

Published in final edited form as:

*Alzheimers Dement.* 2014 October ; 10(0): S277–S283.e10. doi:10.1016/j.jalz.2013.09.005.

## Origin of the PSEN1 E280A mutation causing early-onset Alzheimer's disease

Matthew A. Lalli<sup>a</sup>, Hannah C. Cox<sup>b</sup>, Mary L. Arcila<sup>c</sup>, Liliana Cadavid<sup>c</sup>, Sonia Moreno<sup>c</sup>, Gloria Garcia<sup>c</sup>, Lucia Madrigal<sup>c</sup>, Eric M. Reiman<sup>d</sup>, Mauricio Arcos–Burgos<sup>c,e</sup>, Gabriel Bedoya<sup>c</sup>, Mary E. Brunkow<sup>b</sup>, Gustavo Glusman<sup>b</sup>, Jared C. Roach<sup>b</sup>, Leroy Hood<sup>b</sup>, Kenneth S. Kosik<sup>a,\*</sup>, and Francisco Lopera<sup>c</sup>

<sup>a</sup>Neuroscience Research Institute, Department of Molecular Cellular Developmental Biology, University of California, Santa Barbara, CA, USA

<sup>b</sup>Institute for Systems Biology, Seattle, WA, USA

<sup>c</sup>Grupo de Neurociencias de Antioquia, Universidad de Antioquia, Medellín, Colombia

<sup>d</sup>Banner Alzheimer's Institute, Phoenix, AZ, USA

<sup>e</sup>Translational Genomics Group, Genome Biology Department, John Curtin School of Medical Research, ANU College of Medicine, Biology and Environment, The Australian National University, Canberra, ACT, Australia

### 1. Introduction

A glutamic acid to alanine mutation at codon 280 (E280A) in the presenilin-1 gene (PSEN1) causing early-onset familial Alzheimer's disease (EOFAD) at a mean age of 49 years [1] and displaying a fully penetrant autosomal dominant transmission was first discovered in a kindred in the state of Antioquia, Colombia [2,3]. Subsequently, a large extended kindred from the same region with over 5000 members was identified and described [4]. All affected individuals have the same mutation, present the same phenotype, come from the same geographical region and share common last names.

Recent whole-genome sequencing of six affected individuals from this kindred revealed an extensive stretch of identity-by-descent (IBD) containing the causal E280A mutation, suggesting the introduction of this allele into the population through a recent common founder event [5]. In this study, we investigate a large sample of subjects from this extended kindred through whole-genome sequencing. Our goals were (1) to determine the extent of

© 2013 Elsevier Inc. All rights reserved.

\*Corresponding author: Kenneth S. Kosik, kosik@lifesci.ucsb.edu, Phone: (805) 893-5222.

**Laboratory of Origin:** Kosik Lab, Neuroscience Research Institute, University of California at Santa Barbara. 552 University Rd. Bio II Building 571. Santa Barbara, CA, 93106, USA.

#### Potential conflicts of interest

The authors declare that they have no conflicts of interest.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

IBD surrounding the E280A, (2) to estimate the age of the E280A mutation, and (3) to determine on which genetic background the mutation arose.

## 2. Methods

### 2.1 Participants

Nearly a thousand individuals in the extended kindred have been identified and grouped into 24 sub-pedigrees (Figure 1A). Each pedigree was assigned an identification code consisting of the letter C (referring to Colombia) followed by a number indicating the order in which they were found. To assemble these 24 pedigrees and find the possible first founder of E280A we employed several strategies: interviews with elder healthy individuals of each affected family, interviews with genealogists and historians of the region, study of local historical and genealogical books, and examination of last wills and ecclesiastical records (baptismal, confirmations, marriage, death, census, etc) dating back as far as 1540. At the end of this research, historical records suggested that the affected population has a common ancestor from Spain [4], and we have assembled a large pedigree showing common ancestry for 13 families (Figure 1B).

Individuals in the extended kindred are enrolled in the Colombian Alzheimer's Prevention Initiative (API) registry. This registry comprises more than 2700 living members from the Antioquian early-onset familial Alzheimer's kindred. Around 30% of these members carry the PSEN1 E280A mutation. In total, this kindred is estimated to have approximately 5000 living relatives, including ~1500 mutation carriers.

