University of California
Santa Barbara

The influence of demographic history on the genetics of neurodegenerative illnesses in Colombia

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Molecular, Cellular and Developmental Biology

by

Juliana Acosta Uribe

Committee in charge:

Professor Kenneth S. Kosik, Chair
Professor Julie H. Simpson
Professor Thomas L. Turner
Professor Francisco Lopera, Universidad de Antioquia
Professor Jennifer Yokoyama, University of California San Francisco

September 2021

The Dissertation of Juliana Acosta Uribe is approved.

Professor Julie H. Simpson

Professor Thomas L. Turner

Professor Francisco Lopera, Universidad de Antioquia

Professor Jennifer Yokoyama, University of California San Francisco

---

Professor Kenneth S. Kosik, Committee Chair

August 2021

The influence of demographic history on the genetics of neurodegenerative illnesses in Colombia

Juliana Acosta Uribe

"Nothing in biology makes sense except in the light of evolution"

Theodosius Dobzhansky, 1973

"Nothing in evolutionary biology makes sense except in the light of population genetics"

Michael Lynch, 2007

To my mother,

and to Francisco, Hugo, Orlando and Octavio,

the constant men in my life.

Acknowledgements

It is incredible to look back and see how time flew by so quickly. How all of the sudden, the projects we started in 2013 by traveling to the outskirts of Medellín and gathering samples from all around Colombia became published papers with tremendous impact on how we now understand the genetic landscape of neurodegenerative illnesses in Colombia and Latin America.

First and foremost, I want to thank my two mentors Ken Kosik and Francisco Lopera for being visionaries and creating this unique setting for research and collaboration, this exceptional partnership where we are performing great science with an impact in the community. Furthermore, I want to express my gratitude to my advisor Dr. Kenneth Kosik for believing in me and pushing me to be the best scientist I could be. The transition from clinical research to molecular genetics has not been easy, but thanks to his support and trust I was able to complete this dissertation.

The results I am presenting in this dissertation would not have been possible without the hard work of the amazing team of the Grupo de Neurociencias de Antioquia at the Universidad de Antioquia. I will be forever in debt to David Aguillón and Lucía Madrigal for all the time and effort they invested in making our projects come together, and traveling from the north to the south of the country to gather samples and phenotypic data. I also want to thank Margarita Giraldo and Laura Ramirez for helping with the clinical analyses and the insightful discussions about genotype-phenotype correlation.

I want to thank Nick Cochran at HudsonAlpha for his tremendous collaboration with the sequencing and analyses of over a thousand genomes, and for all his mentorship while I took my first steps in bioinformatics. I acknowledge the support of Jennifer Yokoyama and her lab at UCSF, specially Ethan Geier for his immense patience while teaching me the basics of programming and for introducing me to two of the most fascinating com-

putational tools: Plink and VLookup. Additionally, I appreciate the help from Sharon Browning at the University of Washington, whose thoughtful advice and insight were fundamental for my research.

The cooperative spirit and interdisciplinarity at UCSB is remarkable. I wish to thank Pilar Ramírez from the History department for her enormous help with the historical background of this dissertation. The other researchers at the Kosik Lab, specially Ada Madejska and Israel Hernández, and our undergraduate research assistants Sarah Labib and Rijul Singhal for their collaboration. The Molecular Cellular and Developmental Biology department, especially Kathy Foltz the graduate student advisor and Julie Simpson and Thomas Turner, members of my doctoral dissertation committee.

Ultimately I want to thank my exceptional colleagues at UCSB, Caroline Ackley, Stephanie Khairallah, Carolina Camargo and Brad Killingsworth for their friendship, support and time invested in explaining me the intricacies of molecular and cell biology. But above all, I want to thank my family for their unconditional love and support throughout this journey.

Curriculum Vitæ

Juliana Acosta Uribe

Education

2021	**Ph.D. in Molecular, Cellular and Developmental Biology,** University of California, Santa Barbara.
2019	**M.A. in Molecular, Cellular and Developmental Biology,** University of California, Santa Barbara.
2013	**M.D. (Médico y Cirujano)** Universidad de Antioquia, Medellin, Colombia

Professional Employment

2016 - Present	**Graduate student researcher.** Neuroscience Research Institute. University of California Santa Barbara.
2020-2021	**Project Sponsor and Mentor.** Data Science Capstone Class. Department of Statistics and Applied Probability and the Neuroscience Research Institute. University of California Santa Barbara.
2018-2020	**Teaching Associate.** Neurohumanities: Mind, Brain, Memory And The Musicoliterary. Department of Comparative Literature, University Of California Santa Barbara.
2016-2018	**Teaching Assistant.** Molecular Genetics II: Eukaryotes and Molecular Genetics Laboratory. Department of Molecular, Cellular, and Developmental Biology, University of California Santa Barbara.
2014-2016	**Clinical instructor of Medical Semiology.** School of Medicine, Universidad de Antioquia. Medellin - Colombia
2013 - 2016	**Primary care physician and sub-investigator.** Grupo de Neurociencias de Antioquia, School of Medicine, Universidad de Antioquia. Medellin - Colombia
2007 - 2013	**Undergraduate research assistant (Joven Investigador)**. Grupo de Neurociencias de Antioquia, School of Medicine, Universidad de Antioquia. Medellin - Colombia

Publications

- Acosta-Uribe J∗, Aguillón D∗, Cochran JN, Giraldo M, Madrigal L, Killingsworth BW, Singhal R, Labib S, Alzate D, Velilla L, Moreno S, García GP, Saldarriaga A, Piedrahita F, Hincapie L, López HE, Perumal N, Morelo L, Vallejo D, Solano JM, Reiman EM, Surace EI, Itzcovich T, Allegri R, Sánchez-Valle R, Matallana D, Myers RM, Browning SR, Lopera F, Kosik KS. **A Neurodegenerative Disease Landscape of Rare Mutations in Colombia Due to Founder Effects.**
  [Currently under review as of August 2021]
  ∗Co-first authors

- Cochran JN∗, Acosta-Uribe J∗ , Madrigal L, Aguillón D, García GP, Giraldo MM, Acosta-Baena N, Piedrahita F, Alzate D, López HE, Roberts K, Absher D, Myers RM, Lopera F, Kosik KS. Genetic **Associations with Age at Dementia Onset in the *PSEN1* E280A Colombian Kindred.** Pre-print medRxiv 2020 September 25 doi:https://doi.org/10.1101/2020.09.23.20198424
  ∗Co-first authors

- Arboleda-Velasquez JF, Lopera F, O'Hare M, Delgado-Tirado S, Marino C, Chmielewska N, Saez-Torres KL, Amarnani D, Schultz AP, Sperling RA, Leyton-Cifuentes D, Chen K, Baena A, Aguillon D, Rios-Romenets S, Giraldo M, Guzmán-Vélez E, Norton DJ, Pardilla-Delgado E, Artola A, Sanchez JS, Acosta-Uribe J, Lalli M, Kosik KS, Huentelman MJ, Zetterberg H, Blennow K, Reiman RA, Luo J, Chen Y, Thiyyagura P, Su Y, Jun GR, Naymik M, Gai X, Bootwalla M, Ji J, Shen L, Miller JB, Kim LA, Tariot PN, Johnson KA, Reiman EM, Quiroz YT. **Resistance to autosomal dominant Alzheimer's disease in an APOE3 Christchurch homozygote: a case report.** Nat Med. 2019 Nov;25(11):1680-1683. doi: 10.1038/s41591-019-0611-3. Epub 2019 Nov 4. PMID: 31686034

- Ramirez Aguilar L∗, Acosta-Uribe, J∗ , Giraldo MM, Moreno S, Baena A, Alzate D, Cuastumal R, Aguillón D, Madrigal L, Saldarriaga A, Navarro A, Garcia GP, Aguirre-Acevedo DC, Geier EG, Cochran JN, Quiroz YT, Myers RM, Yokoyama JS, Kosik KS, Lopera F. **Genetic origin of a large family with a novel *PSEN1* mutation (Ile416Thr)**. Alzheimer's & Dementia. 2019 May;15(5):709-719. doi: 10.1016/j.jalz.2018.12.010. Epub 2019 Feb 10. PMID: 30745123
  ∗Co-first authors

- Rios-Romenets, S., Lopez, H., Lopez, L., Hincapie, L., Saldarriaga, A., Madrigal, L., Piedrahita, F., Navarro, A., Acosta-Uribe, J , Ramirez, L., Giraldo, M., Acosta-Baena, N., Sánchez, S., Ramos, C., Muñoz, C., Baena, A., Alzate, D., Ospina, P., Langbaum, J.B., Cho, W., Tariot, P.N., Paul, R., Reiman, E.M. and Lopera, F. (2017), **The Colombian Alzheimer's Prevention Initiative (API) Registry**. Alzheimer's & Dementia, 13: 602-605. https://doi.org/10.1016/j.jalz.2016.09.010

- Lalli MA, Bettcher BM, Arcila ML, Garcia G, Guzman C, Madrigal L, Ramirez L, Acosta-Uribe, J , Baena A, Wojta KJ, Coppola G, Fitch R, de Both MD, Huentelman MJ, Reiman EM, Brunkow ME, Glusman G, Roach JC, Kao AW, Lopera F, Kosik KS. **Whole-genome sequencing suggests a chemokine gene cluster that modifies age at onset in familial Alzheimer's disease.** Mol Psychiatry. 2015 Nov;20(11):1294-300. doi: 10.1038/mp.2015.131. Epub 2015 Sep 1. PMID: 26324103;

Awards

2021	Continuation of the "Charles A. Storke Fellowship" for outstanding research record and interdisciplinary research investigating genomics of neurodegenerative conditions such as Alzheimer's Disease. Awarded by: University of California Santa Barbara
2020	"Charles A. Storke Fellowship" for outstanding research record and interdisciplinary research investigating genomics of neurodegenerative conditions such as Alzheimer's Disease. Awarded by: University of California Santa Barbara
2019	Doctoral Student Travel Grant. Awarded by: Academic Senate, University of California Santa Barbara
	Scholarship to attend the Summer Institute in Statistical Genetics at the University of Washington. Awarded by: University of Washington
	Travel award to attend the Summer Institute in Statistical Genetics at the University of Washington. Awarded by: The Neuroscience research institute, University of California Santa Barbara
	Tau Consortium Fellowship Awarded by: The Tau consortium and Rainwater Foundation
2018	Tau Consortium Fellowship Awarded by: The Tau consortium and Rainwater Foundation

Organizations and Affiliations

2021 - Present	The American Society of Human Genetics
2020 - Present	The Latin America and the Caribbean Consortium on Dementia
2019 - 2021	Graduate Biology Mentorship Association, UCSB
2016 - 2021	Graduate Union of Molecular Biology Investigators, UCSB

Abstract

The influence of demographic history on the genetics of neurodegenerative illnesses in Colombia

by

Juliana Acosta Uribe

High throughput genomic technologies have radically changed the way we understand the genetic landscape of disease. As the field of genomic medicine is growing exponentially, challenges like "lack of transferability" of genetic studies (such as polygenic risk scores) between populations arise. One of the main reasons for this issue, is that recent evolutionary history has created differences in the genetic architecture for disease between human populations. As an example, rare variants show higher geographic clustering and tend to be population-specific. Linkage disequilibrium structure and haplotype blocks around common variants are affected by demographic history as well. One way to circumvent this obstacle is to incorporate diverse populations into genetics studies. Admixed cohorts have proven to be particularly valuable to identify genetic risk for illnesses that are stratified between ancestral origins. This dissertation presents a series of genomic analyses on Colombian individuals with neurodegenerative diseases. We demonstrate that the demographic history of this population affected the genetic burden for neurological disorders, and that by studying individuals with genetic forms of these diseases, we can expand our understanding of the genetic basis of neurodegeneration.

The Colombian population, as well as those in other Latin American regions, arose from a recent tri-continental admixture among Native Americans, Spanish invaders and enslaved Africans, all of whom passed through a population bottleneck due to widespread infectious diseases that left small isolated local settlements. As a result, the current

population reflects multiple founder effects derived from diverse ancestors.

We performed whole genome sequencing in a large cohort of Colombian individuals with Alzheimer's disease (AD), frontotemporal lobar degeneration-motoneuron disease continuum, early onset dementia and healthy participants. We analyzed their global and local ancestry proportions and screened this cohort for deleterious variants in disease-causing and risk-conferring genes. Then, we present a deeper analysis of one of the families with genetic AD due to a novel variant in *PSEN1* (Ile416Thr) of African origin. We describe the phenotype of the symptomatic carriers, as well as their brain accumulation of amyloid-β and Tau prior to disease onset. Lastly, we explore genetic modifiers of the age at onset for dementia in a large family with autosomal dominant AD due to *PSEN1* Glu280Ala. We performed a classic whole genome association study and a novel approach for genetic association using a package that performs a likelihood ratio test with a linear mixed model to adjust for relatedness between individuals.

The genomes revealed multiple rare mutations associated with various forms of adult onset familial dementias. Most of these mutations originated from founders and, remarkably, when the entire founder set of mutations was considered together, the genetic consequences of the local demographic histories emerged. In addition, we identified dozens of genome-wide significant loci that modified the age at onset for AD in the Glu280Ala kindred. We also observed a substantial number of individuals with an age of onset well beyond the typical age of onset for this kindred, some of which had high impact coding variants with effect sizes similar to the *APOE* Christchurch variant (Arg154Ser) recently described in another study.

The results here reveal an unexpected genetic richness in a large Colombian cohort selected for the presence of neurodegenerative conditions affecting cognition. Our results suggests that the demographic history of Colombia is likely to underlie the modern clustering of familial neurodegenerative diseases arising from multi-ancestral rare disease-

associated alleles. It additionally reinforces the value of these large families with genetic neurodegenerative diseases as platforms for genetic discovery. This set is to our knowledge the largest published study in the literature of the genetics of dementia in a Hispanic descent population. Furthermore, this dissertation underscores the numerous insights that can emerge from Latin American population and the importance of inclusiveness in future genetic studies.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Alzheimer's disease (AD) and related dementias are among the 10 leading health conditions associated with disability and death. Unlike the other conditions in this list, there is still no preventive or curative treatment for AD or other neurodegenerative dementias. While deaths associated with heart disease, stroke, breast and prostate cancer have decreased in the last 20 years, AD has risen significantly. The incidence of dementia increases exponentially with age and the prevalence of dementia is expected to rise as the average life expectancy increases. The World Health Organization predicts that 75 million people by 2030 will have dementia, which could grow to 135 million in 2050[1].

Although age is the principal risk factor for dementia, genetic risk factors contribute to 58-79% of AD heritability[2]. Up to 60% of individuals with early onset AD (clinically defined as onset before 65 years) have familial history of AD, and 13% show autosomal dominant inheritance[3]. Despite the advances in understanding the genetic basis of these illnesses, most of the research has been performed in European white non-Hispanic population. While multiple studies have suggested that populations like African and Hispanic have higher per capita risk to develop AD than white/European populations[4,

5, 6], they only represent 2.4% and 1.3% respectively, of individuals included in Genome Wide Association Studies (GWAS)[7, 8] perpetuating European bias in genomic studies.

This dissertation embodies an effort to expand our knowledge of the genetic landscape of neurodegenerative diseases in Latin America, as recent efforts have demonstrated that including diverse populations and multi-ethnic participants in genetic studies increases genetic discovery[9]. **Chapter 2** presents a unique approach to understand the genetic burden for neurogenerative diseases in a Colombian cohort that includes patients with AD, frontotemporal lobe degeneration, motor neuron disease, as well as other types of early onset dementia. We analyzed how the demographic history of the country and its population dynamics led to a significant prevalence of mendelian forms of neurodegenerative illnesses and the confluence of multi-ancestral genetic risk factors. **Chapter 3** dives deeply into the phenotypic characteristics and ancestral origins of a family with a novel genetic variant that causes an autosomal dominant form of AD, while **Chapter 4** focuses in identifying genetic modifiers for age at onset of dementia in a large cohort of patients with genetic AD.

1.1 Genetic basis of Alzheimer's disease

Alzheimer's disease is a neurodegenerative illness characterized by episodic memory loss, often accompanied by executive dysfunction, behavioral symptoms and speech impediments. Although the entire pathophysiology of AD is still not known, extracellular aggregates of amyloid-β and intracellular hyper-phosphorylated Tau [10] appear to trigger neuronal cell death and correlate with the extensive neuronal loss of the illness.

While the precise etiology of AD has not been discovered, disease-causing genetic variants have been reported in the Amyloid Precursor Protein (*APP*) gene in chromosome 21, Presenilin 1 gene (*PSEN1*) in chromosome 14 and Presenilin 2 gene (*PSEN2*) in chro-

mosome 1 [11]. In addition, the most common genetic risk factor for AD is the E4 allele of the apolipoprotein-E (*APOE*) gene. Relative to individuals carrying the most common allele APOE E3, a single copy of APOE E4 increases the risk of illness approximately 4-fold, and two copies of APOE 4 increase it by 12-fold in European descent individuals [11]. More recently, GWAS in late-onset AD have identified risk-conferring variants in new genes involved in diverse biological pathways, such as cholesterol metabolism (*CLU*, *ABCA7*), immune response (*CR1*, *CD33*, *MS4A*, *TREM2*) and endocytosis (*BIN1*, *PICALM*, *CD2AP*, *EPHA1*, *SORL1*)[12, 13, 14]. These pathways continue to unravel the depth of complexity of the pathophysiology of AD and related dementias and shed light on the molecular mechanisms of disease.

1.2 Frontotemporal lobe degeneration

AD shares many of the clinical phenotypes and underlying neurodegeneration pathways, such as Tau and Tar-DNA binding protein (TDP-43) deposition, with a series of neurodegenerative diseases grouped under the umbrella term of Frontotemporal lobe degeneration (FTLD) [15, 16].FTLD is a spectrum of clinically and neuropathologically heterogeneous disorders characterized by significant changes in behavior or language, accompanied by degeneration or dysfunction of the frontal and/or temporal lobes. FTLD encompasses several syndromes including behavioral variant of frontotemporal dementia (bvFTD), semantic variant of primary progressive aphasia (svPPA) and nonfluent/agrammatic variant of primary progressive aphasia (nfvPPA). FTLD is also associated with other motor neuron syndromes such as Amyotrophic Lateral Sclerosis (ALS) progressive supra-nuclear palsy (PSP) and corticobasal degeneration (CBD). These entities have been grouped under the term FTLD and motor neuron disease continuum (FTLD-MND)[17, 18].

