributions. ApoE, amyloid, and synuclein proteins are interactive with Tom40 [33,42]. The Tom40 protein forms the channel by specific protein interactions involving several intrinsic proteins and multiple signaling proteins [43,44]. Basic science studies focusing on the structure and function of *TOMM40* and its associated interactions

are now entering the scientific literature [45–48] and will accelerate drug discovery and development from determining the role of mitochondria and relevant mechanisms of pathogenesis.

RESEARCH IN CONTEXT

1. Systematic review: The authors reviewed the Alzheimer's Disease GWAS literature using traditional (e.g., PubMed) sources, as well as press releases with respect to the interpretation of limited genetic data. The content of commercial GWAS single-nucleotide polymorphism (SNP) platforms and the designation of small odds ratios as evidence for declaring a new "AD gene" of small effect was the primary focus.
2. Interpretation: Our findings highlighted the limitations of the GWAS platforms compared to phylogenetic mapping of sequenced data. The data utilized for genetic identity were contrasted to phylogenetic mapping of the *APOE* and *TOMM40* linkage disequilibrium region. This LD region has been acknowledged as robust, but is usually formally excluded from additional genetic analyses in order to focus on new "genes."
3. Future directions: This Perspective proposes a genetic framework for the confirmation of "new gene" hypotheses using sequencing of proposed LD regions and highlighting the role of highly polymorphic short tandem repeat sequences related to structural variants located throughout the genome.

References

[1] Saunders AM, Roses AD. Apolipoprotein e4 allele frequency, ischemic cerebrovascular disease, and Alzheimer's disease. Stroke 1993;24:1416–7.

[2] Corder E, Saunders A, Strittmatter W, Schmechel D, Gaskell P, Small G, et al. Gene dose of apolipoprotein e type 4 allele and the risk of alzheimer's disease in late onset families. Science 1993;261:921–3.

[3] Marshall E. Snipping away at genome patenting. Science 1997; 277:1752–3.

[4] Holden AL. The snp consortium: Summary of a private consortium effort to develop an applied map of the human genome. Biotechniques 2002;Suppl. 22–24, 26.

[5] Lai E, Riley J, Purvis I, Roses A. A 4-mb high-density single nucleotide polymorphism-based map around human apoe. Genomics 1998; 54:31–8.

[6] Martin ER, Lai EH, Gilbert JR, Rogala AR, Afshari AJ, Riley J, et al. SNPing away at complex diseases: Analysis of single-nucleotide polymorphisms around apoe in Alzheimer disease. Am J Hum Genet 2000;67:383–94.

[7] Martin ER, Gilbert JR, Lai EH, Riley J, Rogala AR, Slotterbeck BD, et al. Analysis of association at single nucleotide polymorphisms in the apoe region. Genomics 2000;63:7–12.

[8] Pacific biosciences names david botstein, ph.D. To its board of directors Menlo Park, CA, 2012. Available at: http://investor.pacificbiosciences.com/releasedetail.cfm?releaseid=694085. Accessed May 7, 2016.

[9] Roses AD, Lutz MW, Amrine-Madsen H, Saunders AM, Crenshaw DG, Sundseth SS, et al. A Tomm40 variable-length polymorphism predicts the age of late-onset Alzheimer's disease. Pharmacogenomics J 2010;10:375–84.

[10] Roses AD, Lutz MW, Saunders AM, Goldgaber D, Saul R, Sundseth SS, et al. African-american tomm40'523-apoe haplotypes are admixture of west african and caucasian alleles. Alzheimers Dement 2014;10:592–601.

[11] Linnertz C, Saunders AM, Lutz MW, Crenshaw DM, Grossman I, Burns DK, et al. Characterization of the poly-t variant in the tomm40 gene in diverse populations. PLoS One 2012;7:e30994.