Enrollment in this study was pursuant to approval by Institutional Review Board at the Universidad de Antioquia and the Western Institutional Review Board at the Institute for Systems Biology. Additional descriptions of the PSEN1 E280A Antioquian population, study participants, and relevant procedures have been previously and extensively provided [1].

### 2.2 Whole genome sequencing

Including six genomes from our previous study, DNA was collected for 102 samples in the Antioquian EOFAD cohort and sequencing was performed by Complete Genomics Incorporated (CGI, Mountain View, CA) using unchained combinatorial probe anchor (cPAL) chemistry [6]. Of the 102 samples, 74 individuals harbored the E280A mutation. Our study analyzes chromosome 14 ( $n = 204$ ) and includes 75 chromosomes bearing the E280A mutation (1 subject was homozygous for the mutation). Reads were aligned to the reference genome (NCBI Build 37, GRCh37) and variants were called by CGI.

### 2.3 Haplotype reconstruction

The presence of an extensive region of IBD around PSEN1 common in all carriers of E280A implies that the E280A allele resides on this haplotype. This allows us to reconstruct an extended haplotype using systematic long-range phasing (SLRP) [7] which is ideal for use in a recent founder population such as the Antioquian EOFAD cohort.

## 2.4 Estimating the age of E280A

After reconstructing haplotype phase using SLRP, we selected informative markers in a ~30 Mb region centered on E280A to estimate the age of E280A. We used two independent methods to determine the age of the E280A allele: a single-marker method [8,9], and the coalescent modeling of intra-allelic variability by DMLE+2.2 software (DMLE) [10]. The single marker method is based on the expected exponential decay of a given haplotype over generation time and estimates the time since the most recent common ancestor (TMRCA) for a given mutation. In contrast, DMLE estimates the actual age of the *de novo* mutation event, thus we expect an older estimated age from DMLE. The single marker method, unlike DMLE, fails to account for population growth rate, which leads to a further underestimation of the mutation age. Additionally, the super-exponential growth of the study population will lead to a further underestimation of the mutation age.

We estimated the mutation age in generations ( $g$ ) with a single-marker method using the following formula [8]:



Here,  $p_d$  is the frequency of a given marker on the haplotype bearing the E280A allele, and  $p_n$  is the overall frequency of a given marker in the sample population ( $n=204$  chromosomes).  $\theta$  is the recombination rate between a given marker and E280A. For this equation to work,  $p_d$  must be greater than  $p_n$ . Physical positions (bp) were converted to genetic distances (cM) using the genetic map obtained from the HapMap2 database lifted over to GRCh37 [11]. Physical positions absent from this map were interpolated.

We then used a second method, modeling the intra-allelic coalescent using DMLE [10], to estimate the age of the mutation. Haplotypes for disease and non-disease chromosomes were constructed with the markers used in the previous section.

Growth rate per generation was estimated by:



where,  $P_t$  is the current population size,  $P_0$  is the initial population size, and  $g$  is the number of generations between the present and the population size at the moment of the mutation origin (assuming 25 years/generation). As an admixture of Native American, European, and African populations, the Antioquian E280A kindred has an historical maximum  $g$  of ~20 since the advent of Europeans to the New World. We estimate an initial population between two–100 founders, and a present day size of around 5000. This yields a super-exponential generational population growth rate of ~0.2–0.4.

We include 75/1500 disease-bearing chromosomes or ~5% of all carriers. We ran DMLE +2.2 with 100,000 burn-in iterations followed by 100,000 simulations of the mutation age.

## 2.5 Local ancestry estimation

We constructed reference panels from HapMap [11], the 1000 Genomes Project [12], and the Human Genome Diversity Panel (HGDP) [13,14]. To infer the locus-specific ancestry along the disease-containing chromosomes in our dataset, we used two state of the art algorithms, LAMP-LD [15] and MULTIMIX [16]. Both programs are well-suited to detect local ancestry from three reference populations and are validated on Latino populations. We used SHAPEIT2 [17] to phase the HGDP dataset to create reference ancestry panels as input for these algorithms.

To resolve the continental origin of E280A, we constructed a reference panel for each of the three continents of potential origin. The Native American reference panel comprises the five Amerindian populations of the HGDP. The African panel contains three West African populations including the HGDP Yoruban and Mandenka and the Yoruba in Ibadan, Nigeria (YRI) from the 1000 Genomes project. We selected the Utah Residents (CEPH) with Northern and Western European ancestry (CEU) to represent European ancestry. We estimated the continental background of each of haplotype spanning E280A in our dataset including those containing the mutation (E280A+) and those without the mutation (E280A-).