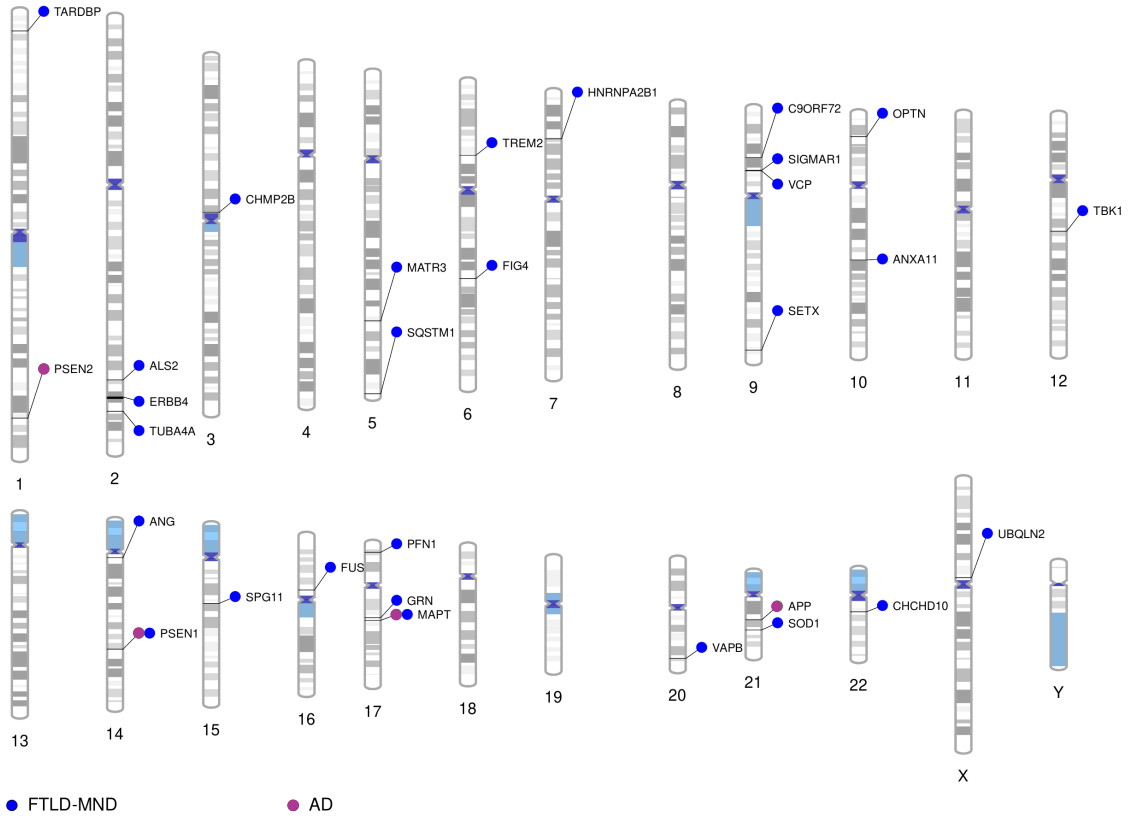


Figure 1.1: Disease causing genes for AD and FTLD-MND

FTLD-MND has demonstrated significant heritability in genetic studies, wherein 40% of the patients have familial history of dementia or ALS, while 10.2-27% have autosomal dominant forms of the illness [19]. However, the mutations associated with these conditions do not always map precisely to the clinical syndromes, and show significant overlap between the multiple syndromes that comprise the FTLD-MND spectrum. The genes most commonly associated with FTLD-MND include chromosome 9 open reading frame 72 (*C9orf72*), progranulin (*GRN*), and microtubule associated protein tau (*MAPT*). When examined from a neuropathologic perspective, both *C9orf72* and *GRN* mutations are associated with TAR DNA-binding protein-43 (TDP-43) inclusions, and *MAPT* pathogenic variants with tau inclusions. In addition, pathogenic variants in

TAR DNA-binding protein 43 (*TARDBP*), RNA-binding protein FUS (*FUS*), charged multivesicular bodyprotein 2B (*CHMP2B*), valosin containing protein (*VCP*), TANK-binding kinase 1 (*TBK1*) and sequestosome 1 (*SQSTM1*) have been associated with these illnesses[20, 21]. *TREM2* has shown association with both FTLD-MND and AD. Homozygous loss-of-function variants can present as FTLD [22], while heterozygous missense variants can be risk-conferring for both AD and FTLD[23, 24].

1.3 The *PSEN1* E280A family

The study of families where a distinct illness or phenotype segregates in a Mendelian fashion has been fundamental for the discovery of the genes that underlie a given disease. In 1997 Lopera et. al described several Colombian families with autosomal dominant AD caused by an adenine to cytosine transversion in *PSEN1* gene[NM_000021: c.839A>C, p.(Glu280Ala)]. This variant is canonically known as *PSEN1* E280A or "Paisa mutation"[25] in reference to their place of discovery.

Thanks to church records of births, baptisms, marriages and deaths, the Neuroscience Group of Antioquia (GNA) was able to reconstruct a large pedigree that linked almost all the carriers of this rare variant, and identified the most recent common ancestor circa 1750. In 2014, a work by Lalli et al[26] demonstrated that all the sequenced carriers of *PSEN1* E280A were identical by descent, and confirmed the hypothesis of a founder effect in the northwest of Colombia. By 2017, the GNA had identified over 6000 members of the family, where ∼1200 of them were confirmed E280A carriers[27] Figure 1.2.

The longitudinal follow-up of this family has proven *PSEN1* E280A to be highly penetrant and with similar phenotype and disease progression (expressivity) in all the carriers. A retrospective study that analyzed the clinical records of 449 *PSEN1* E280A carriers from 1995 to 2010 identified that the median age at onset (AAO) was 44 years

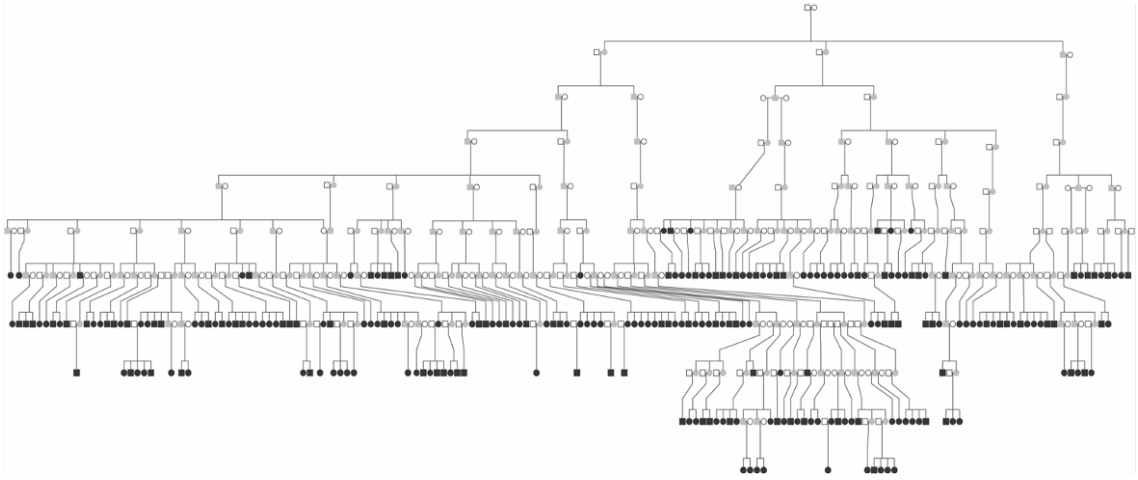


Figure 1.2: Pedigree of the *PSEN1* E280A family

(95% CI 43–45) for mild cognitive impairment (MCI), 49 years (95% CI 49–50) for dementia, and 59 years (95% CI 58–61) was the median age at death[28]. Subsequently, a study used positron emission tomography (PET) scans with Florbetapir to trace the accumulation amyloid-β in 11 symptomatic E280A carriers, 19 pre-symptomatic carriers, and 20 asymptomatic non-carriers (ranging in age from 20 to 56 years). These PET scans demonstrated that the accumulation of fibrillar amyloid-β began at a mean age of 28.2 years (95% CI 27.3–33.4), about 16 years and 21 years before the predicted median AAO for MCI and dementia, and reached plateau at 37.6 years (95% CI 35·3–40·2)[29]. In a similar fashion, PET scans with Flortaucipir were used to trace the Tau accumulation in amyloid-positive E280A carriers, and these results demonstrated medial temporal lobe accumulation of Tau six years before clinical onset of AD[30].

The *PSEN1* E280A family has been fundamental in the process of understanding AD as the variant has shown to be highly penetrant and has low variation around the age at onset, which implies that outliers can be readily detected with high significance. The early onset eliminates the confounding effects of aging when age at onset is studied in older populations and, due to a relatively homogeneous cultural background, many

lifestyle confounders such as diet can also be eliminated. Several projects have tried to identify genetic modifiers of the MCI and dementia AAO for E280A carriers, but have been limited by the number of subjects that were included [26, 31, 32].

The *PSEN1* E280A family is one of the most remarkable examples of deleterious variation arising to significant allelic frequencies as a consequence of a founder effect. The unique demographic history of Colombia favored the discovery of these and other families with similar characteristics, as described thoroughly in Chapter 2.

1.4 Colombia as a platform for genetic discovery

Colombia is located in northwestern South America, in a pivotal location with access to the Atlantic and the Pacific oceans. Since 12,000 BCE it was inhabited by different indigenous groups that settled in the Andean Mountain ranges, the valleys of the Magdalena and Cauca rivers, the Orinoco and Caribbean plains, and the Pacific and Amazon jungles. It is estimated that there were at least 11,332,823 inhabitants at the time of the Spanish invasion[33]. Some of these societies, such as the Cueva, Choco, Panche, Pantágora, Pijao, and Guahíbo, had scattered population patterns, and moved in small groups in search of resources. Others such as the Pasto, Zenú and Quimbaya had settled in villages of hundreds or even thousands of individuals engaged in agriculture. Also notable were the Tairona and Muisca, who were organized into complex political groups of various sizes and maintained medium and long-distance commercial relations with their neighbors[34, 35]. In the 16th century European immigrants deployed military expeditions to conquer the indigenous societies. Most of these immigrants came from the center of present-day Spain, but Portuguese, Italians, Flemish, French and Germans immigrated as well.

Along with the conquerors, thousands of enslaved individuals from sub-Saharan west

Africa were forcibly taken to the region. Although the precise statistics of the Africans who reached the territory of present-day Colombia are not known, 487 ships with 80,000 captives arrived in the port of Cartagena alone from 1570 to 1640[36]. Most of them were from Upper Guinea (present-day Senegal and Sierra Leone) and Angola, and were designated by Portuguese slavers as Mandingas, Fulupos, Balantas, Zapes, Cocolíes, Yolofos, Fulos and Branes. Others, like the Anzicos, Angolas, Congos and Manicongos were speakers of Bantu languages and inhabited the kingdom of Kongo. During the 17th and 18th centuries, the English, Dutch and French introduced slaves mainly from the Cross River in Nigeria, the coast between Elmina and Ghana, the Gold Coast and the Benin gulf through different Caribbean ports, such as Cartagena de Indias, Santa Marta, Riohacha, and Tolú [37, 38, 39, 40]

The forced encounter of these three groups (Native Americans, Europeans and Africans) that had been isolated during thousands of years led to the emergence of a large admixed population. Demographic data from the late 18th century can confirm this: in 1778, the Spanish authorities counted 62,404 slaves (between Africans and descendants of Africans), 156,345 indigenous people, 208,969 whites, and 373,047 were "free people of color" (among mestizos, mulatos and zambos), for a total of 803,452 inhabitants [33, 41]. Soon after European contact, there was a substantial reduction in the Native American population size due mostly to infectious diseases, and to a lesser extent warfare and enslavement. European and African populations had significant mortality as well, reaching a minimum effective population size around 12 generations ago[42, 43]. This bottleneck created serial founder effects, where the survivors were geographically isolated and had an exponential population growth during the following generations.

The shaping of the current Colombian population by admixture and genetic drift offers an incisive discovery platform for genotype-phenotype associations as populations that have undergone bottlenecks and founder effects have a higher mutation load[44].

Further studies of the genomic landscape for a given illness in Latin-American individuals entails the understanding of the fine-scale population structure and demography, as genetic variants that are disease causing or risk conferring depend heavily on the recent history of the population

Chapter 2

A Disease Landscape of Rare Mutations in Colombia due to Founder Effects

2.1 Introduction

The circumstances related to Latin America's unique demographic history led to numerous genetic founders that expanded rare genetic variation. The regional populations of Colombia originated from varying proportions of a recent tri-continental admixture consisting of diverse indigenous peoples, Spanish invaders and enslaved Africans, all of whom had been geographically separated for tens of thousands of years. During the Spanish conquest, these individuals suffered massive mortality from numerous infectious diseases, including smallpox, influenza, syphilis, hepatitis, measles, tuberculosis, diphtheria, cholera, typhus, scarlet fever, and meningitis, which created a narrow bottleneck

with a minimum effective population size approximately 12 generations ago[43]. Survivors were geographically dispersed in a patchwork of relatively isolated small founder populations. Following the first decades of the Spanish invasion and European expansion throughout various territories, the second half of the sixteenth century saw a large and continuous growth of an admixed population, especially in the Andean region of the country [Appendix 2.6]. The population growth amplified the effects of genetic drift confined to highly local settings that marked a fine-grained geographic map with a local genetic stamp[45].

Demographic history and local ancestry have gained significant interest in genomic studies aiming to understand the disease burden of underrepresented populations and transferability of risk scores from research done in European cohorts. However, most of these studies have focused on genome wide association studies (GWAS) and polygenic risk scores that usually rely on the sequencing of common genetic variants[46, 47], while missing those rare alleles absent from European genomes [48]. Rare variants are likely to play a role in the problem of *missing heritability*, have larger effect sizes[49], and are more susceptible to population dynamics and genetic drift.

Rare mutations contribute to the occurrence of neurodegenerative disease, which prompted a search for individuals with young onset familial dementia and related neurodegenerative disorders. We suspected that genetic drift stamped local populations with unique sets of rare variants. Numerous rare genetic conditions converge under this phenotypic label, and therefore as a population indicator of rare variation, dementia represents a readily identifiable trait with a great deal of genetic variation. Among the many genes in which disease mutations fit the phenotypic label are *PSEN1*, *PSEN2*, *APP*, *C9orf72*, *GRN*, *MAPT* , *TARDBP*, *FUS*, *VCP*, *CHMP2B*, and *TBK1*. Rare variants in these genes offer novel perspectives on the breadth of their associated clinical phenotypes and

the underlying molecular pathways. Here we describe a cohort of 900 Colombian individuals with neurodegenerative diseases and report the genetic variants associated with neurodegeneration in the context of their ancestral origins and admixture.

2.2 Subjects and methods

2.2.1 Participants

Nine hundred nineteen individuals from 573 families were recruited or referred to the "Grupo de Neurociencias de Antioquia", University of Antioquia, Colombia for "The Admixture and Neurodegeneration Genomic Landscape" (TANGL) study. The project was approved by the School of Medicine, Universidad de Antioquia Internal Review Board and informed consent following the guidelines of the Code of Ethics of the World Medical Association (Declaration of Helsinki) was obtained from all participants or their legally authorized proxies. The recruitment targeted patients with early onset dementia and families in which multiple first-degree relatives were affected. All the individuals were born in Colombia [Appendix 2.6]. All subjects were evaluated following a standard protocol including physical and neurological examination, as well as population validated neuropsychological assessment [50]. Family history was obtained from the patients and their relatives and was considered positive if at least one first or second degree relative presented dementia or motor neuron disease (MND). Families were classified as Autosomal Dominant if at least three first degree relatives suffered from dementia or MND in two consecutive generations. When patients had familial forms of dementia, their relatives with neurological and psychiatric disorders were recruited along with healthy family members. Based on their clinical diagnosis participants were divided in four cohorts:

• The Alzheimer's Disease (AD) cohort included individuals with early onset AD (AAO<65 years) and individuals with autosomal dominant late onset AD. Patients with atypical presentations of AD, such as primary progressive aphasia - logopenic variant (lvPPA), posterior cortical atrophy, and spastic paraparesis associated with *PSEN1* pathogenic variants were included in this cohort. AD was diagnosed according the NINCDS-ADRDA criteria [51].

• The Frontotemporal lobar degeneration and motor neuron disease (FTLD-MND) spectrum cohort comprised patients with multiple presentations of frontotemporal lobar degeneration (FTLD), which include behavioral variant of frontotemporal dementia (bvFTD), primary progressive aphasia-semantic variant (svPPA), primary progressive aphasia-non-fluent/agrammatic variant (navPPA) and FTLD with amyotrophic lateral sclerosis (FTLD-ALS). Diagnosis of FTLD variants was done according to Gorno-Tempini et al, 2011[17], and Rascovsky et al, 2011[52]. Patients with cortico-basal degeneration (CBD), progressive supranuclear palsy (PSP) diagnosed according to The Movement Disorder Society Criteria [53]) and with amyotrophic lateral sclerosis (ALS), diagnosed according to Strong et al, 2017 [54], were included in this cohort.

• The Early Onset Dementia not otherwise specified (EOD) cohort included patients with early onset dementia (AAO <65 years) that did not fully meet criteria for AD or FTLD at the time of evaluation and did not have secondary causes that explain their neurodegeneration. Some of these individuals were relatives of the patients from the other cohorts but presented with conditions such as Parkinson's Disease, bipolar disorder, or Lewy body disease.

• The Healthy participant cohort included individuals related and unrelated to the patients. These subjects had a CDR score of 0 in their last examination, and no evidence of neurodegenerative dementia or motor neuron disease. The complete demographic

information of the final cohort with high-quality genomic sequencing (n=900) can be found in Table 2.1 and 2.6.

Cohort	**n**	**AAO**		**Female**	
		mean	range	n	%
AD	376	59	30-90	249	66.2
FTLD-MND	197	58.8	21-82	92	46.7
EOD	73	54.5	25-75	49	67.1
Healthy Participants	254	60	18-100 **	159	62.6

Table 2.1: Demographic information of the TANGL cohort. AD: Alzheimer's disease, FTLD-MND: frontotemporal lobar degeneration and motor neuron disorder, EOD: Early onset dementia not otherwise specified. AAO: Age at onset, ** Age at evaluation.

2.2.2 Genome sequencing

Peripheral blood from the participants was obtained by standard phlebotomy and genomic DNA was isolated from leukocytes using the Gentra Puregene Blood Kit (Qiagen). Genome sequencing (WGS) was performed for 800 samples at the HudsonAlpha Institute for Biotechnology on either the Illumina HiSeq X platform, or the Illumina NovaSeq platform. A subset of individuals was sequenced at the Human Longevity Institute on the Illumina HiSeq X platform (119 samples). The combined dataset had a mean read depth of 34X and an average of 92% of bases covered at 20X. Sequencing libraries at HudsonAlpha were prepared by Covaris shearing, end repair, adapter ligation, and PCR using standard protocols. Library concentrations were normalized using KAPA qPCR prior to sequencing. Sequencing reads from both centers were aligned to the hg19 reference genome with bwa-0.7.12. BAMs were sorted and duplicates were marked with Sambamba 0.5.4 [55] Indels were realigned, bases were re-calibrated, and gVCFs were generated with GATK 3.3 [56] Variants were called across all samples in a single batch with GATK 3.8 using the -newQual flag to minimize false negative singleton calls. Genome annotation was

performed using SnpEff 4.3s[57] after splitting multi-allelic sites with Vt. The genome was annotated with the gene definitions from human genome build Ensembl GRCh37.75. All single nucleotide variants and indels were annotated with CADD v1.3 [58]. Population database frequency annotations included 1000 Genomes Phase 3 (1000GP) [59], TOPMed Bravo (lifted over from hg38 to hg19 using CrossMap 0.2.7 [60]), and several population database sets annotated using WGSA 0.7 [61]) including ExAC [62], gnomAD [63], ESP, and UK10K. Variants were also annotated with dbSNP release 151[64].

Calls were filtered with vcftools (v0.1.12b) [65] to retain sites with quality scores equal or greater than 20 and mean read depth scores equal or greater than 30. KING (v2.2.4) [66] was used to verify disclosed familiar relationships and pedigree structures, and individuals with unexplained relatedness were removed. For duplicate samples and monozygotic twin pairs, only one genome was kept. PLINK v.1.90 [67, 68] was used to identify and exclude individuals with discordant X-chromosome sex and those with more than 5% missing data [69]. Mendel errors were set to missing before removing autosomal variants with missingness >5% obtaining a total of 41,123,431 variants and 900 individuals from 566 families available for analysis. [Figure 2.1]

We conducted additional array genotyping measurements using the Array-8+ v1.0 Kit + neuro booster array consortium (NBA) content, beadchip 20042459 Illumina Global Diversity (Catalog 20031816). Imputation was performed using the TOPMed Imputation Panel and Server (version 1.3.3), which includes 97,256 references samples and 308,107,085 variants and uses Minimac4 for imputation. Pre-imputation scripts (version 4.3.0 from William Rayner at the University of Oxford) were run using default settings, which filtered out palindromic single nucleotide variants (SNVs) with minor allele frequency (MAF) >0.4 or variants with >0.2 MAF difference from the TOPMed reference panel. The exome from the Spanish c.1189C>T (p.Pro397Ser) carrier[70]) was processed

from fastq to VCF using a standard clinical alignment pipeline from the HudsonAlpha Institute for Biotechnology Clinical Services Laboratory that uses Sentieon version 201808.07 (a computational wrapper for common tools such as bwa), including alignment with Sentieon-BWA (version 201808.07; identical to bwa mem 0.7.15-r1140) and variant calling with Illumina Strelka2 (version 2.9.10).