[12] Crenshaw DG, Gottschalk WK, Lutz MW, Grossman I, Saunders AM, Burke JR, et al. Using genetics to enable studies on the prevention of alzheimer's disease. Clin Pharmacol Ther 2013;93:177–85.

[13] Lutz MW, Sundseth SS, Burns DK, Saunders AM, Hayden KM, Burke JR, et al. A genetics-based biomarker risk algorithm for predicting risk of alzheimer's disease. Alzheimers Dement (N Y) 2016;2:30–44.

[14] Marx J. New alzheimer's theory stirs controversy. Science 1993;262: 1210–1211.5.

[15] Cummings JL, Vinters HV, Cole GM, Khachaturian ZS. Alzheimer's disease: Etiologies, pathophysiology, cognitive reserve, and treatment opportunities. Neurology 1998;51(1 Suppl 1):S2–17. discussion S65–17.

[16] Schellenberg GD, Montine TJ. The genetics and neuropathology of Alzheimer's disease. Acta Neuropathol 2012;124:305–23.

[17] Schellenberg GD, D'Souza I, Poorkaj P. The genetics of Alzheimer's disease. Curr Psychiatry Rep 2000;2:158–64.

[18] Guerreiro RJ, Gustafson DR, Hardy J. The genetic architecture of Alzheimer's disease: Beyond app, psens and apoe. Neurobiol Aging 2012; 33:437–56.

[19] Post SG, Whitehouse PJ, Binstock RH, Bird TD, Eckert SK, Farrer LA, et al. The clinical introduction of genetic testing for Alzheimer disease. An ethical perspective. JAMA 1997;277:832–6.

[20] Mahley RW, Huang Y. Apolipoprotein e sets the stage: Response to injury triggers neuropathology. Neuron 2012;76:871–85.

[21] Li H, Wetten S, Li L, St Jean PL, Upmanyu R, Surh L, et al. Candidate single-nucleotide polymorphisms from a genome-wide association study of Alzheimer disease. Arch Neurol 2008;65:45–53.

[22] Guerreiro RJ, Hardy J. Tomm40 association with alzheimer disease: Tales of apoe and linkage disequilibrium. Arch Neurol 2012; 69:1243–4.

[23] Roses AD. APOE and TOMM40 in Alzheimer's Disease: A case of mistaken identity, using precision tools but with loss of accuracy. J Pharmacogenomics Pharmacoproteomics 2015;6:3.

[24] Jun G, Ibrahim-Verbaas CA, Vronskaya M, Lambert JC, Chung J, Naj AC, et al. A novel Alzheimer disease locus located near the gene encoding tau protein. Mol Psychiatry 2016;21:108–17.

[25] Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet 2013; 45:1452–8.

[26] Roses AD, Lutz MW, Crenshaw DG, Grossman I, Saunders AM, Gottschalk WK. Tomm40 and apoe: Requirements for replication studies of association with age of disease onset and enrichment of a clinical trial. Alzheimers Dement 2013;9:132–6.

[27] Roses AD, Akkari PA, Chiba-Falek O, Lutz MW, Gottschalk WK, Saunders AM, et al. Structural variants can be more informative for disease diagnostics, prognostics and translation than current SNP