To resolve on which specific European background E280A arose, we constructed additional reference ancestry panels from the HGDP and the 1000 Genomes Project to represent North, Western, and Southern Europe.

## 3. Results

### 3.1 Haplotype reconstruction

Reconstructing variant phase reveals a common extended haplotype among all carriers of the E280A mutation comprising 580 single nucleotide polymorphisms (SNPs) and spanning a minimum interval of 1.8 Mb from rs2158987 to rs10135303 and extending through approximately 20 genes (Figure 2, Supp. Table E1). None of the 129 chromosomes lacking E280A share this distinct segment of IBD. Some individuals share a common extended mutation-bearing haplotype spanning over 60 Mb. The presence of this conserved haplotype indicates a single founder event for the E280A allele in Antioquia and the length of the haplotype suggests recent origin. A recombination ~1 Mb telomeric to the mutation marks the beginning of the minimal common haplotype. This recombination likely happened very early in the founding of the population because the mutation-bearing haplotypes are almost equally split between two extended haplotypes upstream of this recombination (Supp. Figure 1). This confounds the determination of the haplotype of the original founder haplotype telomeric to E280A. Centromeric to E280A, the original founder haplotype can be determined by consensus. Carriers of E280A share  $11.7 \pm 6$  Mb of the ancestral haplotype telomeric to E280A.

### 3.2 Estimated age of E280A

Having reconstructed the phase of the haplotype containing E280A, we next sought to estimate the age of this mutation. Estimating the age of E280A provides insights into history and demography of the Antioquian founder population. The single marker method produced an average estimated TMRCA of 9.9 [95%CI 7.2–12.6] generations (Table 1). This corresponds to approximately 250 [95%CI 179–316] years ago using an average rate of 25 years/generation. This estimate is consistent with genealogical records of this kindred that trace back to ca. 1780 (Figure 1B), coincident with the founding of Yarumal in 1787, a town where many E280A carriers reside.

Modeling the intra-allelic coalescent with DMLE yielded an estimated age of 15.3 [95%CI 11–25] generations (Figure 3), which places the estimated time of origin of the mutation in the early 17th century, a time when waves of Spanish colonizers admixed with the indigenous population of Antioquia. Taken together, these estimates reflect a recent origin of the E280A in Antioquia that is consistent with genealogy and historical events in Antioquia.

### 3.3 Geographic origin of E280A

Latin American populations are recent admixtures of Native American, European, and African ancestry. The Antioquian population is characterized by the admixture of European males with Native American females, with a small proportion of African ancestry introduced through the Atlantic slave trade [18]. After colonization of the region, Antioquia remained fairly isolated [19]. An interesting question is whether the E280A arose on a European, Native American, or African genetic background.

The goal of local ancestry estimation is to determine the ancestral origin of each segment along a diploid chromosome. As a first approach to determine the ancestry of the E280A allele, we constructed a phased reference panel and simply checked whether any individuals in this panel shared the common extended haplotype identified in our study. Remarkably, only one individual in the entire reference panel comprising >2000 genomes shared the common haplotype. Sequenced in the 1000 Genomes Project, this individual was selected from Colombia in Medellin (CLM), the capital of Antioquia. Because this individual is a member of a sequencing trio, deterministic phasing guarantees a high accuracy of haplotype reconstruction. Thus, while not providing information on the geographic origin of the E280A allele, this finding reflects the accurate haplotype reconstruction in our study.

We constructed a reference panel for each of the three continents of potential origin of E280A (Figure 4A) and estimated the continental background of each haplotype spanning E280A in our dataset including those containing the mutation (E280A+) and those without the mutation (E280A-). Both algorithms of local ancestry estimation unambiguously estimate a European ancestry of E280A (Figure 4B) allowing us to exclude the possibility of an African or Native American origin of this allele. In contrast, the local ancestry estimation of the haplotypes not containing E280A had a mixture of continental genetic backgrounds, in similar proportions to overall admixture estimates of other samples from Antioquian populations [20, 21].