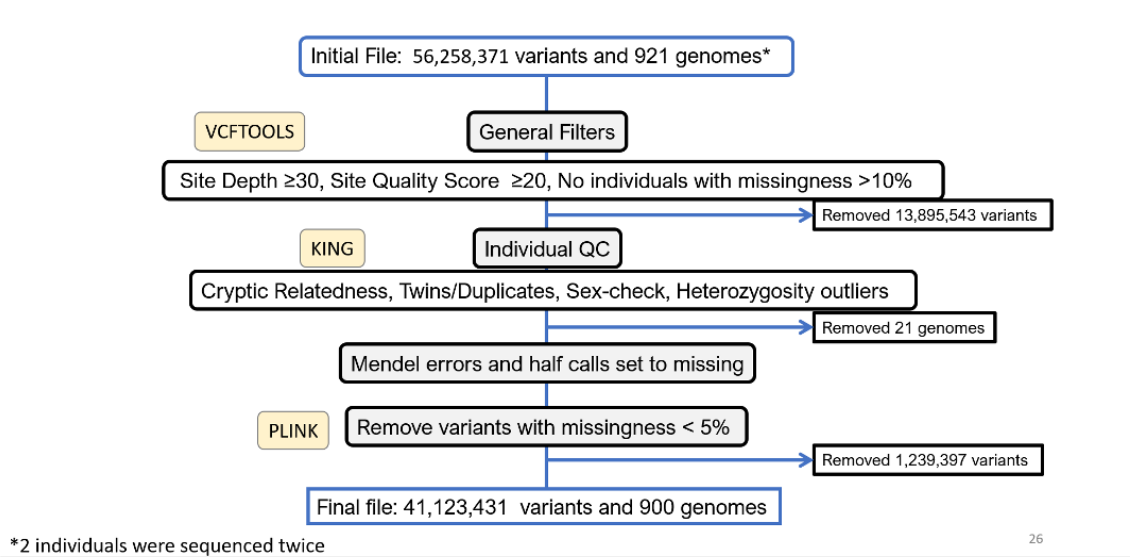


Figure 2.1: Pipeline for whole genome sequence data quality control (QC)

2.2.3 Population structure analysis

We implemented protocols similar to those previously developed for ancestry estimation in admixed populations [71, 46]. We merged the 900 genomes (TANGL cohort) with the 1000 Genomes Project (1000GP) Phase 3 genomes generating the TANGL.1000GP dataset (n=3404). Then, we created a subset including only the TANGL cohort, the non-admixed African Populations (ESN, GWD, LWK, MSL, YRI, N=504) and European populations (CEU, GBR, FIN, IBS and TSI, N=503). We merged these genomes with Native American samples (AYMARAN, MAYAN, NAHUAN, QUECHUAN, N=43) from Mao

et al.[72] inferred to have >0.99 Native Ancestry, and created the TANGL.AFR.EUR.NAT dataset. After removing monomorphic variants, triallelic sites that were not due to a strand flip in either dataset and those sites with missingness greater than or equal to 1%, we retained 845,950 autosomal variants and 1950 individuals for further analysis.

2.2.4 Global ancestry inference

A subset of unrelated samples from TANGL.AFR.EUR.NAT was selected by keeping only the proband of each family and, using KING (v2.2.4)[66] with '–related' and '–degree 3' settings to identify cryptic relatedness. Only sample pairs with kinship coefficient less than 0.044 were retained for TANGL, AFR and EUR. The NAT individuals showed significant relatedness between them, and the threshold for that population was set to '—degree 2' to retain the most NAT samples with kinship less than 0.0884. The final TANGL.AFR.EUR.NAT -Unrelated dataset comprised 1611 unrelated individuals (TANGL N=566, AFR N=501, EUR N=503, NAT=41).

We calculated global ancestry using ADMIXTURE (v.1.3.0)[73] independently for the unrelated TANGL individuals (n=566) and for the TANGL.AFR.EUR.NAT-Unrelated cohort. As recommended by ADMIXTURE, PLINK (v.1.9)[67, 68] was used to perform pair-phased linkage disequilibrium (LD) pruning; excluding variants with an r2 value of greater than 0.2 with any other SNP within a 50-SNP sliding window, advancing by 10 SNPs each time (–indep-pairwise 50 10 0.2). The LD-pruned dataset contained 203,810 variants. We then performed an unsupervised analysis modeling from one to ten ancestral populations (K = 1 – 10) using the random seed option and replicating each calculation 20 times. We selected the run with the best Loglikehood value for each K and compared the Cross Validation (cv) error values to determine the model with the lowest cv value. Ancestral proportion statistics of mean and standard deviation were calculated using

the statistical software R. In addition, we determined mitochondrial and Y-chromosome haplogroups of the TANGL-unrelated cohort using HaploGrep 2 with Phylotree 17[74], and yHaplo respectively [75]

2.2.5 Local ancestry inference

We phased the combined TANGL.AFR.EUR.NAT dataset with SHAPEIT (v.2.r900) [76] using the haplotype reference panel of the 1000GP. We used the parameters –duohmm and a window of 5MB (-W 5), which takes advantage of the inclusion of families, pedigree structure and the large amount of IBD shared by close relatives, leading to increased accuracy [77]. We used the PopPhased version of RFMix (v1.5.4)[78] to estimate the local ancestry using the following flags: -w 0.2, -e 1, -n 5, –use-reference-panels-in-EM, –forward-backward as recommended by Martin et al[46] for estimating local ancestry in admixed populations. To determine the carrier haplotype and local ancestry of a rare variant of interest, we used PLINK(v.1.9)[67, 68]. We identified other single nucleotide variants (SNVs) in linkage disequilibrium with the variant of interest and used them as tags to identify the carrier haplotypes in the phased dataset, and then searched for the local ancestry of the specific locus in the RFMix output.

2.2.6 Principal Component Analysis (PCA)

For PCA, we used the subset of unrelated samples with LD-pruning of variants as described in the methods for "Global ancestry inference". We performed a PCA using the smartpca package from EIGENSOFT(v7.2.1)[79], with 3 outlier removal iterations (numoutlieriter: 3) and flag "altnormstyle: NO" to match EIGENSTRAT normalization formulas[79]. The PCA results were plotted using the PCAviz package[80] for R. For the PCA with the Ancestral populations we retained variants with MAF >10%. For

the PCA of the TANGL-unrelated cohort we extracted a common variant set, retaining those with MAF >10%, and then a lower frequency variant set, keeping only variants with MAF between 5 – 10%.

2.2.7 Genetic screening for disease causing variants

Each individual was initially screened for pathogenic variants in the most recognized genes associated with AD and FTLD according to AD/FTLD mutation databases (https://www.molgen.vib-ua.be/ADMutations, https://www.alzforum.org/mutations); *PSEN1*, *PSEN2*, *APP*, *MAPT*, *GRN VCP*, *FUS*, *CHMP2B*, *TARDBP* and *TBK1*. (The molgen.vib-ua.be/ADMutations database is not available as of July 2021). A secondary genetic analysis was done to identify possibly pathogenic variants in other genes associated with similar or overlapping phenotypes. For the secondary screening, we chose the disease-causing genes reported in the following OMIM phenotypic series and phenotypes; Frontotemporal dementia and/or Amyotrophic Lateral Sclerosis [MIM: PS105550, PS167320, PS105400], Parkinson disease, Adult-Onset Leukoencephalopathies and, Ceroid lipofuscinoses. We retained variants with MAF of 0.001 or less in the ExAC database if the gene had autosomal dominant or X-linked inheritance, and 0.01 or less if the gene had autosomal recessive inheritance. The remaining variants were discarded if they were more prevalent in controls than cases or if they had a CADD Phred score less than 20. The selected protein altering variants, defined as nonsynonymous single nucleotide variants, splicing altering variants, insertions or deletions were manually curated by searching in the databases described before as well as ClinVar and Litvar[81]. The previously unreported (novel) variants were classified according to the guidelines published by the American College of Medical Genetics and Genomics and the Association for Molecular Pathology[82]. Variants in *TBK1* and *PSEN2* were also classi-

fied according the Guerreiro algorithm [83]. Additionally, subjects were screened for *C9orf72* hexanucleotide expansion using repeat-primer following the protocol described in DeJesus-Hernandez et al., 2011 [84]. We searched for large copy number variations using four callers: DELLY[85], ERDS[86], CNVnator[87], and BIC-seq2[88]. Events called by multiple callers were inspected for validity using Integrative Genomics Viewer[89]. More details about the curation process in Figure 2.2.

2.2.8 Genetic screening for risk associated variants

We used publications in the literature to identify genes in which rare variants were associated with increased risk for AD and/or FTLD-MND with an odds ratio higher than 2. *TREM2* [90, 91] , *ABCA7* [92, 93] and *SORL1* [90] were selected as intermediate effect risk genes. We retained variants that were known to be risk conferring, led to premature truncation of the protein (PTV), or that were classified as strictly damaging (SD) according to previous published criteria [90]. Strictly damaging variants had MAF <0.01 in the ExAC database, and were unanimously classified as deleterious by three different in silico prediction algorithms; SIFT [94], Polyphen-2 [95] (Hum Div.), and MutationTaster [96]. In addition to this strategy, we included *ADAM10* [MIM: 602192] c.510G>T (p.Gln170His) and c.541A>T (p.Arg181Gly) variants as they have been reported to confer intermediate risk for AD[97, 98].Variant nomenclature is according to the HGVS Recommendations[99]; the GenBank reference transcripts used for each disease causing and risk conferring variant can be found in appendix 2.5.

2.2.9 Identity by descent

If any of the disease-conferring or risk-associated variants were shared by two or more unrelated individuals, we used hap-IBD[91] v1.0 to search for Identity by Descent (IBD)

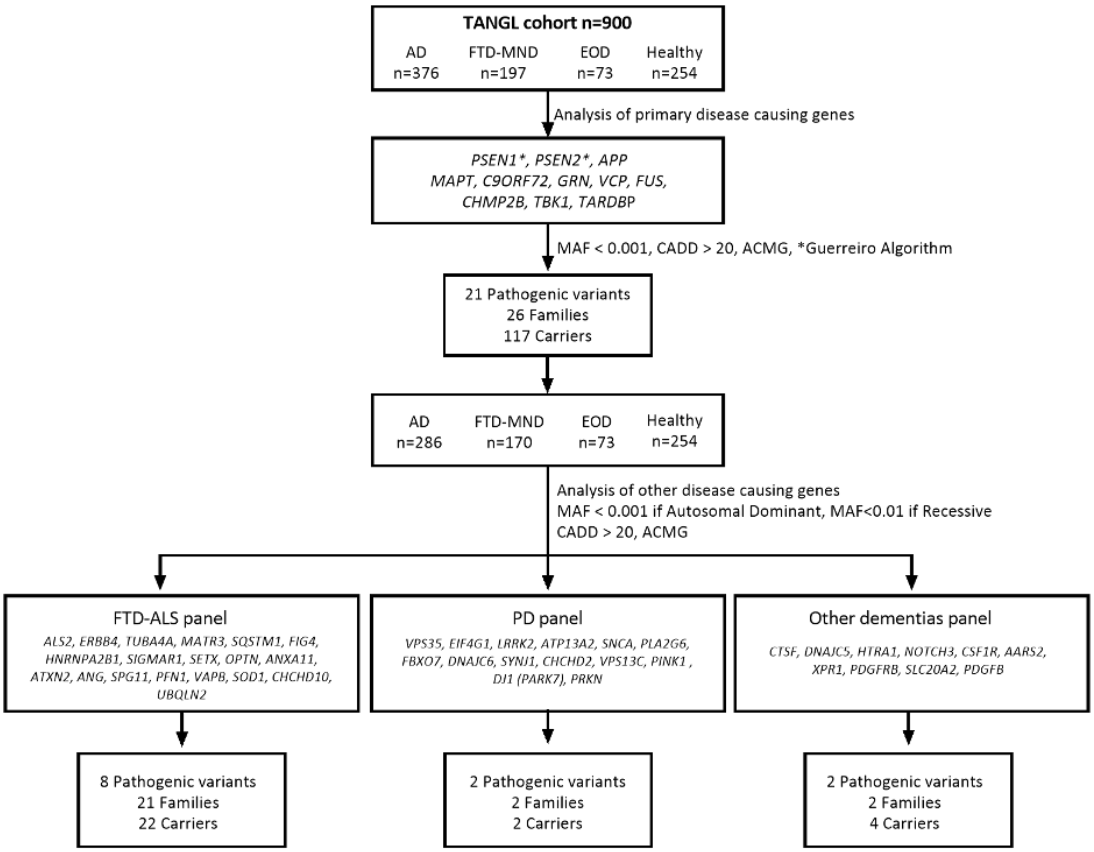


Figure 2.2: Genes associated with AD and FTD were selected from the AD/FTD mutation databases (https://www.molgen.vib-ua.be/ADMutations and https://www.alzforum.org/mutations). A secondary genetic analysis was done to identify possibly pathogenic variants in other genes associated with similar or overlapping phenotypes. For the secondary screening, the disease-causing genes included were those reported in the following OMIM phenotype series and phenotypes: Frontotemporal dementia and/or Amyotrophic Lateral Sclerosis (PS105550, PS167320, PS105400), Parkinson disease (PS168600), Adult-Onset Leukoencephalopathies (PS125310, 221820) and, Ceroid lipofuscinoses (PS256730).

around the locus. Because this software detects IBD of 2cM and higher, we additionally performed an alignment of the haplotypes carrying the variants of interest to search for smaller IBD segments between the TANGL and 1000 Genomes Project (1000GP) carriers. Autozygosity (homozygosity by descent) was determined using the same methods.

2.3 Results

2.3.1 Ancestral origins of the TANGL cohort

Nine hundred Colombian individuals with high-quality genome sequences were included in "The Admixture and Neurodegeneration Genomic Landscape" (TANGL) study. The individuals were divided into four different cohorts: Alzheimer's disease (AD), Frontotemporal lobar degeneration and motor neuron disease (FTLD-MND), early-onset dementia not otherwise specified (EOD), and healthy participants [Table 2.1 and appendix 2.6]. These 900 individuals represented 566 independent families, which were classified into the same four cohorts according to the diagnosis of the proband.

Because the sample set was highly selected, we first sought to determine the genomic similarity between the TANGL cohort and other Colombian individuals. We initially merged the TANGL and the 1000 Genomes Project (1000GP) phase 3 [100] datasets and performed a principal component analysis (PCA). The TANGL cohort had a similar distribution in the first three principal components (PC) to the "Colombians from Medellín" (CLM) of the 1000GP, allowing us to conclude that both populations are genetically similar [Appendix 2.7]. To take a closer look into the ancestral origins of the TANGL cohort, we used the software ADMIXTURE to estimate the number of ancestral populations (K) from which the cohort arose. The lowest cross validation (cv) error was obtained when assuming the cohort was derived from three ancestral populations (k=3), which agrees

with the history of the tri-continental admixture after the Spanish conquest. To analyze the global and local ancestry of the TANGL cohort, we merged the TANGL genomes with the European and African populations from the 1000GP and Native American genomes from Mao et al[72] and repeated the ADMIXTURE analysis(2.3 A). In this joint dataset, K=3 accurately differentiated Native American, European and African cohorts, but the lowest CV error was obtained for K=6. Modeling for six ancestral populations allowed the detection of substructure within the African and European cohorts, and created an additional cluster described by Moreno-Estrada et al [71] as a "Latino-specific European component" [Figure 2.3 C]. Consistent with previous studies[101], the ancestral population with the highest proportion in our cohort was European (mean of 64%, SD = 15%), followed by Native American (mean of 27%, SD = 11%), and African being the least represented (mean of 9%, SD= 11%) [Figure 2.3B and Figure 2.8 B]. These individual admixture values (q-values) at K=3 correlated with the sum of local ancestries estimated by RFMix (Pearson's r >0.99), allowing us to conclude that the local ancestry inferred for each individual matches the percentages of global ancestry obtained by an orthogonal method) [Appendix 2.8]. However, the regional differences in the fine structure of the Colombian population make these global ancestry proportions highly region dependent. For example, the three individuals whose global ancestry was nearly 90% African were from the Pacific coast of the country where former enslaved Africans settled and most of the population self identifies as Afro-Colombian [Appendix 2.6].

After calculating the proportions of global ancestry, we evaluated the TANGL cohort for sex biased admixture, a genetic trait previously described in the Colombian population[102, 103]. We used HaploGrep2 and yHaplo to determine mitochondrial and Y-chromosome haplogroups. The mitochondrial haplogroups of the probands (n=566) were predominantly Native American (83.4%) while the Y-chromosome haplogroups

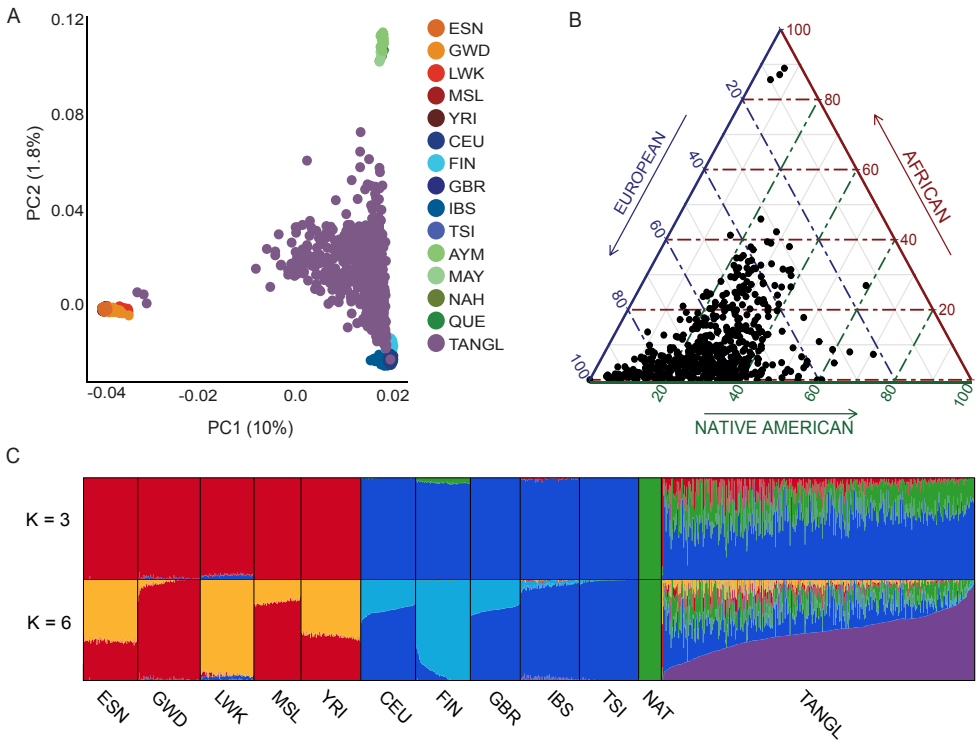


Figure 2.3: Population structure and admixture analyses of the TANGL cohort.(A) PC1 vs PC2 of the Principal component analysis of the TANGL cohort (purple) with the European (blue) and African (orange) individuals from the 1000GP and 43 Native American genomes (green). (B) Ternary plot representing the global ancestry of each of the individuals in the TANGL cohort Values according to sum of local ancestries calculated by RFMix. (C) Q plot of ADMIXTURE results assuming 3 and 6 ancestral populations (K). ESN: Esan in Nigeria. GWD: Gambian in Western Divisions in the Gambia. LWK: Luhya in Webuye, Kenya. MSL: Mende in Sierra Leone. YRI: Yoruba in Ibadan, Nigeria. CEU: Utah Residents (CEPH) with Northern and Western European Ancestry. FIN: Finnish in Finland. GBR: British in England and Scotland. IBS: Iberian Population in Spain. TSI: Toscani in Italy. AYM: Aymara. MAY: Mayan, NAH: Nahuatl. QUE: Quechua. NAT: Native American

(n=224) were mostly of European and of Mediterranean origins (92.8%), thus supporting the conclusion than multiple cohorts of Colombian origin show sex-biased admixture with Native American maternal lineages and paternal lineages from Europe [Appendices

2.6 and 2.7]. Overall, these analyses let us conclude that despite recruiting the TANGL cohort based upon neurodegenerative conditions from the Andes region of Colombia, it recapitulated the admixture patterns previously described in the country.

The TANGL cohort was distributed between the three ancestral populations in the PCA, clustering closer to Europeans and Native Americans (Figure 2.4). To determine if the clustering of the admixed individuals was driven by their percentages of global ancestry, we compared the values of the principal components (PC) with the percentage of global ancestry attributed to each of the three ancestral populations by ADMIXTURE. PC1 correlated with the percentage of African ancestry (Pearson's r2 of 1), and PC2 showed a correlation with the level of Native American ancestry (Pearson's r2 of 0.87). To determine whether the Colombian population clustered according to their global ancestry without including the ancestral populations in the analyses, we retained the 566 unrelated probands from the TANGL cohort and performed two PCAs, one with common variants (MAF >10%) and one with less frequent variants (MAF 5-10%). Both PCAs showed correlation of the PCs with the global admixture proportions, regardless of the inclusion of the ancestral population.