mapping and exon sequencing. Expert Opin Drug Metab Toxicol 2016; 12:135–47.
[28] Roses AD. Polyallelic structural variants can provide accurate, highly informative genetic markers focused on diagnosis and therapeutic targets: Accuracy vs. precision. Clin Pharmacol Ther 2016;99:169–71.
[29] Mani J, Meisinger C, Schneider A. Peeping at toms-diverse entry gates to mitochondria provide insights into the evolution of eukaryotes. Mol Biol Evol 2016;33:337–51.
[30] Wang W, Karamanlidis G, Tian R. Novel targets for mitochondrial medicine. Sci Transl Med 2016;8:326rv3.
[31] Roses AD, St Jean PL, Ehm MG. Use of whole-genome association scans in disease gene identification, drug discovery and development. IDrugs 2007;10:797–804.
[32] Hollingworth P, Harold D, Sims R, Gerrish A, Lambert JC, Carrasquillo MM, et al. Common variants at abca7, ms4a6a/ms4a4e, epha1, cd33 and cd2ap are associated with alzheimer's disease. Nat Genet 2011;43:429–35.
[33] Bender A, Desplats P, Spencer B, Rockenstein E, Adame A, Elstner M, et al. Tom40 mediates mitochondrial dysfunction induced by alpha-synuclein accumulation in parkinson's disease. PLoS One 2013; 8:e62277.
[34] Ribeiro M, Rosenstock TR, Oliveira AM, Oliveira CR, Rego AC. Insulin and igf-1 improve mitochondrial function in a pi-3k/akt-dependent manner and reduce mitochondrial generation of reactive oxygen species in Huntington's disease knock-in striatal cells. Free Radic Biol Med 2014;74:129–44.
[35] 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;102:18694–9.
[36] Roses AD, Saunders AM, Lutz MW, Zhang N, Hariri AR, Asin KE, et al. New applications of disease genetics and pharmacogenetics to drug development. Curr Opin Pharmacol 2014;14:81–9.
[37] Huang Y, Mahley RW. Apolipoprotein e: Structure and function in lipid metabolism, neurobiology, and alzheimer's diseases. Neurobiol Dis 2014;72 Pt A:3–12.
[38] Roses AD, Saunders AM, Huang Y, Strum J, Weisgraber KH, Mahley RW. Complex disease-associated pharmacogenetics: Drug efficacy, drug safety, and confirmation of a pathogenetic hypothesis (Alzheimer's disease). Pharmacogenomics J 2007; 7:10–28.
[39] Chen HK, Ji ZS, Dodson SE, Miranda RD, Rosenblum CI, Reynolds IJ, et al. Apolipoprotein e4 domain interaction mediates detrimental effects on mitochondria and is a potential therapeutic target for Alzheimer disease. J Biol Chem 2011;286:5215–21.
[40] 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;105:1343–6.
[41] Reiman EM, Caselli RJ, Yun LS, Chen K, Bandy D, Minoshima S, et al. Preclinical evidence of alzheimer's disease in persons homozygous for the epsilon 4 allele for apolipoprotein e. N Engl J Med 1996;334:752–8.
[42] Anandatheerthavarada HK, Biswas G, Robin MA, 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;161:41–54.
[43] Hohr AI, Straub SP, Warscheid B, Becker T, Wiedemann N. Assembly of beta-barrel proteins in the mitochondrial outer membrane. Biochim Biophys Acta 2015;1853:74–88.
[44] Gornicka A, Bragoszewski P, Chroscicki P, Wenz LS, Schulz C, Rehling P, et al. A discrete pathway for the transfer of intermembrane space proteins across the outer membrane of mitochondria. Mol Biol Cell 2014;25:3999–4009.
[45] Melin J, Schulz C, Wrobel L, Bernhard O, Chacinska A, Jahn O, et al. Presequence recognition by the tom40 channel contributes to precursor translocation into the mitochondrial matrix. Mol Cell Biol 2014; 34:3473–85.
[46] Shiota T, Imai K, Qiu J, Hewitt VL, Tan K, Shen H-H, et al. Molecular architecture of the active mitochondrial protein gate. Science 2015; 349:1544.
[47] Kuszak AJ, Jacobs D, Gurnev PA, Shiota T, Louis JM, Lithgow T, et al. Evidence of Distinct Channel Conformations and Substrate Binding Affinities for the Mitochondrial Outer Membrane Protein Translocase Pore Tom40. J Biol Chem 2015;290:26204–17.
[48] Allende C, Sohn E, Little C. Treelink: Data integration, clustering and visualization of phylogenetic trees. BMC Bioinformatics 2015; 16:414.