To further resolve the specific background on which E280A arose, we constructed reference panels representing North, West, and South Europe (Figure 4C). Both LAMP-LD and MULTIMIX estimated that E280A arose on a Western European genetic background (Figure 4D). Due to the low numbers of haplotypes sampled in our reference populations, and the lack of an exact haplotype match in any of these panels, it is not possible at this time to further resolve the geographic origin of E280A within Europe using these reference panels.

Finally, we asked whether E280A arose on a background from a genetically isolated population such as Basque, Bedouin, Adygei, and Sardinian. While previous studies have suggested possible Basque or Sephardic contributions to the Antioquian population [18], our local ancestry analysis did not support these as likely genetic backgrounds on which the E280A mutation arose.

## 4. Discussion

We have estimated the age and geographic origin of the autosomal-dominant point mutation E280A in PSEN1 that causes early-onset familial Alzheimer disease. An extended and conserved haplotype containing E280A in all carriers of this mutation indicates a recent single founder event that defines an extended kindred. With an estimated TMRCA of E280A of ten generations ago and a predicted de novo mutation event around 15 generations ago, this mutation likely arose in the post-Columbian era, early in the 17th century during the Spanish colonization of Colombia.

The predicted Western European ancestry of the haplotype on which E280A resides is consistent with a Spanish origin of the initial mutation carrier and founder of the E280A extended kindred. Given the recent origin and founder effect in the EOFAD cohort, genome-wide local ancestry estimation and subsequent admixture mapping may reveal additional variants that influence the disease in this population [22].

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

Support for this work came from NIH Fogarty grant R01 AG029802 (KSK), the Errett Fisher Foundation (KSK), the University of Luxembourg – Institute for Systems Biology Program, Colciencias Grants 111540820543 and 111540820512 (FL). The funding sources had no role in the data collection, data analysis, data interpretation, or writing of the report. The corresponding author had access to all the data in the study and had final responsibility for the decision to submit for publication.

## References

1. Acosta-Baena N, Sepulveda-Falla D, Lopera-Gomez CM, Jaramillo-Elorza MC, Moreno S, Aguirre-Acevedo D, et al. Pre-dementia clinical stages in presenilin 1 E280A familial early-onset Alzheimer's disease: a retrospective cohort study. *Lancet Neurol.* 2011; 10:213–20. [PubMed: 21296022]
2. Cornejo W, Lopera F, Uribe CS, Salinas M. 1987 Descripcion de una familia con demencia presenil tipo. *Alzheimer Acta Med Colomb.* 1987; 12:55–61.