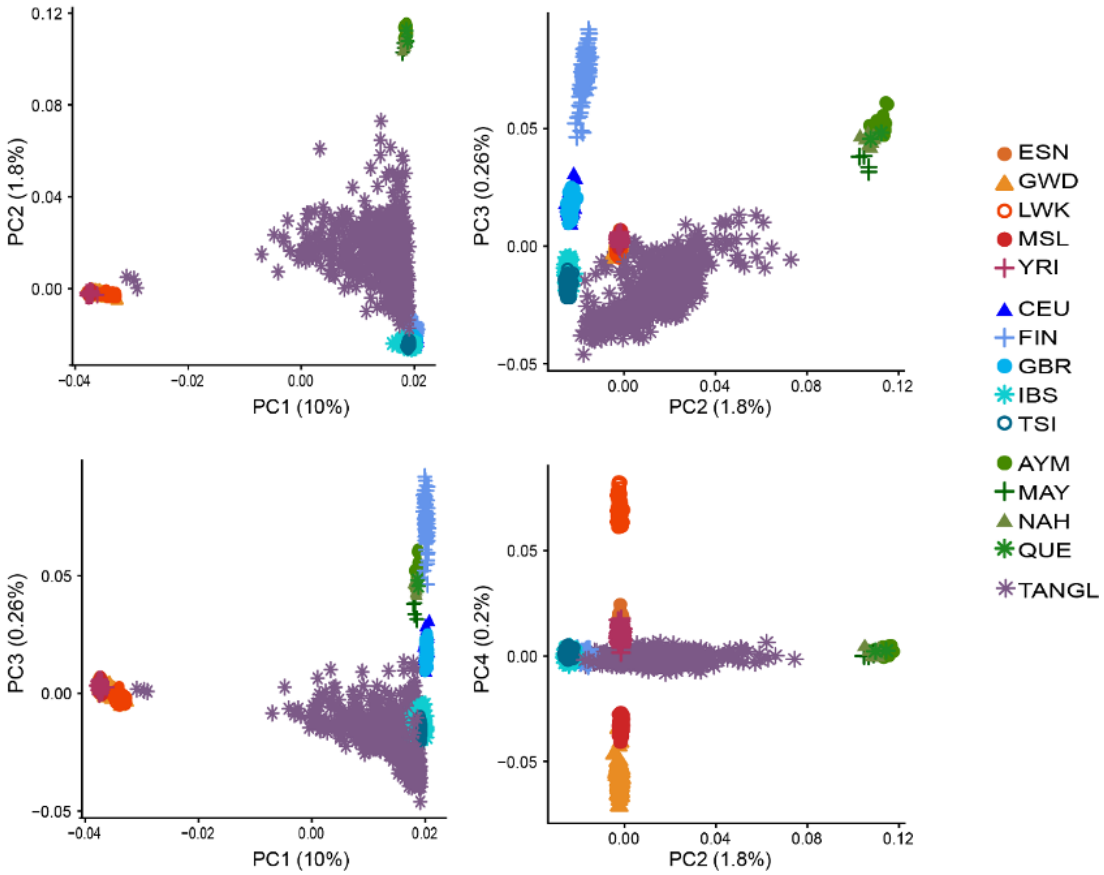


Figure 2.4: Principal component analyses of the African and European cohorts of the 1000GP, along with 43 Native American genomes and the TANGL cohort. African are displayed in orange, European in blue, native American in Green and TANGL in purple. ESN: Esan in Nigeria. GWD: Gambian in Western Divisions in the Gambia. LWK: Luhya in Webuye, Kenya. MSL: Mende in Sierra Leone. YRI: Yoruba in Ibadan, Nigeria. CEU: Utah Residents (CEPH) with Northern and Western European Ancestry. FIN: Finnish in Finland. GBR: British in England and Scotland. IBS: Iberian Population in Spain. TSI: Toscani in Italy. AYM: Aymara. MAY: Mayan, NAH: Náhuatl. QUE: Quechua

2.3.2 Disease causing variants

AD-associated genes

The 900 genomes were initially examined for pathogenic variants in AD-associated genes (*PSEN1*, *PSEN2* and *APP*) and the protein altering variants were curated according to the ACMG guidelines for the interpretation of genetic variants[82] and the algorithm proposed by Guerreiro et al [83] [Figure 2.2]. Eleven pathogenic variants were identified in the *PSEN1* gene [Table 2.2]. Three of these were novel and classified as definite pathogenic c.485T>G (p.Ile162Ser), c.667C>A (p.Gln223Lys) and c.782C>T (p.Val261Ala), four had been previously identified in the Colombian population c.349C>G (p.Pro117Ala), c.428T>C (p.Ile143Thr), c.839A>C (p.Glu280Ala) and c.1247T>C (p.Ile416Thr)[25, 104, 105, 106], and four had been described in families outside Colombia with diverse ancestries c.356C>T (p.Thr119Ile)[107], c.488A>G (p.His163Arg)[108], c.791C>T (p.Pro264Leu)[108] and c.851C>T (p.Pro284Leu) ("Mutations Database, Alzforum.,"). *PSEN1* c.839A>C (p.Glu280Ala)[25], of European origin, is the largest family in the world with familial Alzheimer's disease and living nearby is a family with the *PSEN1* variant c.1247T>C (p.Ile416Thr) [104] that originated in Africa. c.782T>C (p.Val261Ala) was identified in a singlet without confirmed paternity and it was classified as pathogenic despite the lack of family history due to the report of three different pathogenic mutations in the same codon c.780G>T (p.Val261Phe)[109], c.780G>A (p.Val261Ile)[110], c.780G>C (p.Val261Leu)[111]. All the reported variants, with the exception of c.851C>T (p.Pro284Leu), presented as early onset amnestic AD. The c.851C>T (p.Pro284Leu) carriers developed spastic paraparesis (SP), which is an atypical form of AD occasionally associated with certain *PSEN1* mutations[109, 112, 113]. All the families with pathogenic *PSEN1* mutations had autosomal dominant inheritance; however, the singlet c.782T>C (p.Val261Ala) was indeterminate. Among these *PSEN1* variants,

six were of European origin, three were Native Americans and one African [Table 2.2].

All the carriers of each variant, except c.791C>T (p.Pro264Leu), reported a known common ancestor. Several families from the harbored the *PSEN1* c.791C>T (p.Pro264Leu) variant but we could not connect them by family history. Therefore, to prove that c.791C>T (p.Pro264Leu) was the result of a founder effect, we used the hap-IBD software to identify identical by descent (IBD) segments between the variant carrying chromosomes. All the *PSEN1* c.791C>T (p.Pro264Leu) carrier haplotypes shared an IBD segment of 2.79 cM around the *PSEN1* locus, supporting the hypothesis of a common ancestor for all three families originating at about the same time. *PSEN1* c.791C>T (p.Pro264Leu) has been described in multiple populations, and while the European carriers of this variant often present SP[114], this phenotype was not observed in the Colombian carriers of the variant. To determine if this phenotypic heterogeneity is related to the ancestral haplotype wherein the variant arose, we used RFMix to estimate the ancestry of the variant carrier haplotype. In the TANGL cohort, *PSEN1* c.791C>T (p.Pro264Leu) resided on a Native American haplotype, which suggests that the haplotype of origin may play a role in the different expressivity and clinical manifestations between the variant carriers. Six of the other pathogenic *PSEN1* variants resided on European haplotypes, two variants were present in Native American and one in an African background. The multi-ancestral origins of the *PSEN1* variants suggest that the admixture process contributed to the introduction of pathogenic variants to a population.

Two of the *PSEN1* variants described in this cohort had been previously identified in other families in Colombia [c.428T>C (p.Ile143Thr)(Arango2001), c.356C>T (p.Thr119Ile)] and in Argentina [c.356C>T (p.Thr119Ile)[107]]. We performed additional array genotyping to test for IBD between the members of these families and those from the TANGL cohort. The Colombian carriers of c.428T>C (p.Ile143Thr) and c.356C>T

(p.Thr119Ile) showed IBD overlapping the *PSEN1* locus. Interestingly, the Colombian individuals who harbored c.356C>T (p.Thr119Ile) with whom no shared ancestor could be determined by history carried a small IBD segment shared with the Argentinian carrier of the same variant. The geographical expanse over which these variants reside could reveal small population migratory streams from Europe or within South America.

Most of the variants observed in *PSEN2* were either benign or had been previously classified as risk factors for AD. Only the variant c.487C>T (p.Arg163Cys), which had been described in a Chinese patient with AD(Gao2019), was classified as pathogenic. Interestingly, this variant resided on an African haplotype in the Colombian carrier. No pathogenic variants were observed in *APP*; but one individual with AD had CNV spanning *APP*[114] [Chromosome 21 g.(26253828_30011000)dup]. These results confirm *PSEN1* as the most prevalent gene associated with genetic AD in our cohort, mostly as the result of founder effects, and that the current genetic burden of the TANGL cohort is influenced by the genetic diversity of its founders.

FTLD-MND associated genes

We performed the same curation process for FTLD-MND associated genes (*MAPT*, *C9orf72*, *GRN*, *VCP*, *FUS*, *CHMP2B*, *TBK1*, *TARDBP*). Most of the individuals with genetic forms of FTLD-MND in the TANGL cohort had deleterious variants in *MAPT* and *TARDBP* [Table 2.2]. The *MAPT* c.1189C>T (p.Pro397Ser) variant was identified in three independent families from the same geographic region that shared IDB segment of 2.89 cM overlapping the locus. This variant had been previously reported in five apparently unrelated Spanish families[70], and like the Spanish counterpart, the Colombian *MAPT* c.1189C>T (p.Pro397Ser) carriers had variable expressivity of the illness. To elucidate whether the Colombian *MAPT* c.1189C>T (p.Pro397Ser) carriers were IBD with

the Spanish families, we used exome sequencing data from a Spanish patient to search for similarities in the variant carrying haplotype. We identified a minimal shared haplotype of 2.65cM including the *MAPT* locus, which suggests that the Colombian families share a common ancestor with the Spanish carriers of *MAPT* c.1189C>T (p.Pro397Ser).

Two siblings with FTLD-MND born of consanguineous parents were homozygous for the *TBK1* c.1717C>T (p.Arg573Cys) variant. Haploinsufficiency of *TBK1* has been previously associated with familial MND and FLTD and is a known mechanism of pathogenicity(Freischmidt2015). Homozygosity of nonsense *TBK1* variants has been proven to be lethal in mice(Bonnard 2000). A second variant in *TBK1* was c.1257_1258del (p.Val421Cfs*26), identified in two unrelated individuals that shared an IBD segment of 3.1 cM including the *TBK1* locus. We identified two variants in *TARDBP* that had been previously reported in European populations with diagnosis of ALS[115, 116], and in contrast with these cohorts, Colombian *TARDBP* c.1147A>G (p.Ile383Val) carriers had significant intra-familial variability with heterogeneous FTLD-MND spectrum disorders. Our study identified only one carrier of *C9orf72* expansion, a single carrier of a pathogenic variant in *GRN*, and no disease-causing variants in *CHMP2B*, *FUS* or *VCP*. While the frequency of the identified mutations differs from those reported in European descent cohorts[84, 117], all the identified pathogenic variants in these FTLD-MND associated genes resided on European haplotypes.

Genes associated with MND

To explore the phenotypic and genetic overlap between FTLD and MND, we searched for pathogenic variants in nineteen additional genes associated with MND, with or without FTLD. The *SQSTM1* [MIM: 601530] c.1175C>T (p.Pro392Leu) variant, was present in 11 unrelated cases and two controls of the TANGL cohort. These cases were unre-

lated and were clinically heterogeneous: six had diagnosis of AD, three of FTLD, one of CBD and one PSP [Table 2.2]. Eight of the eleven cases had family history of dementia or neurodegenerative disease, and none of them carried other pathogenic mutations in the explored disease-causing genes. The *SQSTM1* c.1175C>T (p.Pro392Leu) has been widely studied, mostly in European populations [118, 119] and it was present in five individuals from the European cohort of the 1000GP. We used HAP-IBD to search for IBD between the Colombian and the 1000GP carriers of *SQSTM1* c.1175C>T (p.Pro392Leu). Ten carriers of the TANGL cohort shared IBD segments >2cM overlapping the variant, which resided in a European haplotype as well. To determine IBD at a smaller scale, we did a manual alignment of all the variant-carrying haplotypes and detected an IBD segment of ∼1cM between all the TANGL cohort and 1000GP European *SQSTM1* c.1175C>T (p.Pro392Leu) carriers. This observation suggests that *SQSTM1* c.1175C>T (p.Pro392Leu) shows the signature of a founder effect that pre-dates the Spanish invasion. Variants with higher allelic frequency also show IBD between the TANGL cohort and with other carriers outside of Colombia.

In contrast to the pathogenic variants in the FTLD-MND associated genes, five of the eight disease associated variants identified in the MND panel were of Native American origin while only two were of European ancestry [Table 2.2]. However, most of these individuals with pathogenic variants in Native American haplotypes presented with FTLD phenotypes. For example, the *TUBA4A* [MIM: 191110] c.820C>G (p.Pro274Ala) variant was identified in two independent families with positive family histories of dementia and diagnosis of bvFTD and EOD without motor neuron disease. As described previously for other variants, these families shared a long IBD haplotype of 15.54cM overlapping the locus, suggesting a recent common ancestor. The *SOD1* [MIM: 147450] c.63C>G (p.Phe21Leu) variant was identified in one patient with sporadic navPPA who

did not have any motor or ALS-associated symptoms. This variant and others in this same amino acid [c.62T>G (p.Phe21Cys)] had been previously reported in patients with ALS[120, 121]. Additional pathogenic variants in *ANXA11* [MIM: 602572] and *HNRNPA2B1* [MIM: 600124] residing in Native American haplotypes were identified in patients with svPPA and bvFTD. These results further intertwine MND and FTLD with several genes previously associated exclusively with MND that may also be responsible for a FTLD phenotype in a different ancestral context. The genetic and clinical heterogeneity of MND associated genes had been previously described in European population[122], but the inclusion of diverse individuals expands the extent of genetic overlap between FTLD and ALS.

A patient with PSP was homozygous by descent for a European haplotype harboring the *FIG4* [MIM: 609390], c.122T>C (p.Ile41Thr). Although this gene has been associated with autosomal dominant forms of ALS, this same specific variant has been reported in compound heterozygosity with nonsense variants in European individuals with autosomal recessive cases of Charcot-Marie-Tooth's disease[123][MIM: 611228]. A family presenting with FTLD-ALS was shown to have a novel c.724G>A (p.Ala242Thr) variant in *UBQLN2* [MIM: 300264]. *UBQLN2*, found on the X-chromosome, is associated with MND or FTLD-MND, with a lower penetrance in females[124]. The family phenotype was of late onset bvFTD presentation in the female carrier, and FLTD-MND in the male carrier.

Genes associated with other neurodegenerative disorders

Several families with EOD were explained by variants in other non AD-FTD-MND genes [Figure 2.2]. A family with an unspecified autosomal dominant EOD had a novel mutation in *DNAJC5* [MIM: 611203] c.347T>G (p.Leu116Arg) residing on an African haplotype. Their phenotype and postmortem brain tissue histopathology was compati-

ble with adult-onset ceroid neuronal lipofuscinosis-4B (CNL4B) [MIM: 162350]. A novel pathogenic variant in *LRRK2* [MIM: 609007] c.4334C>G (p.Ser1445Cys) was identified in a patient with a European background and non-motor symptoms in Parkinson's disease and dementia. One patient with a family history of cancer and dementia carried the *CSF1R* [MIM: 164770] c.2068G>A (p.Gly690Ser) variant in a Native American haplotype. *CSF1R* variants have been associated with Hereditary Diffuse Leukoencephalopathy with Spheroids (HDLS)[125] [MIM: 221820]. A postmortem brain tissue examination supported HDLS diagnosis for the *CSF1R* c.2068G>A (p.Gly690Ser) variant carrier. These families provide novel insights on genetic-phenotypic relationships.

Despite an extensive evaluation of known genes previously reported for Mendelian forms of dementia, we were not able to identify a disease-causing variant in all families with autosomal dominant inheritance of the illness. Of the 566 families included in the present study, 59 had autosomal dominant inheritance defined as three or more affected individuals in two consecutive generations. For the 18 families in which all individuals had early onset of symptoms (<65 years), we could identify disease causing variants in all but three, and 13 of them carried pathogenic *PSEN1* variants. In families with both early and late onset cases, we identified disease causing variants in seven of 33. No disease-causing variant was identified in the 12 individuals from the eight families where everyone had late onset, but 10 of them carried at least one *APOE* [MIM: 107741] E4 allele. (Two were *APOE* E3/E3, six were E3/E4 and four E4/E4). In conclusion, a pathogenic variant was identifiable in the families with autosomal dominant inheritance in which most of the affected individuals had disease onset before 65 years.

Gene	Coding Change	CADD	Local Ancestry
Alzheimer's Disease Genes			
APP	g.(26253828_30011000)dup	-	European
PSEN1	c.349C>G (p.Pro117Ala)	26.9	European
	c.356C>T (p.Thr119Ile)	24.4	European
	c.428T>C (p.Ile143Thr)	26.8	European
	c.485T>G (p.Ile162Ser)	32	Native American
	c.488A>G (p.His163Arg)	23.4	European
	c.667C>A (p.Gln223Lys)	33	Native American
	c.782T>C (p.Val261Ala)	25.9	Undetermined
	c.791C>T (p.Pro264Leu)	35	Native American
	c.839A>C (p.Glu280Ala)	29.3	European
	c.851C>T (p.Pro284Leu)	33	European
	c.1247T>C (p.Ile416Thr)	25.9	African
PSEN2	c.487C>T (p.Arg163Cys)	35	African
FTLD Genes			
C9ORF72	(GGGGCC)n Repeat Expansion		European
GRN	c.709-2A>G (p.Ala237fs)	23.1	European
MAPT	c.902C>T (p.Pro301Leu)	34	European
	c.1189C>T (p.Pro397Ser)	25	European
TARDBP	c.881G>T (p.Gly294Val)	18.89	European
	c.1147A>G (p.Ile383Val)	0.308	European
TBK1	c.1257_1258del (p.Val421Cfs*26)	-	European
	c.1717C>T (p.Arg573Cys)	29.6	European
ALS Genes			
ANXA11	c.904C>T (p.Arg302Cys)	31	Native American
FIG4	c.122T>C (p.Ile41Thr)	26.5	European
HNRNPA2B1	c.965G>A (p.Gly322Glu)	23.6	Native American
SOD1	c.63C>G (p.Phe21Leu)	22.9	Native American
SQSTM1	c.1175C>T(p.Pro392Leu)	34	European
TUBB4A	c.811G>A (p.Ala271Thr)	22.8	Native American
UBQLN2	c.724G>A (p.Ala242Thr)	25.9	Undetermined
Other Neurodegeneration Associated Genes			
CSF1R	c.2068G>A (p.Gly690Ser)	23.1	Native American
DNAJC5	c.347T>G (p.Leu116Arg)	27.2	African
LRRK2	c.4334C>G (p.Ser1445Cys)	24.3	European
	c.6055G>A (p.Gly2019Ser)	35	European

Table 2.2: Disease-causing variants in the TANGL cohort. CADD corresponds to the Phred score

2.3.3 Genetic variation associated with AD risk genes

Both rare and common variants can have a small effect size on Alzheimer's disease risk[126]. To explore rare variants conferring intermediate risk for the illness, we selected three genes (*TREM2*, *SORL1* and *ABCA7*) that have shown odds ratio higher than two (OR >2) in disease association studies[90]. Using the criteria suggested by [90]), we identified 14 protein truncating variants (PTV) and 16 strictly damaging (SD) variants in *TREM2*, *SORL1* and *ABCA7* [Table 2.3].

The most common risk-conferring variants in the TANGL cohort resided on *TREM2*, with over a hundred individuals carrying SD or PVT in this gene. The most prevalent variant was c.469C>T (p.His157Tyr), with 50 heterozygous and seven homozygous carriers. All the c.469C>T (p.His157Tyr) carriers were IBD for a Native American haplotype. Two out of three algorithms classified His157Tyr as definitely pathogenic, while a metanalysis determined *TREM2* c.469C>T (p.His157Tyr) has an OR= 3.65[127] and therefore, it qualified for the present study. Additionally, we identified 33 *TREM2* c.140G>A (p.Arg47His) carriers in our cohort; three of them were homozygous for this variant. All the *TREM2* c.140G>A (p.Arg47His) carriers from the TANGL cohort shared an IBD European haplotype overlapping the *TREM2* locus, and this same variant-carrying haplotype was present in five European individuals from the 1000GP who showed IBD with the Colombian carriers. Besides risk conferring variants in Native American and European haplotypes, an African *TREM2* haplotype [GenBank: NM_001271821] carrying c.572G>A (p.Trp191), c.632T>C (p.Leu211Pro) and c.287C>A (p.Thr96Lys) was identified in 10 individuals. This haplotype was previously associated with an increased risk in African-American cohorts[128]. Unlike the previous cases of homozygosity, one individual with early onset AD was a compound heterozygote with both the Thr96Lys/Trp191*/Leu211Pro haplotype and the c.469C>T (p.His157Tyr) variant, sug-

gesting that genetic risk factors from different ancestral origins may coexist in admixed individuals and populations.

Rare variants in *TREM2* are population specific. For example, *TREM2* c.140G>A (p.Arg47His) is associated with increased risk for AD in European descent populations[129, 130] but not in African[128] or Asian[131, 132], while *TREM2* c.469C>T (p.His157Tyr) shows association with AD in Asian[127, 133] but not in European[134] or African[128]) cohorts. Interestingly, the c.469C>T (p.His157Tyr) variant was found in Colombia on a Native American haplotype, raising the possibility that this allele arrived from Asia to the American continent close to the time when the Americas were first populated 15,000–20,000 years ago. To support this hypothesis, we searched for this variant in the Human Genome Dating database [135], which uses coalescent modeling to estimate the time to the most recent common ancestor (TMRCA) between the variant carriers and the age of the variant. The estimated age of the c.469C>T (p.His157Tyr) allele is 1,265 generations (95% confidence interval of 1,108.5–1,430.9), which corresponds to 31,625 years by setting one generation equals 25 years (https://human.genome.dating/snp/rs2234255.jnt). In contrast, c.140G>A (p.Arg47His) emerged more recently, as it was estimated to be 425 generations old or 10,625 years (https://human.genome.dating/snp/rs75932628), dating to a time before gene flow from Europe to the Americas occurred. These results lead us to conclude that the disease burden in this population is not only affected by the recent admixture after the conquest of the Americas, but was also affected by migrations(Benton2021) during the original populating of the continent.

Risk-conferring variants in *ABCA7* and *SORL1* were less prevalent than those in *TREM2*. Most of the variants detected in *ABCA7* consisted in PTV and resided on African haplotypes. The majority in *SORL1* were SD variants of European origin, two homozygous carriers of *ABCA7* variants c.2124_2130del (p.Glu709fs) and c.4886C>T

(p.Ser1629Leu), and a compound heterozygote of risk variants from different ancestral origins. There were no compound heterozygous or homozygous variants for *SORL1*, and the c.6550G>A (p.Ala2184Thr) variant was only found in a healthy centenarian. Additionally a search for risk associated variants in *ADAM10* [97, 98], identified c.510G>C (p.Gln170His) in ten individuals, including one homozygous patient. These reported variants in *TREM2*, *SORL1*, *ABCA7* and *ADAM10* were IBD in carriers of the same variant. In summary, the characteristics we described for disease-causing variants such as IBD between carriers, multiple ancestral origins of deleterious variants within the same gene, and autozygosity were present in variants with higher allelic frequencies in risk-associated genes.

The high allelic frequency of some risk conferring variants in the TANGL cohort allowed the detection of individuals who were homozygous by descent and raised the hypothesis of consanguinity between their parents, as was the case for the two families with recessive dementias [*TBK1* c.1717C>T (p.Arg573Cys) and FIG4 c.122T>C (p.Ile41Thr)]. We used Hap-IBD and manual haplotype alignment to estimate the autozygosity of the homozygous individual for risk-associated variants in *ABCA7* [c.2124_2130del (p.Glu709fs) and c.4886C>T (p.Ser1629Leu)], *TREM2* [c.140G>A (p.Arg47His) and c.469C>T (p.His157Tyr)] and *ADAM10* [c.510G>C (p.Gln170His)]. Five individuals from three families who were the offspring of related parent had autozygous segments >30cM overlapping the risk associated variant. The remaining individuals had smaller autozygous segments, suggesting background relatedness of the population due to a small effective population size or bottlenecks [136, 137].

Gene	Coding Change	Classification	CADD	Local Ancestry
ABCA7	c.2T>C	PTV	24.9	Native American
	c.236A>C (p.Asn79Thr)	SD	24.5	African
	c.1180_1190del (p.Leu396fs)	PTV	.	African
	c.1531G>T (p.Glu511*)	PTV	39	African
	c.1776G>T (p.Trp592Cys)	SD	26	African
	c.2124_2130del (p.Glu709fs)+	PTV	.	European
	c.2194C>T (p.Gln732*)	PTV	36	European
	c.2552+11_2552+58del	PTV	.	African
	c.2611G>C (p.Asp871His)	SD	24.8	African
	c.3781delC (p.Pro1261fs)	PTV	.	Native American
	c.4208delT (p.Leu1403fs)	PTV	.	European
	c.4465C>T (p.Arg1489*)	PTV	39	European
	c.4886C>T (p.Ser1629Leu)+	SD	35	Native American
	c.4895C>T (p.Pro1632Leu)	SD	34	African
	c.5302delC (p.Leu1768fs)	PTV	.	Native American
	c.5463+2T>C	PTV	23.7	European
	c.5794C>T (p.Arg1932C)	SD	34	African
SORL1	c.994C>T (p.Arg332Trp)	SD	35	European
	c.1432G>C (p.Ala478Pro)	SD	28.2	European
	c.1496C>T (p.Ser499Leu)	SD	35	European
	c.2200G>A (p.Asp734Asn)	SD	34	European
	c.2230C>T (p.Arg744*)	PTV	39	European
	c.2710C>T (p.Arg904Trp)	SD	33	Native American
	c.3679G>T (p.Gly1227Cys)	SD	34	European
	c.4520C>T (p.Pro1507Leu)	SD	26.2	Undetermined
	c.6550G>A (p.Ala2184Thr)	SD	34	African
TREM2	c.140G>A (p.Arg47His)+	SD	33	European
	c.469C>T (p.His157Tyr)+	SD	23.1	Native American
	c.572G>A(p.Trp191*)	PTV	-	African
	c.594G>A (p.Trp198*)	PTV	39	Undetermined
ADAM10	c.510G>C (p.Gln170His)+	SD	19.17	European

Table 2.3: Risk-Conferring variants in the TANGL cohort. CADD corresponds to the CADD Phred score, SD: Strictly Damaging, PTV: Protein truncation variant. Variants with + were observed in hetero and homozygous states

2.4 Discussion

Genetic drift has been one of the main forces shaping human genomic variation[138, 139] While populations that emerge from a bottleneck will harbor reduced genetic variation, over time such a population can accumulate higher numbers of deleterious variants due to random fluctuations in allele frequencies[44]. Furthermore, deleterious allele frequencies decrease more slowly in smaller populations because natural selection acts on fitness differences and therefore requires genetic variation[44]. The Colombian tricontinental admixture among the Native Americans, Europeans and Africans combined a portion of the genetic disease burden that was previously limited to each of these ancestral populations. Within the backdrop of an admixed population, numerous infectious diseases extracted a very steep mortality. As a consequence, the small isolated settlements that survived the bottleneck rapidly expanded locally during the colonial period[43]. These multiple isolated bottlenecks each with their own rare variants added to the diversity over the entire population. The TANGL cohort recapitulated the admixture patterns previously described in the Colombian population, suggesting that the country's demographic history is likely to underlie the modern clustering of familial neurodegenerative diseases arising from multi-ancestral rare disease-associated alleles.

In this cohort, most familial early onset AD cases were caused by variation in the *PSEN1* gene. We identified eleven different pathogenic *PSEN1* variants from multiple ancestral origins, nearly all attributed to founder effects. The *PSEN1* mutations emerged from a small effective population in each of the early settlements that constituted a patchwork of bottlenecks dispersed throughout the country. Because people tended to remain geographically isolated, the rare variants represent a local genetic footprint. Survivors who emerged from the bottleneck had escaped the large number of infectious diseases responsible for decimating the population. During the historical period of colonization,

populations in these settlements grew rapidly as the incidence of diseases diminished, which favored the segregation of potentially damaging variants at higher rates. The question arises as to whether the *PSEN1* mutations could be under positive selection or are the mutations completely explained by drift. Because *PSEN1* mutant phenotypes do not appear until after the age of child-bearing, it is unnecessary to invoke trade-off effects for maintaining the mutation in the population. Positive selection for Alzheimer risk in the context of infectious burden has been previously attributed to the *APOE* E4 risk allele[140]. *PSEN1* mutations cause the production of excess amyloid-β which may function as an anti-microbial peptide (AMP)[141]. In this manner, *PSEN1* mutations may have been positively selected as protection against the enormous mortality of infectious diseases. AMPs function as an ancient component of the innate immune system that target bacteria, mycobacteria, enveloped viruses, fungi, and protozoans[142]. Amyloid-β is active against at least eight common and clinically relevant microorganisms, and several anti-amyloid-β clinical trials have reported increased rate of infections among the participants [141]. However, given the short $\sim$500-year interval since the selective pressure occurred and the $\sim$100-year pulse-like nature of the selection, the possibility of positive selection must remain speculative. Without a sufficient time interval for the mutation to spread widely through the population, the only indirect support for positive selection might consider the collective fitness conferred by all of the *PSEN1* mutations due to their shared phenotypic effect of increasing amyloid-β as an AMP. Whether these mutations represent a statistical excess will require further study, but given the population size at the time to which the mutations can be historically traced, (see ancestry data for each mutation), it is likely that the mutations derived from a small effective population, thus supporting their possible over-representation. A comparison comes from large catchment groups for clinics with an interest in familial dementias—one in Alabama had no *PSEN1* cases in their series[143] and another in San Francisco had six *PSEN1* cases (personal

communication, Jennifer Yokoyama). In one study that sought Early-Onset Alzheimer Patients from 28 university hospitals across France spanning the dates 1993 to 2016, 17 carried a *PSEN1* mutation. However, any comparison with our cases is problematic because ten of these arose de novo, which was not the case in the TANGL cohort, and some were of unknown pathogenicity.

In addition to the *PSEN1* variants, we identified multiple rare variants causing autosomal dominant early onset dementia. Variants were usually found in one locality and likely derived from a common ancestor [Figure 2.5]. Previous studies had reported disease causing variants for other neurological disorders with the signature of founder effects; among these are four different CADASIL [MIM: 125310] associated variants in *NOTCH3* [MIM: 600276, c.307C>T(p.Arg103Cys), c.421C>T (p.Arg141Cys), c.484T>A (p.Cys162Ser), c.1363T>C (p.Cys455Arg)][144, 145], a familial episodic pain syndrome [MIM: 615040] with a variant in *TRPA1* [MIM: 604775, c.2564A>G (p.Asn855Ser)][146], Huntington's disease[147][MIM: 143100], a Parkinson disease variant in *LRRK2* [c.6055G>A (p.Gly2019Ser)] [148], blepharophimosis–ptosis–epicanthus inversus syndrome (BPES) [MIM: 110100] type 1 with a *FOXL2* [MIM: 605597, c.157C>T (p.Gln53 $*$)] variant and BPES type 2 with *FOXL2* in-frame 30 bp duplication (c. 909–938dup)[149], a complex ataxia due to a *KIF1A* variant [MIM: 601255, variant c.304G>C (p.Gly102Arg)], generalized epilepsy with febrile seizures plus (GEFS) [MIM: 604403] with *SCN1A* [MIM: 182389 c.5225A>G (p.Asp1742Gly)] variant[150], and non–syndromic hearing loss [MIM: 220290] due to a *GJB2* variant [MIM: 121011 c.35delG (p.Gly12Valfs$*$][151] . Founder effects can also be detected in other non–neurologic conditions: *BRCA1/2* variants [MIM: 113705, 600185] among Colombian women with breast and ovary cancer increased the prevalence of these variants in the studied population [152]. Most of these mutations map to small distinct locales that when, taken together, demonstrate the remarkable overlap

of the genetic and geographic maps.

This study underscores the numerous genetic insights that can emerge from Latin American populations. Another example is the putative modifier gene—homozygosity of the Christchurch variant in ApoE3—that may strongly delay the onset of Alzheimer's disease [32]. This gene variant and many of the rare large effect size mutations reported here arose due to the unique genetic history of the region. Ongoing interest in Latin American genetic studies, akin to all genetic studies in under-represented populations, must consider the ethical implications of the research. Over the many years these were obtained, the research was conducted with the full involvement of the community and extensive interactions with and informed consent from the contributing families.

2.5 Permissions and Attributions

The content of Chapter 2 and is the result of a collaboration with Dr. David Aguillón at the University of Antioquia.

Acosta-Uribe J∗, Aguillón D∗, Cochran JN, Giraldo M, Madrigal L, Killingsworth BW, Singhal R, Labib S, Alzate D, Velilla L, Moreno S, García GP, Saldarriaga A, Piedrahita F, Hincapie L, López HE, Perumal N, Morelo L, Vallejo D, Solano JM, Reiman EM, Surace EI, Itzcovich T, Allegri R, Sánchez-Valle R, Matallana D, Myers RM, Browning SR, Lopera F, Kosik KS. A Neurodegenerative Disease Landscape of Rare Mutations in Colombia Due to Founder Effects. [Manuscript under review as of August 2021]
∗Co-first authors

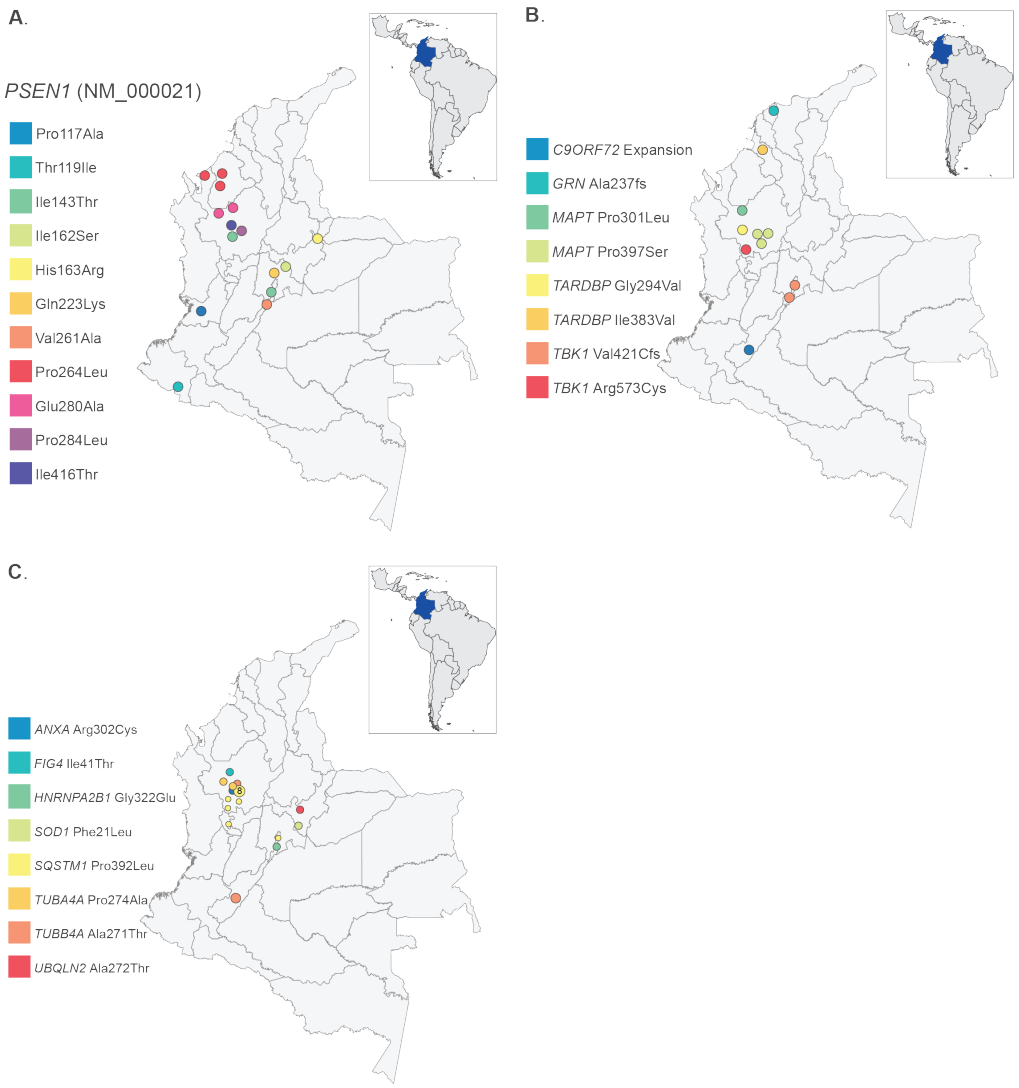


Figure 2.5: Place of origin of the families with disease causing variants. Numbers inside circles represent numbers of families carrying the variant who are original from the same geographic region.

2.6 Supplementary material

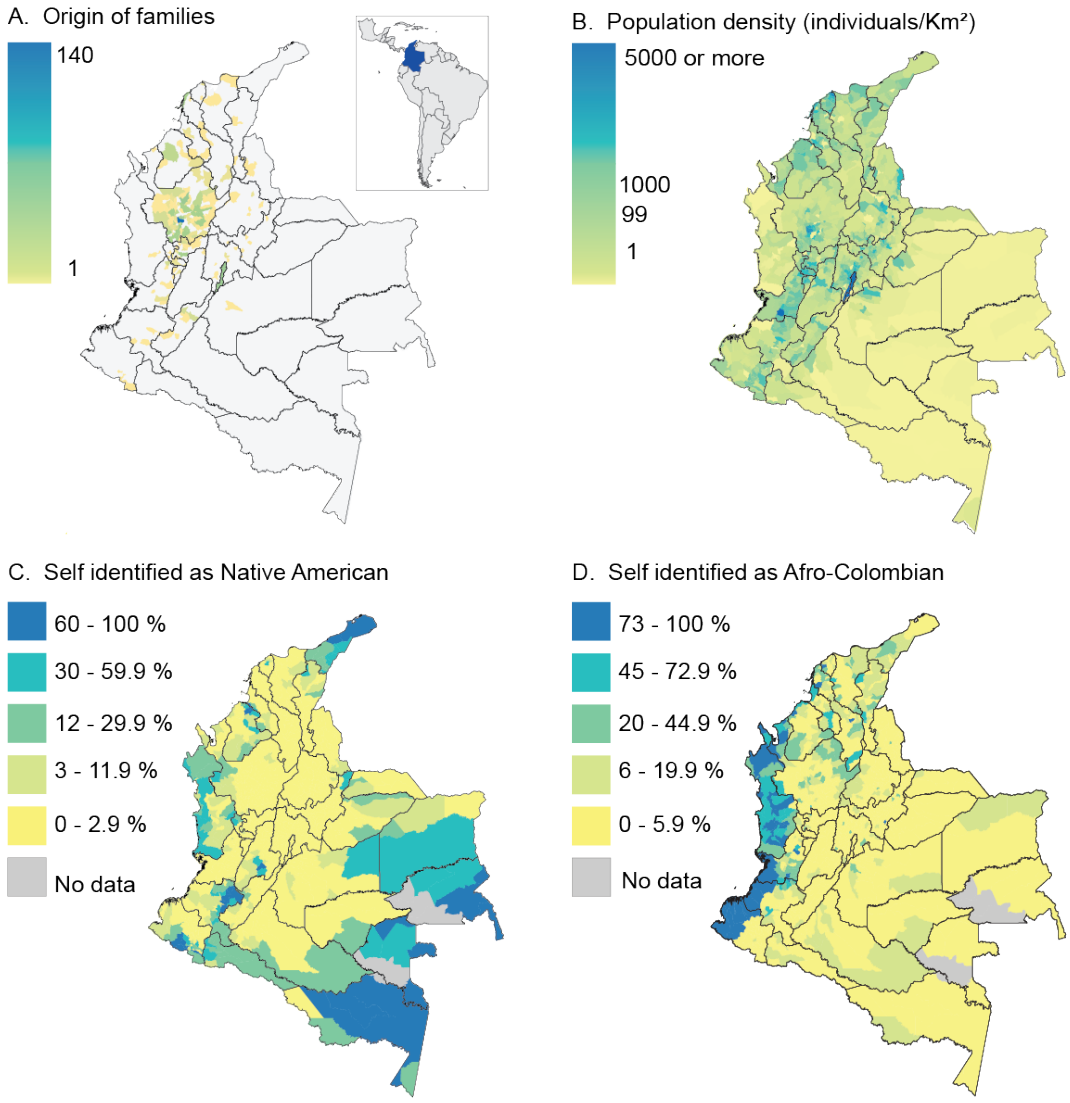


Figure 2.6: Demographic information of the TANGL cohort and the Colombian population: (A) Place of origin of the included families in the map of Colombia. (B) Population density in Colombia according to the 2018 census (C and D) Colombian population that self identifies as Native American or Afro-Colombian according to the 2005 census. Census data from 2018 and 2005 is publicly available by the "Departamento Administrativo Nacional de Estadística DANE": www.dane.gov.co