3. Lopera F, Arcos M, Madrigal L, Kosik KS, Cornejo W, Ossa J. Demencia tipo Alzheimer con agregacion familiar en Antioquia, Colombia. *Acta Neurol Colomb.* 1994; 10:173–87.
4. Lopera F, Ardilla A, Martinez A, Madrigal L, Arango-Viana JC, Lemere CA, et al. Clinical features of early-onset Alzheimer disease in a large kindred with an E280A presenilin-1 mutation. *JAMA.* 1997; 277:793–9. [PubMed: 9052708]
5. Lalli MA, Garcia G, Madrigal L, Arcos-Burgos M, Arcila ML, Kosik KS, et al. Exploratory data from complete genomes of familial Alzheimer disease age-at-onset outliers. *Hum Mutat.* 2012; 33:1630–4. [PubMed: 22829467]
6. Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science.* 2010; 327:78–81. [PubMed: 19892942]
7. Palin K, Campbell H, Wright AF, Wilson JF, Durbin R. Identity-by-descent-based phasing and imputation in founder populations using graphical models. *Genet Epidemiol.* 2011; 35:853–60. [PubMed: 22006673]
8. Risch N, Deleon D, Ozelius L, Kramer P, Almasy L, Singer B, et al. Genetic analysis of idiopathic torsion dystonia in Ashkenazi Jews and their recent descent from a small founder population. *Nat Genet.* 1995; 9:152–9. [PubMed: 7719342]
9. Ichida K, Hosoyamada M, Kamatani N, Kamitsuji S, Hisatome I, Shibasaki T, et al. Age and origin of the G774A mutation in SLC22A12 causing renal hypouricemia in Japanese. *Clin Genet.* 2008; 74:243–51. [PubMed: 18492088]
10. Reeve JP, Rannala B. DMLE+: Bayesian linkage disequilibrium gene mapping. *Bioinformatics.* 2002; 18:894–5. [PubMed: 12075030]
11. International HapMap Consortium. A second generation human haplotype map of over 3. 1 million SNPs. *Nature.* 2007; 449:851–61. [PubMed: 17943122]
12. 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature.* 2012; 491:56–65. [PubMed: 23128226]
13. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, et al. Genetic structure of human populations. *Science.* 2002; 298:2381–5. [PubMed: 12493913]
14. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, et al. Ancient admixture in human history. *Genetics.* 2012; 192:1065–93. [PubMed: 22960212]
15. Baran Y, Pasaniuc B, Sankararaman S, Torgerson DG, Gignoux C, Eng C, et al. Fast and accurate inference of local ancestry in Latino populations. *Bioinformatics.* 2012; 28:1359–67. [PubMed: 22495753]
16. Churchhouse C, Marchini J. Multiway admixture deconvolution using phased or unphased ancestral panels. *Genet Epidemiol.* 2013; 37:1–12. [PubMed: 23136122]
17. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods.* 2013; 10:5–6. [PubMed: 23269371]
18. Carvajal-Carmona LG, Soto ID, Pineda N, Ortiz-Barrientos D, Duque C, Ospina-Duque J, et al. Strong Amerind/white sex bias and a possible sephardic contribution among the founders of a population in northwest Colombia. *Am J Hum Genet.* 2000; 67:1287–95. [PubMed: 11032790]
19. Bedoya G, Montoya P, Garcia J, Soto I, Bourgeois S, Carvajal L, et al. Admixture dynamics in Hispanics: a shift in the nuclear genetic ancestry of a South American population isolate. *Proc Natl Acad Sci U S A.* 2006; 103:7234–9. [PubMed: 16648268]
20. Galanter JM, Fernandez-Lopez JC, Gignoux CR, Barnholtz-Sloan J, Fernandez-Rozadilla C, Via M, et al. Development of a panel of genome-wide ancestry informative markers to study admixture throughout the Americas. *PLoS Genet.* 2012; 8:e1002554. [PubMed: 22412386]
21. Bryc K, Velez C, Karafet T, Moreno-Estrada A, Reynolds A, Auton A, et al. Colloquium paper: genome-wide patterns of population structure and admixture among Hispanic/Latino populations. *Proc Natl Acad Sci U S A.* 2010; 2:8954–61. [PubMed: 20445096]
22. Chakraborty R, Weiss KM. Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. *Proc Natl Acad Sci U S A.* 1988; 85:9119–23. [PubMed: 3194414]

**Systematic Review**

We performed a systematic review by searching Pubmed for the term “E280A.” We included all studies concerning the Antioquian kindred affected by early-onset familial Alzheimer’s disease. Although a Spanish origin was suggested for the E280A mutation, no genetic study to date definitely demonstrates this.

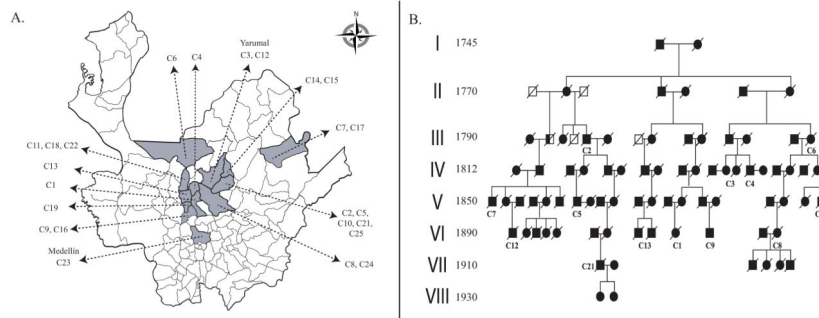
**Interpretation**

Our study estimates an age and geographic origin of E280A consistent with a single founder dating from the time of the Spanish Conquistadors who began colonizing Colombia in the early 16th century.