Cohort	Diagnosis	n	AAO		Female	
			mean	range	n	%
AD n=376	Alzheimer's Disease Dementia	343	58.9	30 - 90	227	66.2
	Logopenic variant PPA	3	60.3	54 - 69	2	66.7
	Mild cognitive impairment	3	64.6	61 - 67	3	100
	Asymptomatic carriers of pathogenic PSEN1 variants	27	30.5**	14-60**	17	63
FTLD MND n= 197	Amyotrophic lateral sclerosis	30	52.1	21 - 71	14	46.7
	Corticobasal Degeneration	13	58.8	47 - 73	7	53.8
	Behavioral variant of FTD	78	59	35 - 78	39	50
	Semantic variant of PPA	24	60	45 - 82	10	41.7
	non-fluent/agrammatic (PPA	26	60.7	48 - 74	15	57.7
	Right Variant FTD	3	65.3	63 - 67	1	33.3
	FTD with MND	6	66.3	59 - 71	1	16.7
	Progressive supranuclear palsy	6	59.8	45 - 78	3	50
	Asymptomatic carriers of MAPT and TARDBP variants	11	62**	43-77**	2	18.2
EOD n= 73	Bipolar Disorder without cognitive impairment	7	51.2**	34-58**	4	57.1
	Bipolar Disorder without cognitive impairment	13	61.3	51 - 75	11	84.6
	Neuronal Ceroid Lipofuscinosis 4B (CLN4B)	3	32.3	30 - 34	3	100
	EOD	40	55.1	25 - 66	28	70
	Hereditary leucoencephalopathy	1	48	-	1	100
	Lewy Body Disease	3	59.3	55 - 67	1	33.3
	Parkinson's Disease	2	57.5	55 - 60	-	-
	Primary lateral sclerosis + optic atrophy + dementia	4	54**	47-61**	1	25
Healthy Individuals n = 254	Healthy Related to a patient	168	54.8**	14-85**	111	66
	Healthy Unrelated	86	80.5**	51-101**	48	55.8
		900	-	-	549	61

Table 2.4: Demographic information of TANGL sub-cohorts: AD: Alzheimer's disease; FTLD: frontotemporal lobar degeneration; MTD: Motor neuron disorder; EOD: Early onset dementia not otherwise specified; PPA: Primary progessive aphasia; FTD: Frontotemporal dementia; AAO: Age at onset; ** Age at evaluation

Gene Name	MIM ID	GenBank Accesion number
ABCA7	605414	NM_019112
ADAM10	602192	NM_001110
ANXA11	602572	NM_001157
APOE	107741	NM_000041
APP	104760	NM_000484
C9ORF72	614260	NM_001256054
CHMP2B	609512	NM_014043
CSF1R	164770	NM_005211
DNAJC5	611203	NM_025219
FIG4	609390	NM_014845
FOXL2	605597	NM_023067
FUS	137070	NM_001170634
GJB2	121011	NM_004004
GRN	138945	NM_002087
HNRNPA2B1	600124	NM_002137
KIF1A	601255	NM_004321
LRRK2	609007	NM_198578
MAPT	157140	NM_005910
NOTCH3	600276	NM_000435
PSEN1	104311	NM_000021
PSEN2	600759	NM_000447
SCN1A	182389	NM_001165963
SOD1	147450	NM_000454
SORL1	602005	NM_003105
SQSTM1	601530	NM_003900
TARDBP	605078	NM_007375
TBK1	604834	NM_013254
TREM2	605086	NM_018965
TRPA1	604775	NM_007332
TUBA4A	191110	NM_006000
TUBB4A	602662	NM_006087
UBQLN2	300264	NM_013444
VCP	601023	NM_007126

Table 2.5: GenBank accession numbers for the genes reported in chapter 2. MIM gene accesion number was taken from https://www.omim.org, GeneBank accession was obtained from https://www.ncbi.nlm.nih.gov/refseq

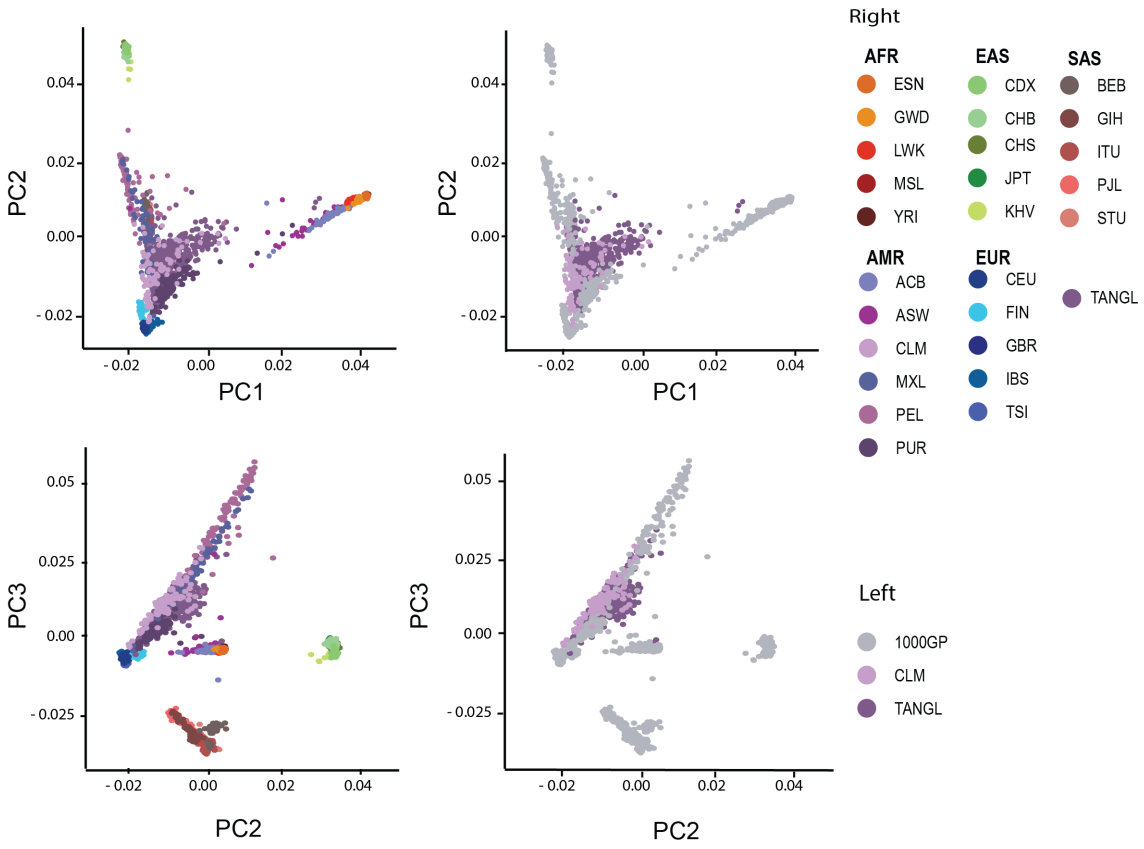


Figure 2.7: Principal Component Analysis of whole genomes from 1000 Genomes project and the TANGL cohort. ESN: Esan in Nigeria. GWD: Gambian in Western Divisions in the Gambia. LWK: Luhya in Webuye, Kenya. MSL: Mende in Sierra Leone. YRI: Yoruba in Ibadan, Nigeria. ACB: African Caribbean in Barbados. ASW: Americans of African Ancestry SW USA. CLM: Colombians from Medellin, Colombia. MXL: Mexican Ancestry from Los Angeles USA. PEL: Peruvians from Lima, Peru. PUR: Puerto Ricans from Puerto Rico. CDX: Chinese Dai in Xishuangbanna, China. CHB: Han Chinese in Beijing, China. CHS: Southern Han Chinese. JPT: Japanese in Tokyo, Japan. KHV: Kinh in Ho Chi Minh City, Vietnam. CEU: Utah Residents with Northern and Western European Ancestry. FIN: Finnish in Finland. GBR: British in England and Scotland. IBS: Iberian Population in Spain. TSI: Toscani in Italia. BEB: Bengali from Bangladesh. GIH: Gujarati Indian from Houston, Texas. ITU: Indian Telugu from the UK. PJL: Punjabi from Lahore, Pakistan. STU: Sri Lankan Tamil from the UK

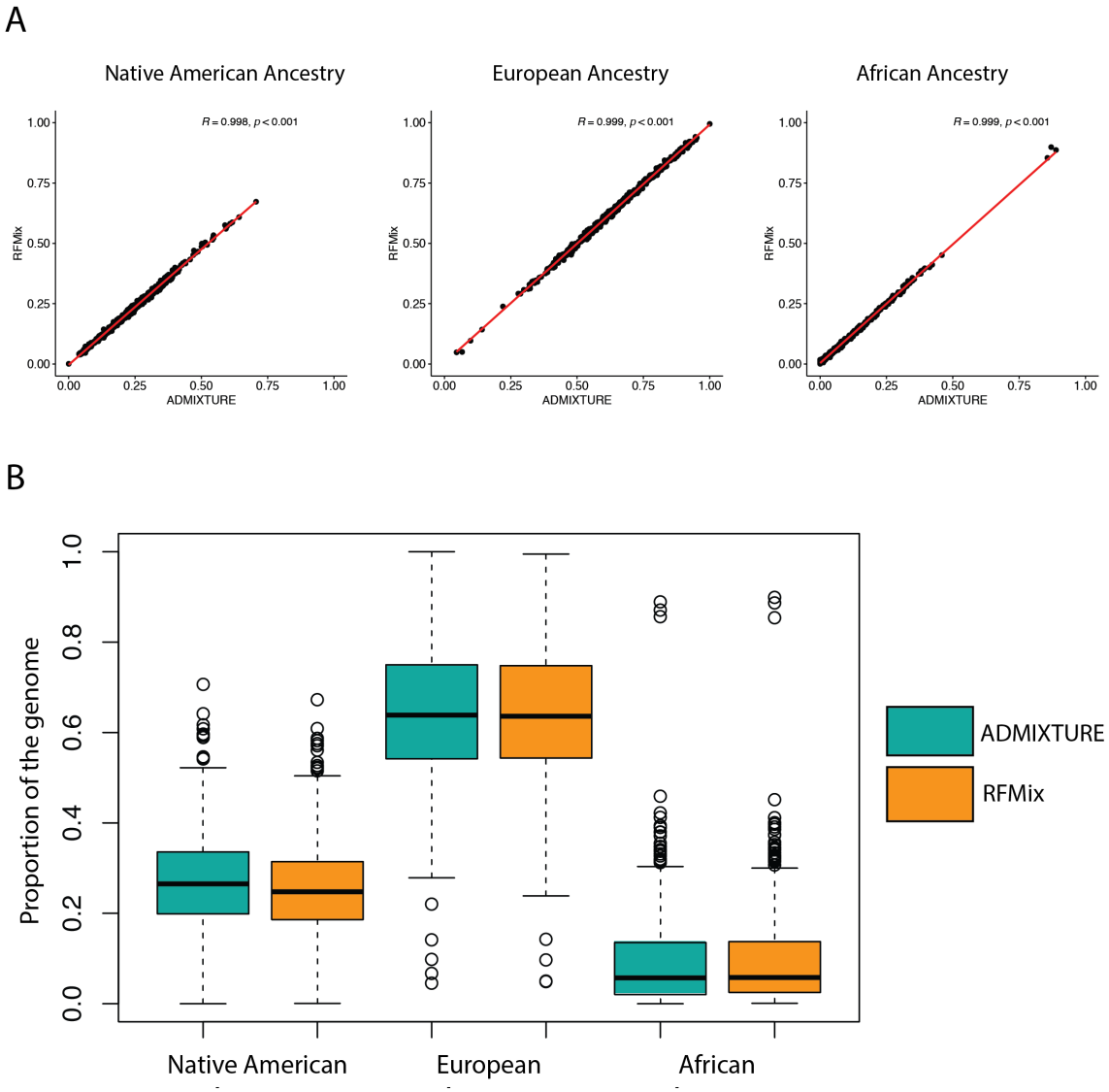


Figure 2.8: Global ancestry proportions of the TANGL cohort calculated by ADMIXTURE and sum of RFMix local ancestry estimation. (A) Correlation of global ancestry proportions calculated for each individual by two different software, RFMix sum of local ancestries (Y axis) vs ADMIXTURE (X axis) ; R: Pearson correlation coefficient. (B) Global ancestry proportions for the TANGL cohort calculated by ADMIXTURE and RFMix.

Haplogroup	Subclade	Origin	n	%
L0	L0a	Subsaharian Africa	3	0.5
L1	L1b	West Africa	8	1.4
	L1c	West and Central Africa	8	1.4
L2	L2a	Subsaharian Africa	7	1.2
	L2b	West Africa	1	0.2
	L2c	West Africa	6	1.1
	L2d	West Africa	1	0.2
L3	L3b	West-Central Africa	5	0.9
	L3d	East-Central Africa	3	0.5
	L3e	West-Central Africa	16	2.8
	L3f	Northeast Africa/Arabic peninsula	2	0.4
L4	L4b	East Africa	1	0.2
M	M1a	North Africa/Middle East/Mediterranean	1	0.2
	M1b	North Africa/Middle East/Mediterranean	1	0.2
C	C	America	44	7.8
D	D1	America	15	2.6
A	A2	America	236	41.7
R	R	Arabian peninsula/North Africa	1	0.2
B	B	America	177	31.2
H	H	Europe	1	0.2
	H1b	European/Near East/Caucasus/Jewish diaspora	6	1.1
	H1c	European/Near East/Caucasus/Jewish diaspora	1	0.2
	H1e	European/Near East/Caucasus/Jewish diaspora	1	0.2
	H1j	Western Europe	1	0.2
	H14a	European/Near East/Caucasus	1	0.2
	H20	Eastern Europe and the Caucasus	1	0.2
	H6a	Eastern Europe and the Caucasus	1	0.2
V	V1a	Europe	1	0.2
J	J1b	European/Near East/Caucasus	5	0.9
	J1c	European	3	0.5
	J1d	European/Near East/Caucasus	1	0.2
	J2a	European/Near East/Northern Africa	1	0.2
T	T2b	Europe/Middle East/North Africa.	3	0.5
U	U2e	Europe	1	0.2
	U5b	Central Europe	1	0.2
K	K2a	European/Near East	1	0.2
			994	100

Table 2.6: Mitochondrial haplogroups of the probands. Haplogroup nomenclature according to Phylotree 17

Haplogroup	Subclade	Origin	n	%
E	E1a2	West Africa	3	0.9
	E1b1	Middle East, Northern Africa	28	13.7
G	G2a1	Caucasus (Mediterranean basin)	2	0.6
	G2a2	Caucasus (Mediterranean basin)	4	1.4
I	I1a2	Northern Europe	2	0.6
	I1a3	Northern Europe	1	0.3
	I2a1	Sardinia	4	1.4
	I2a2	Balkans	2	2.8
J	J1a2	Semitid/Bedouinid Arabids (Mediterranean basin)	12	4.0
	J2a1	Syrid/Nahrainid Arabids (Mediterranean basin)	10	4.3
	J2a2	Syrid/Nahrainid Arabids (Mediterranean basin)	3	1.1
	J2b2	Syrid/Nahrainid Arabids (Mediterranean basin)	3	1.4
T	T1a1	Iberian	11	4.3
Q	Q1a2	America	13	6.0
R	R1b1	West Europe	126	57.3
			224	100

Table 2.7: Y chromosome haplogroups of the male probands. Haplogroups classified according to ISOGG Y-DNA Haplogroup Tree 2019-2020

Chapter 3

Genetic origin of a large family with a novel *PSEN1* variant

3.1 Introduction

Many genetic variants contribute to Alzheimer's disease (AD); however a small percentage of AD cases have highly penetrant autosomal dominant disease causing variants in *PSEN1*, *PSEN2* or *APP* [153]. Over 280 different autosomal dominant pathogenic variants are currently recognized worldwide [154]. The clinical phenotype of autosomal dominant AD is often similar to sporadic AD; however some cases have atypical presentations such as spastic paraparesis [155]. Although the age of onset is usually in the fifth to sixth decade the range of onsets is broad from 20 to 70 years [156]. The early onset has allowed investigators to study pre-clinical disease and the transition from pre-clinical to clinical AD [29, 157]

The first family with autosomal dominant AD identified in Colombia has a *PSEN1*

variant resulting in a Glu280Ala amino acid change in *PSEN1* [25]. This family is by far the largest family in the world with familial AD consisting of approximately 6000 family members clustered in the state of Antioquia, Colombia. Their age at onset for mild cognitive impairment (MCI) is 44 years (95% CI 43-45) and 49 years for dementia (95% CI 49-50) [28]. We report here a second large family with autosomal dominant AD who also live Antioquia and harbor a novel pathogenic variant c.1247T>C in codon 416 of *PSEN1* resulting in an Ile416Thr substitution. We describe the sociodemographic, clinical characteristics and neuropsychological profile of this family as well as the historical genetic context of their pathogenic variant. These data will broaden knowledge of the clinical spectra of autosomal dominant AD, extend the genetic loci in *PSEN1* capable of causing disease, reveal patterns of historic gene flow shaping the worldwide map of autosomal dominant AD, and possibly provide the participants an opportunity for prevention trials [158].

3.2 Subjects and methods

3.2.1 Participants

Sociodemographic, clinical, and genetic data from 93 family members 18 years or older were gathered between 2002 and 2016. (Figure 3.1) The pedigree was built with the information given by the family members. Participant evaluation was done in accordance to The Code of Ethics of the World Medical Association (Declaration of Helsinki) and the individuals signed an informed consent approved by the Institutional Review Board of the School of Medicine of University of Antioquia (Colombia). For cognitively impaired individuals, their proxies signed the informed consent.

The participants underwent medical, neurological, and neuropsychological evaluations

including the CERAD (Consortium to establish a registry for Alzheimer's disease) test battery with additional neuropsychological tests and dementia functional scales, as previously described [159, 50]. Diagnosis of Mild Cognitive Impairment (MCI) and dementia was done without knowledge of the carrier status according to Petersen 2011 [160] and DSM-IV criteria (American Psychiatric Association 2013) respectively. Collected data were stored in medical records software (SISNE V 2.0). Statistical analyses of the clinical data were done with Stata Statistical Software (Release 15, StataCorp 2017).

Sociodemographic, clinical and neuropsychological characteristics were presented using frequency and percentage (%) for categorical variables, and median and interquartile range (25th percentile-75th percentile) for quantitative variables.

3.2.2 Genome sequencing

A DNA sample from the proband was sent to a genetic laboratory (Athena Diagnostics Massachusetts, USA) for *PSEN1* variant testing. The analysis of *PSEN1* was performed by PCR amplification and automated uni-directional DNA sequencing of the coding region (10 exons, 1204 base pairs), plus 20 additional bases of intronic DNA flanking every exon. All abnormal sequence variants were confirmed by bi-directional sequencing. Peripheral blood samples from the 93 participants were obtained in EDTA tubes. Restriction fragment length polymorphism (RFLP) variation was used to identify the Ile416Thr carriers; DNA was extracted through a modified salting-out technique (Gentra Puregene Blood Kit by Qiagen), *PSEN1* Exon 11 was amplified by PCR (forward primer ACAGCAGCATCTACAGTTA, reverse primer TCAGGGCAGAGCTTATAGTT) and the amplicon was digested with VspI to detect NM_000021.3: c.1247T>C (p.Ile416Thr).

In addition, whole genome sequencing of 31 individuals (9 asymptomatic non-carriers, 12 asymptomatic carriers and 10 symptomatic carriers.) was completed using the Il-

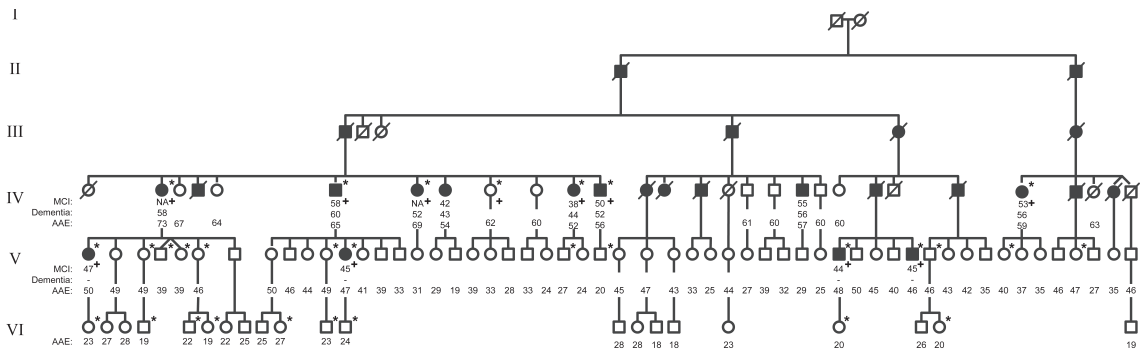


Figure 3.1: The diagram represents the extended *PSEN1* Ile416Thr family, square for male and circle for female. Individuals with MCI and dementia have been classified as symptomatic and are represented as filled icons, empty icons represent asymptomatic individuals. Individuals from whom we have information but died before the study were represented as deceased (crossed out). All 93 participants in the study were labeled with age at onset for MCI (if applicable), age at onset for Dementia (if applicable), and age at the clinical evaluation. If the age at onset was unknown the data was labeled as Not Available (NA). The symptomatic individuals from the V generation met the MCI criteria, but none of them met dementia criteria, therefore AAO for dementia was not registered. MCI: Mild cognitive impairment AAE: age at examination NA: data not available ∗Individual with whole genome sequencing data + Individual used for zygosity mapping

lumina HiSeqX platform (Human Longevity Inc., San Diego, CA). Samples were sequenced to reach an average 30X coverage with single indexed paired-end 150 base pair reads. Raw sequencing data was processed using the Broad Institute's Genome Analysis Toolkit (GATK v3.2.2) best-practices pipeline [161]. Reads were aligned to the hg19 reference assembly of the human genome using the Burrows-Wheeler Aligner (BWA-MEM v0.78)[162]. Joint single nucleotide polymorphism (SNP) variant calling was performed in all 31 samples using GATK HaplotypeCaller. After joint-calling, SNPs were filtered to retain those with genotype quality (GQ) scores greater than 30 and read depth (DP) scores greater than 20. Heterozygous sites were also filtered to retain SNPs with an allelic balance (AB) greater than 0.25 and less than 0.75. In PLINK [67], SNPs in linkage equilibrium (r2 <0.2) with genotyping rates above 99% and minor allele frequencies (MAF) greater than 5% were retained for estimating identity-by-state (IBS) and relatedness (Pi-hat) by calculating genome-wide IBS distances for all pairs of individuals. Multi-dimensional scaling (MDS) analysis of the IBS pairwise distances was conducted in PLINK utilizing overlapping SNPs from the 31 participants and individuals from the 1000 Genomes Project (phase 3 data release) that represent European (CEU, n = 103), African (YRI, n = 111), and East Asian/Native American (CHB, n= 108)[100].

To identify a shared haplotype among the 22 carriers of the pathogenic *PSEN1* Ile416Thr variant, SNPs were phased using Beagle v4.1 [163] with 1000 Genomes Project populations used as a reference panel. The frequency of the Ile416Thr-associated haplotype in reference populations was determined using phased genotype data for chromosome 14, and all haplotype frequencies were confirmed using LDlink [164]. Recombination rates were obtained from the hg19 genetic map provided within Eagle v2.4 [165].

LOD score calculation assumed a 100% penetrant association between genotype and phenotype, where LOD = $\log_{10}\left(\frac{1}{0.5^i}\right)$ and i is the number of individuals. For identifying

the genomic region associated with the phenotype (zygosity mapping), we first filtered variants for quality to single nucleotide variants in the PASS tranche (GATK 99%) with a total depth of at least 30, and genotype quality of 99, not in a RepeatMasker region, and present at least once in the 1000 genomes reference. The presence of variants in the 1000 genomes reference was required to identify the haplotype that flanks the variant.

3.2.3 Brain imaging

Pittsburgh Compound B (PiB), flortaucipir (FTP) positron emission tomography (PET) and structural magnetic resonance imaging (MRI) measurements were acquired from three members of the family (2 *PSEN1* Ile416Thr carriers and 1 non-carrier) at Massachusetts General Hospital, Boston. 11C PiB PET data were expressed as the distribution volume ratio (DVR) with cerebellar grey as reference tissue, using a large cortical ROI aggregate that included frontal, lateral temporal and retrosplenial cortices (FLR) as described previously [30]. 18F FTP regional binding was expressed in FreeSurfer ROIs as the standardized uptake value ratio (SUVR) to cerebellum.

3.3 Results

3.3.1 Clinical description of the proband

The proband was a 55-year-old woman who complained of slowly progressive anterograde memory impairment with onset at age 40 (individual IV-7 in Figure 3.1). She misplaced objects, forgot conversations, and repeated herself. She was no longer able to perform her daily activities without supervision. Her relatives reported she had insomnia, visual and auditory hallucinations, had become irritable and had mood swings. Her family recalled that her father, her grandfather, and other family members had similar

symptoms. Palmomental reflex and dysdiadochokinesia were the only findings on her physical examination. Her Mini Mental State Examination (MMSE) score was 18/30, 9 points below expected for her age and schooling; the CERAD neuropsychological battery showed severe compromise of verbal (5 points below expected in the recall of the CERAD word list) and non-verbal memory. The patient also had mild anomia and her verbal fluency was below expected. Her executive function, attention, praxis and visual perceptual skills were also impaired. She was not able to copy the Rey-Osterrieth complex figure. Her score for the Functional Assessment Staging Test (FAST) was 5, equivalent to moderate dementia.

The next clinical evaluation was performed when the patient was 62 years old. She scored 9/30 in the MMSE but was unable to comprehend the other tests. Her FAST score was 6e, equivalent to moderately severe dementia. The patient became bedridden when she was 64 years old; she was incontinent, aphasic, and completely dependent for activities of daily living. She had frequent myoclonus and tonic-clonic seizures with partial response to valproic acid. The patient died at the age of 70 due to sepsis from infected pressure ulcers.

Genetic testing of the patient identified a Thymine to Cytosine transition in exon 11 NM_000021: c.1247T>C (p.Ile416Thr). This missense variant affects a highly conserved residue in the eighth transmembrane domain of Presenilin1 (Figure 3.2). Every computational prediction algorithm checked predict deleteriousness, which included a CADD PHRED score of 25.9, a 99.8% conservation score by phyloP in 100 vertebrates, a 98.6% rankscore by GERP, and deleterious prediction by SIFT, PolyPhen2 and MutationTaster. The high level of conservation at this site, the predicted damage likely due to a change from a hydrophobic to a polar amino acid in the transmembrane domain, compounded by the location of the variant near a splice site all supported pathogenicity.

3.3.2 Demographics and medical assessment

A total of 93 family members, including the proband, were evaluated (Figure 3.1). Twenty-six (27.9%) participants were identified as carriers of the Ile416Thr variant. The majority of them were women (59.2%), but the male:female ratio was similar in carriers compared to non-carriers (Table 3.1).

	Non-carriers	**Ile416Thr carriers**			
	Total	Total	Asymptomatic	MCI	Dementia
	n = 67	n = 26	n = 14	n = 4	n = 8
Age at onset	NA	NA	NA	47.6 (5.8)	51.6 (5.0)
Age at evaluation	36.5 (14.3)	46.0 (14.4)	36.5 (12.0)	50.8 (3.8)	59.1 (8.7)
CDR					
0 (None)	63 (94%)	14 (53,9%)	14 (100%)	0	0
0.5 (Questionable)	4 (6%)	4 (15.4%)	0	4 (100%)	0
1 (Mild)	0	3 (11.5%)	0	0	3 (37.5%)
2 (Moderate)	0	2 (7.7%)	0	0	2 (25%)
3 (Severe)	0	3 (11.5%)	0	0	3 (37.5%)
Depression	7 (10.4%)	8 (30,8%)	1 (7.1%)	1 (25%)	6 (75%)
Anxiety	0	6 (23,1%)	1 (7.1%)	1 (25%)	4 (50%)
Delusions	0	3 (11.5%)	0	1 (25%)	2 (25%)
Hallucinations	0	3 (11.5%)	0	1 (25%)	2 (25%)
Insomnia	0	4 (15.4%)	0	1 (25%)	3 (37.5%)
Myoclonus	0	5 (19.2%)	0	0	5 (62.5%)
Seizures	0	4 (15.4%)	0	0	4 (50%)
TBI	9 (13,4%)	2 (7.7%)	1 (7.1%)	1 (25%)	0
Headache	0	4 (15.4%)	1 (7.1%)	2 (50%)	1 (12.5%)
Alcoholism	2 (3%)	4 (15.4%)	1 (7.1%)	0	3 (37.5%)
Tobacco use	14 (20,9%)	11 (42.3%)	5 (35.7%)	2 (50%)	4 (50%)
Drug abuse	6 (9%)	2 (7.7%)	2 (14.3%)	0	0

Table 3.1: CDR: clinical dementia rating. MCI mild cognitive impairment, TBI: Traumatic Brain Injury

After medical and neuropsychological evaluation four participants were diagnosed with MCI and eight participants with dementia; 37.5% were in early stages (CDR 1), 25% were in moderate stage (CDR 2) and 37.5% were severely demented (CDR 3), all of them were later confirmed as heterozygous Ile416Thr carriers. Symptomatic carriers had a mean age at onset of 42.35 years old (S.D. 6.28) for memory complaints, 47.6

years old (S.D. 5.83) for MCI, and the mean age of onset of dementia was 51.6 years old (S.D. 5.03) (Figure 3.3). Eight older healthy controls ($\geq$ 55 years) did not carry the variant. As the variant co-segregated with the disease in three or more cases within the family it was classified as definite pathogenic according to the Guerreiro algorithm [83]. Two cognitively normal Ile416Thr carriers had PET-amyloid and PET-Tau scans with evidence of preclinical amyloid plaques and tau accumulation while an age-matched non-carrier control did not show this pathology (Figure 3.4). The cerebral pattern of amyloid-β deposition in the carriers resembled that found in clinically affected individuals who are at risk for late-onset Alzheimer's disease and other cases of AD-causing *PSEN1* variants citeFleisher2012 and the PET-Tau scans resembled those previously reported citeFleisher2012. The findings include preferential PiB binding in posterior cingulate, precuneus, parietotemporal, frontal, and basal ganglia regions. Elevated levels of FTP binding in medial temporal lobe regions were evident within the context of increased amyloid pathology, before estimated years to symptom onset.

Depression and anxiety were more frequent in Ile416Thr carriers than in non-carriers, and more prevalent in those with MCI and dementia when compared to asymptomatic carriers (Table 3.1). Other neuropsychiatric symptoms such as delusions, hallucinations and insomnia were also reported and the neuropsychiatric inventory scores [166] were higher as the disease progressed. Myoclonus and bilateral tonic-clonic seizures were frequent in early stages as well as in patients with severe dementia (CDR 3). Neuropsychological assessment identified amnesic as the most common type of MCI. Patients with dementia had a relatively better performance in language and attention than praxis and executive function.

3.3.3 Zygosity mapping

To assess genetic variants associated with the phenotype in an unbiased manner, we performed zygosity mapping and calculated a LOD score assuming 100% penetrance and 100% concordance between genotype and phenotype (see methods for formula) using whole genome data filtered strictly for quality (see methods for filtering criteria) for 10 symptomatic individuals and one asymptomatic family member past the latest age of onset ($\geq$ 55 years). These 11 individuals yielded 215,030 variants meeting strict quality filters spread approximately evenly across the genome. The goal in selecting very high-quality variants (including presence in 1000 genomes at least once) was not to identify the variant itself, but rather to identify ancestral variants in the haplotype that flank the variant that are high quality reliable indicators. We filtered for variants with a genotype consistent with phenotype, i.e. a homozygous reference in the asymptomatic individual past the age of onset, and a heterozygous variant in all 10 symptomatic individuals (consistent with our assumption that the causative variant is dominant and 100% penetrant). This filter yielded 249 variants with 223 variants on chromosome 14 and with 222 variants lying between 62 MB and 76 MB, which is highly enriched by Fisher's exact for variants on chromosome 14 ($*p < 0.0001$, OR=223, OR 95% CI = 112-∞) when compared to 249 variants randomly distributed across the genome. On average, 9 out of 249 variants would be expected on chromosome 14 if a set of 249 variants were distributed at the same ratios of numbers of variants per chromosome for the unfiltered set of 215,030 variants meeting the original criteria (Figure 3.5). When the calculation was restricted to the 11 individuals with whole genome sequencing data available, the LOD score was 3.3 at concordant sites (equivalent to a p value of 0.00049). As the Ile416Thr carrier status was also known in two additional symptomatic individuals and seven additional healthy individuals beyond the latest age of onset without whole genome sequencing data available,

a total of 20 informative individuals could be used for LOD calculation, which yielded a LOD score of 6.0 given the stated assumptions for LOD calculation in the methods (equivalent to a p value of 9.5x10-7), providing strong evidence for linkage of this site to phenotype (3 orders of magnitude above the typical genome-wide LOD score cutoff of 3). No disease-associated variants in *TREM2*, *SORL1*, or *ABCA7* were observed in the 10 carriers past age of onset with WGS data, and there was no detectable contribution of *APOE* E4 alleles to modify age at onset.

The Ile416Thr variant had not been documented in disease specific or population databases such as the Genome Aggregation Database (gnomAD) of 138,632 individuals (17,210 of Latin American ancestry) [62] (gnomAD; http://gnomad.broadinstitute.org), the 1000 Genomes Project [100] (1000G; http://www.internationalgenome.org/), TOPMed database (62,784 individuals), UK10K, or ESP6500 suggesting the possibility of a founder effect. One caveat is that the admixture of the Colombian population from ancestral populations in which more rare variants are present (African and Native American ancestries) combined with the isolation of this group of individuals in Antioquia means that more rare variants may cluster by chance. Therefore, while the absence of this variant in over 200,000 healthy individuals in population databases is certainly a contributing source of evidence for this variants pathogenicity, is not sufficient evidence alone.

One approach for assessing the pathogenicity of this variant is to consider criteria set forth by the American College of Medical Genetics (ACMG) for this very purpose [82]. The segregation of the variant with the illness is scored as strong evidence (PP1S), the absence in population databases is scored as moderate evidence (PM2) and computational deleteriousness prediction is scored as supporting evidence (PP3). Because computational predictions can be overlapping (and we present several scores that overlap above, for example by reliance on conservation), they are scored at this lowest level of evidence

(supporting). Importantly, we were conservative here in not assigning the ACMG score of PM1 (location in a mutational hot spot and/or critical functional domain) despite the fact that this variant does lie in a transmembrane domain of *PSEN1* where other variants have been described [142] because there is also some population variation in this domain. Taken together, these three observations are consistent with a Likely Pathogenic assertion (1 strong and at least 1 moderate), which is diagnostic. Even if we consider the absence in population databases as only a supporting piece of evidence because of the caveats stated above, the variant would still meet Likely Pathogenic by ACMG criteria (1 strong and 2 or more supporting).

3.3.4 Origin of the Ile416Thr variant

MDS analysis suggested a primarily African and European admixed ancestry for this family (Figure 3.6). The haplotype structure around the Ile416Thr variant, after phasing the genotype data for 22 Ile416Thr variant carriers and nine non-carriers (n = 31 family members), identified a 16 SNP, 86.7 kilobase haplotype shared by all carriers spanning most of *PSEN1*. This haplotype encompassed the *PSEN1* disease variant locus and was absent in all non-carriers (Figure 3.7). Therefore, it seems likely that the variant occurred on this haplotypic background. All Ile416Thr carriers were heterozygous for 16-SNPs on the same strand as the variant, which further supports that the pathogenic variant occurs on this haplotypic background.

To identify the ethnic ancestry of the shared haplotype, we queried phased genotype data from participants in the reference populations of the 1000 Genomes Project. This background haplotype was observed in 25 individuals of African or admixed African ancestry in the 1000 Genomes Project, all of whom did not carry the variant. The haplotype occurred most frequently in Gambian individuals (n = 6; 2.7% frequency in

Gambian population). Interestingly, this haplotype was not observed in any European or Asian reference populations. Furthermore, the three SNPs most proximal to Ile416Thr in the shared haplotype (rs362375, rs177413, and rs17126104) occur most frequently in African populations, with rs362375 and rs17126104 found almost exclusively in these populations. Together, these results suggest that the Ile416Thr pathogenic variant occurs on a haplotypic background that is African in origin (Figure 3.7 and Table 3.2).

3.4 Discussion

This study describes a private *PSEN1* (Ile416Thr) variant leading to early onset clinical Alzheimer's disease in a large family. There are caveats to portions of the evidence for this variant's pathogenicity – for example, more rare variants are present in this admixed and isolated group leading to the potential a false positive signal despite the variant's absence in population databases, and the computational predictions we report are highly overlapping because of their dependence on conservation. However, even considering these caveats, the strong segregation data for this variant supports a Likely Pathogenic assertion by ACMG criteria. So far, we have registered 26 carriers of the Ile416Thr variant, 14 of them are in pre-symptomatic stages of the illness. Individuals with MCI have an amnestic profile and depressive symptoms are frequent. In the dementia stage, most of the patients have myoclonus and seizures at late stages of disease. As of today more than 200 variants in *PSEN1* gene have been described [154], including another private variant, Glu280Ala, which affects 25 related families with a total of 6000 individuals. Interestingly, both families come from different but nearby towns in the state of Antioquia, Colombia.

A founder from the Iberian peninsula was reported as the origin of the Glu280Ala variant [26], and as described here, a founder from Africa gave rise to the family with

the Ile416Thr variant. These vastly distinct geographic origins highlight the complex genetic admixture of contemporary Colombians. Whereas the Colombian population is a three-way admixture of European, Amerindian and African populations [101], the proportions of the admixture show great variation over relatively small geographic sectors [45]. Therefore, the fine structure of the population, defined as the mapping of familial genetic features onto the geographic locales where families reside, reflects specific population histories. Because this admixture occurred over a relatively short time period—less than 500 years—the historical genetic imprint has remained readily detectable. Further marking the historical genetic trace was the tendency, until very recently, for low geographic mobility among the progeny of the founder populations.

Consistent with the historic record, the family harboring the Ile416Thr variant arose from a mini-population bottleneck. Girardota is located in a distinct geographic sector known as "Tierras del Río Abajo" or Lands of the Lower River, just north of the modern Antioquia capital, Medellín. The region, where currently the towns Girardota and Copacabana are located, was legally designated Río Abajo in 1598 and the town which came to be known as Girardota (originally Hatogrande) was founded in 1620. This location created a degree of geographic isolation from surrounding population groups. In 1665, 103 slaves imported from Guinea, Angola, Congo and Cabo Verde were recorded in treasury of the Municipality of Girardota [167]. As the anti-slave movement took root, many of these individuals fled to a mountainous location in the region and were further isolated. The genetic findings, in which the variant lies within a unique haplotype, fit well with the historical record of the region. This novel *PSEN1* pathogenic variant supports the view that founder effects and genetic drift can explain the contemporary geographic dispersion of these rare early-onset Alzheimer variants.

The well-defined age at onset and the moderately large size of the family are ideal

for future prevention trials as appropriate pharmacological interventions emerge. Both the Ile416hr family reported here and the very large Glu280Ala family have similar ages at onset. The progression of disease is similar to the extensively studied progression of patients with Glu280Ala variant, the largest known family with genetic AD, and for which there is extensive data about the neuropsychological performance in different stages of the illness [28, 168].

3.5 Permissions and Attributions

The content of Chapter 3 is the result of a collaboration with Dr. Laura Ramirez Aguilar at the University of Antioquia.