**Future Directions**

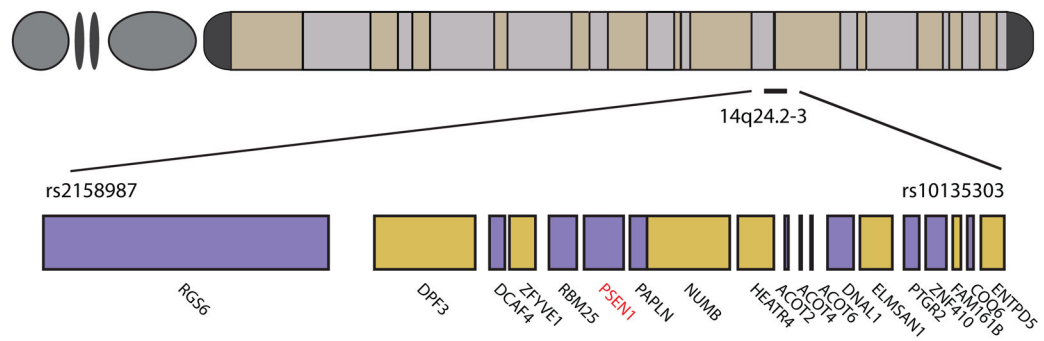
Using the same methods of local ancestry estimation demonstrated in this work, we will expand our analysis genome-wide in the same individuals. Genome-wide local ancestry estimation and subsequent admixture mapping may reveal additional variants that influence the disease in this population.





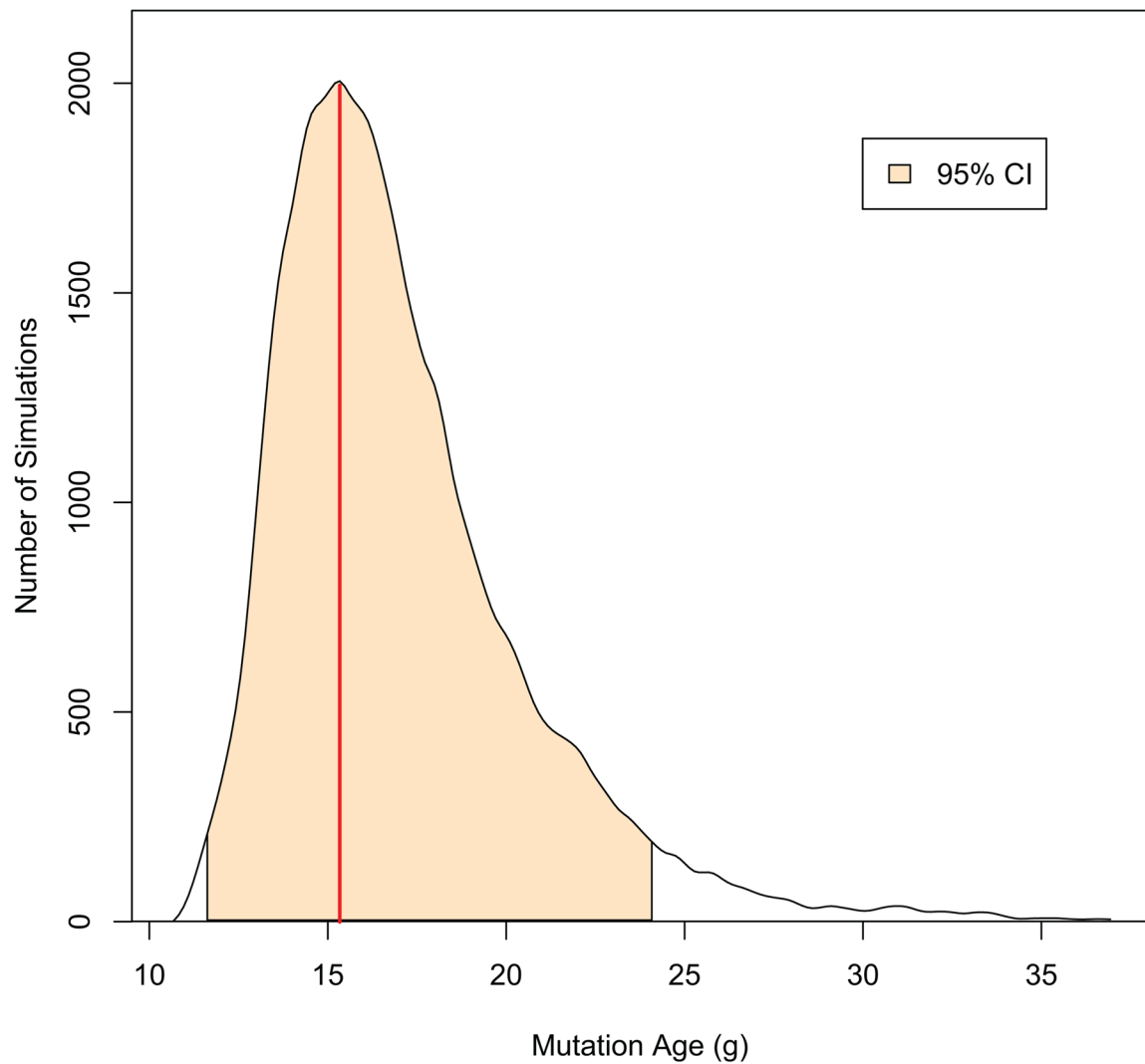
**Figure 1. Geographical map of E280A carriers and a pedigree reconstructed from historical and genealogical accounts**

A) Twenty-four families affected with EOFAD caused by E280A are highlighted on the map of the Department of Antioquia, Colombia. B) For 13 of the recruited families with E280A, genealogical records demonstrate a common pair of founders in 1745. The living members of this family are distributed throughout most of the regions indicated in panel A, consistent with the hypothesis that all affected individuals in Antioquia share an even more distant common founder with a similar dispersal pattern of descendants. EOFAD, early-onset familial Alzheimer's disease.



**Figure 2. Minimal haplotype containing E280A**

All carriers of E280A share a minimal common haplotype spanning 1.8 Mb surrounding PSEN1 E280A from rs2158987-rs10135303 (chr14:72696625-74484310). Positions are given in GRCh37 coordinates. For a complete list of the SNPs on this haplotype, please refer to Supp. Table E1.



**Figure 3. Estimation of mutation age by DMLE**

After 100,000 burn-in runs, the age of the mutation was simulated 100,000 times by DMLE. Mutation age is estimated in generation time, g. A 95% credible interval (CI) is shaded between 11–24 generations. A red line indicates the maximum likelihood estimated age of the mutation at 15.3 generations.



#### Figure 4. Geographic origin of the haplotype spanning E280A

Using the ancestry reference panel comprising the populations depicted in (A), LAMP-LD and MULTIMIX unanimously estimate a European ancestry of E280A in (B) for the haplotypes containing E280A (E280+). Continental ancestry estimation is also shown for haplotypes spanning the region that do not contain the mutation (E280A-). Similarly, using the reference ancestry panels depicted (C), both algorithms strongly support a Western European origin for haplotypes bearing E280A (D). Abbreviations: CEU, Utah Residents (CEPH) with Northern and Western European ancestry. FIN, Finnish in Finland. GBR, British in England and Scotland. HGDP, Human Genome Diversity Project. IBS, Iberian population in Spain. TSI, Toscani in Italia. YRI, Yoruba in Ibadan, Nigeria.

Table 1

Single marker estimates of time to most recent common ancestor (TMRCA) of E280A.

Marker	Pd*	Pn <sup>†</sup>	θ <sup>‡</sup>	Generations	Age <sup>§</sup>
rs1504606	0.68	0.41	0.15	4.8	120
rs11850859	0.73	0.57	0.13	7.2	180
rs61992164	0.39	0.36	0.12	24.1	600
rs191808201	0.35	0.17	0.12	12.0	300
rs1106494	0.81	0.59	0.10	5.9	150
rs55921409	0.89	0.61	0.07	4.4	110
rs2332457	0.87	0.62	0.03	15.3	380
rs8020803	0.89	0.5	0.03	8.6	210
<b>E280A</b>	-	-	-	-	-
rs205813	0.79	0.5	0.05	10.6	270
rs117346379	0.72	0.31	0.07	7.5	190
rs1652593	0.67	0.57	0.10	14.2	350
rs61985647	0.57	0.28	0.12	7.0	180
rs61988101	0.53	0.34	0.14	8.2	210
rs36038796	0.47	0.36	0.17	9.5	240

\* Pd, frequency of a marker on a haplotype bearing E280A.

<sup>†</sup> Pn, overall frequency of a marker in the sample population.

<sup>‡</sup> θ, recombination rate between a given marker and E280A.

<sup>§</sup> Conversion from generation to age uses a rate of 25 years/generation. Age rounded to nearest decade.