Published material was reprinted with permission from the publisher:

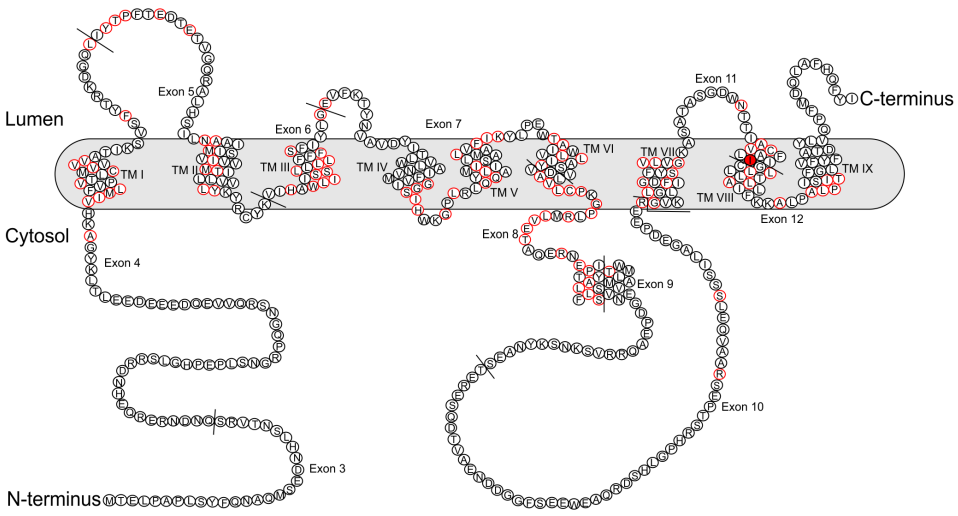
Ramirez Aguilar L∗, Acosta-Uribe, J∗ , Giraldo MM, Moreno S, Baena A, Alzate D, Cuastumal R, Aguillón D, Madrigal L, Saldarriaga A, Navarro A, Garcia GP, Aguirre-Acevedo DC, Geier EG, Cochran JN, Quiroz YT, Myers RM, Yokoyama JS, Kosik KS, Lopera F. Genetic origin of a large family with a novel *PSEN1* mutation (Ile416Thr). Alzheimer's & Dementia. 2019 May;15(5):709-719. doi: 10.1016/j.jalz.2018.12.010. Epub 2019 Feb 10. PMID: 30745123

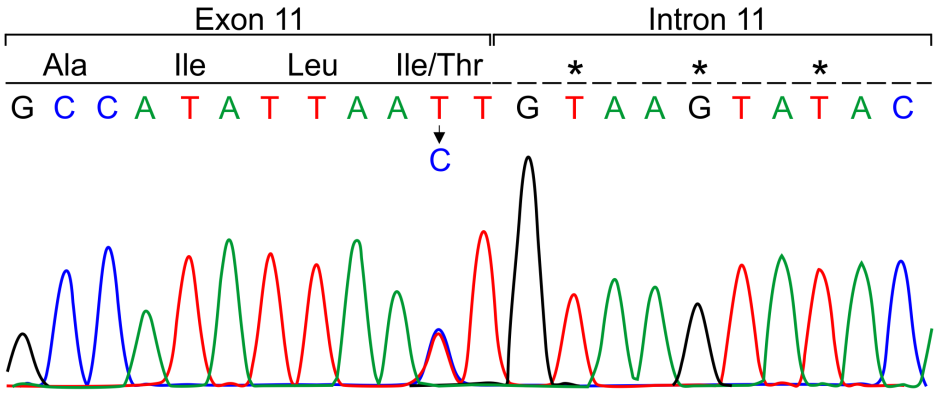
∗Co-first authors

3.6 Supplementary material

1000Genomes Super Population	1000 Genomes Population	I416T Haplotype Count	I416T Haplotype Frequency
AFR	ACB	2	0.01
(Frequency = 0.011)	ASW	3	0.025
	ESN	5	0.025
	GWD	6	0.027
	LWK	1	0.005
	MSL	3	0.018
	YRI	4	0.019
AMR	CLM	1	0.005
(Frequency = 0.0013)	MXL	0	0
	PUR	0	0
	PEL	0	0
EAS	CHB	0	0
(Frequency = 0)	JPT	0	0
	CHS	0	0
	CDX	0	0
	KHV	0	0
EUR	CEU	0	0
(Frequency = 0)	TSI	0	0
	FIN	0	0
	GBR	0	0
	IBS	0	0
SAS	GIH	0	0
(Frequency = 0)	PJL	0	0
	BEB	0	0
	STU	0	0
	ITU	0	0

Table 3.2: Frequency of Ile416Thr-associated haplotype in all 1000 genomes project reference populations. AFR, African; ACB, African Caribbeans in Barbados; ASW, Americans of African Ancestry in Southwest USA; ESN, Esan in Nigeria; GWD, Gambian in Western Divisions of Gambia; LWK, Luhya in Webuye, Kenya; MSL, Mende in Sierra Leone; YRI, Yoruba in Ibadan, Nigeria; AMR, Ad Mixed American; CLM, Colombians from Medellin; MXL, Mexican ancestry from Los Angeles, PUR, Puerto Ricans from Puerto Rico; PEL, Peruvians from Lima, Peru; EAS, East Asian; CHB, Han Chinese in Beijing, China; JPT, Japanese in Tokyo, Japan; CHS, Southern Han Chinese; CDX, Chinese Dai in Xishuangbanna, China; KHV, Kinh in Ho Chi Minh City, Vietnam; EUR, European, CEU, Utah Residents (CEPH) with Northern and Western European Ancestry; TSI, Toscani in Italia; FIN, Finnish in Finland; GBR, British in England and Scotland; IBS, Iberian Population in Spain; SAS, South Asian; GIH, Gujarati Indian from Houston, Texas; PJL, Punjabi from Lahore, Pakistan; BEB, Bengali from Bangladesh; STU, Sri Lankan Tamil from the UK; ITU, Indian Telugu from the UK.





NM_000021.3: c. 1247T>C (p.Ile416Thr)

Human	413	AILIGLCLTLLLLAIFKKALPALPISITFGLVFYFATDYLVQPFMDQLAFHQFYI
Rhesus macaque	413	AILIGLCLTLLLLAIFKKALPALPISITFGLVFYFATDYLVQPFMDQLAFHQFYI
Mouse	413	AILIGLCLTLLLLAIFKKALPALPISITFGLVFYFATDYLVQPFMDQLAFHQFYI
Chicken	414	AILIGLCLTLLLLAIFKKALPALPISITFGLVFYFATDNLVQPFMDQLAFHQFYI
Zebrafish	402	AILIGLCLTLLLLAIFKKALPALPISITFGLVFYFATDNLVRPFMDQLAVHQFYI

Figure 3.2: (A) schematic representation of Presenilin 1 in the cell membrane. Residues that have variants are known to cause Alzheimer's disease are represented with a red border. Residue 416 in TM VIII is colored in red. TM: Transmembrane (B) Color chromatogram depicting the missense variant in the *PSEN1* gene.(C) Alignment of *PSEN1* orthologs (uniprot.org). Residue 416 is highlighted in yellow. ∗Residues that are conserved among the 5 species. • Residues that are conserved at least in 4 species. : Residues that are conserved at least in 3 species.

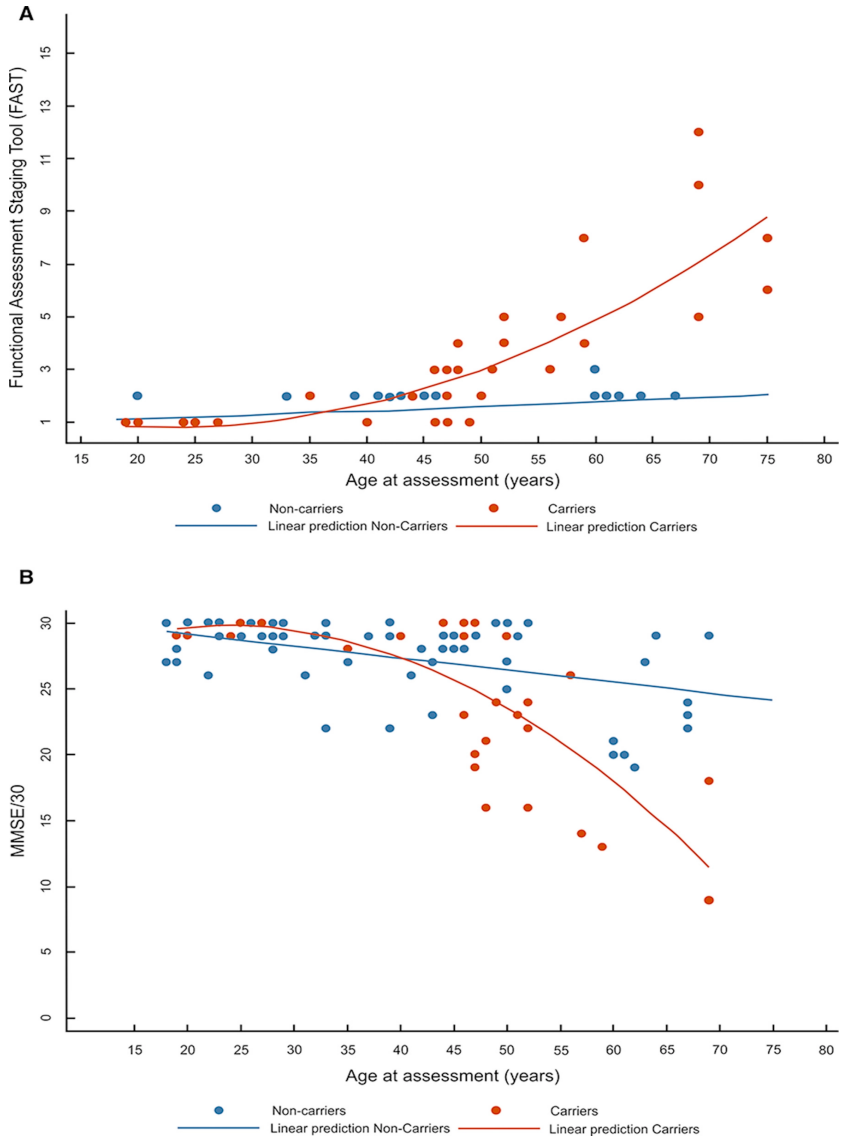


Figure 3.3: Age dependant cognitive decline in I416Thr carriers. FAST: Functional Assesment Staging Tool. FAST score was converted into a linear scale, where 6a corresponds to 6, 6b to 7 and so on until 16. Lineal and curvilinear models were used to represent, in non-carriers and carriers, respectively, the relation between the FAST (A) and the Minimental State Examination (MMSE) score (B) and the age at assessment (in years) centered at 40 years. Individuals with overlapping ages and scores were represented as a single point.

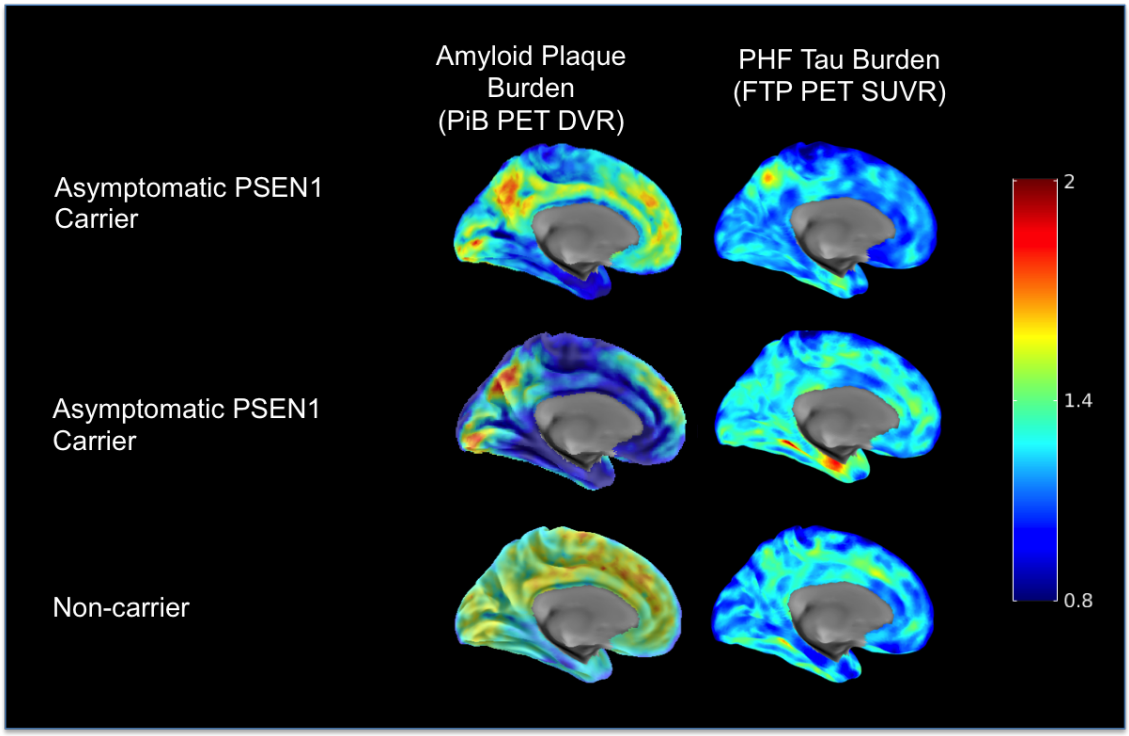


Figure 3.4: PET maps are shown for two cognitively unimpaired carriers (MMSE>27) and one age-matched non-carrier family member. Sagittal PiB DVR maps are shown on the left, and sagittal FTP SUVR maps are presented on the right. Images are displayed in standardized atlas space, along with whole-brain surface renderings, with a left hemisphere view. Row (A) A cognitively unimpaired carrier with high cortical amyloid (DVR=1.39) and with FTP binding in entorhinal cortex (SUVR=1.35). (B) A cognitively unimpaired carrier with higher cortical amyloid (DVR=1.41), and FTP binding in entorhinal cortex (SUVR=1.12). (C) An age-matched unimpaired non-carrier with low amyloid by PiB PET (DVR=1.10) and low FTP binding in entorhinal cortex (SUVR=1.11). PET: positron emission tomography, MMSE: Minimental State Exammination, PiB: Pittsburgh Compound B , DVR: Distribution volume ratio, FTP: Flortaucipir, SUVR: Standardized uptake value ratio.

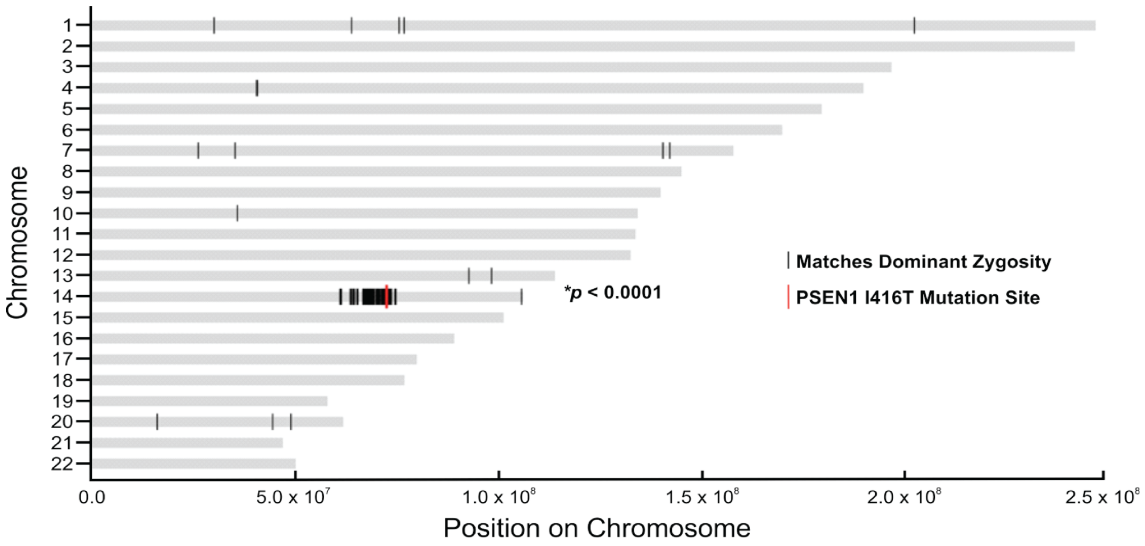


Figure 3.5: Whole genome data (WGD) from 11 individuals was filtered to high quality variants with zygosity matching phenotype (heterozygous alternate for 10 affected individuals and homozygous reference for 1 healthy individual past the latest age of onset). Variants are highly enriched on chromosome 14 ($*p < 0.0001$, OR=223, OR 95% CI = 112-∞). Note the location of the candidate variant in the block of variants with zygosity matching phenotype between 62 MB and 76 MB on chromosome 14.

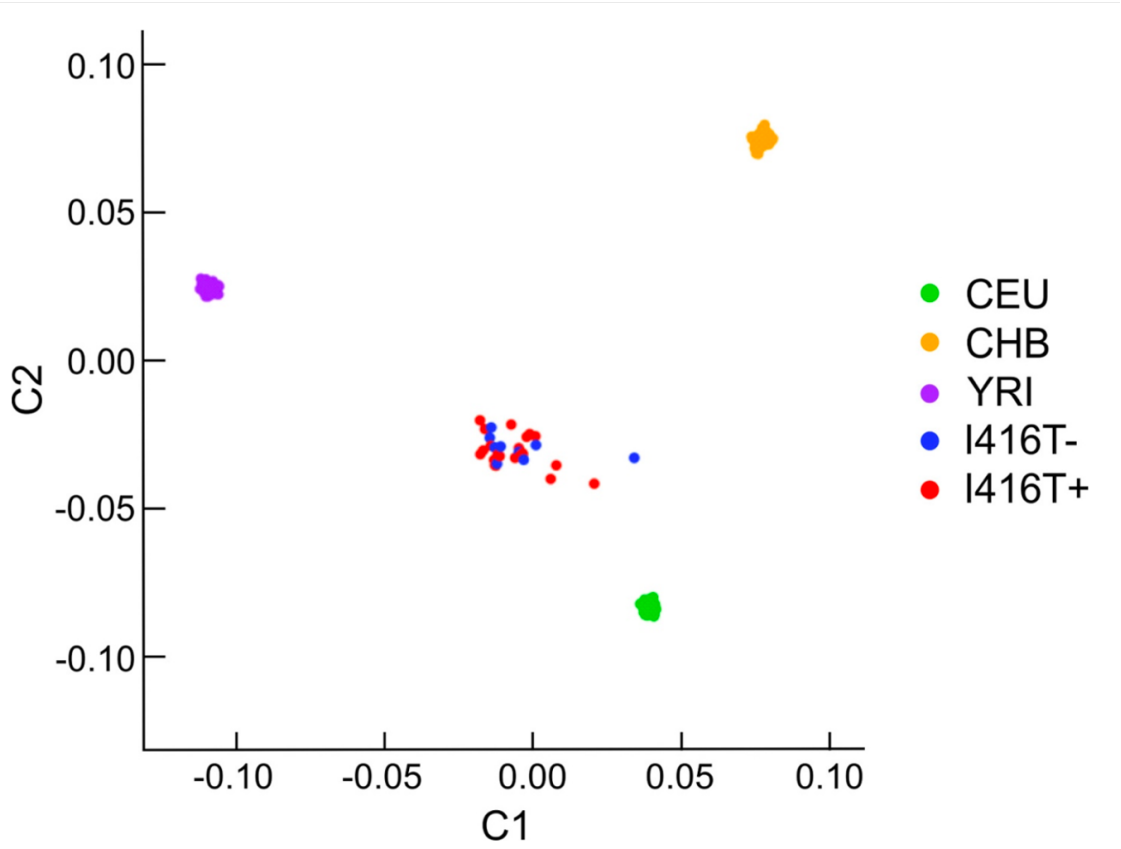


Figure 3.6: Population substructure in 22 Ile416Thr variant carriers (red) and nine non-carriers (blue) was assessed by multidimensional scaling analysis (MDS) and visualized by plotting the first two dimensions. Individuals from the CEU (green), YRI (purple), and CHB (yellow) populations in the 1000 Genomes Project were used as references. (CEU: Utah Residents with Northern and Western European Ancestry. CHB: Han Chinese in Beijing, China. YRI: Yoruba in Ibadan, Nigeria)

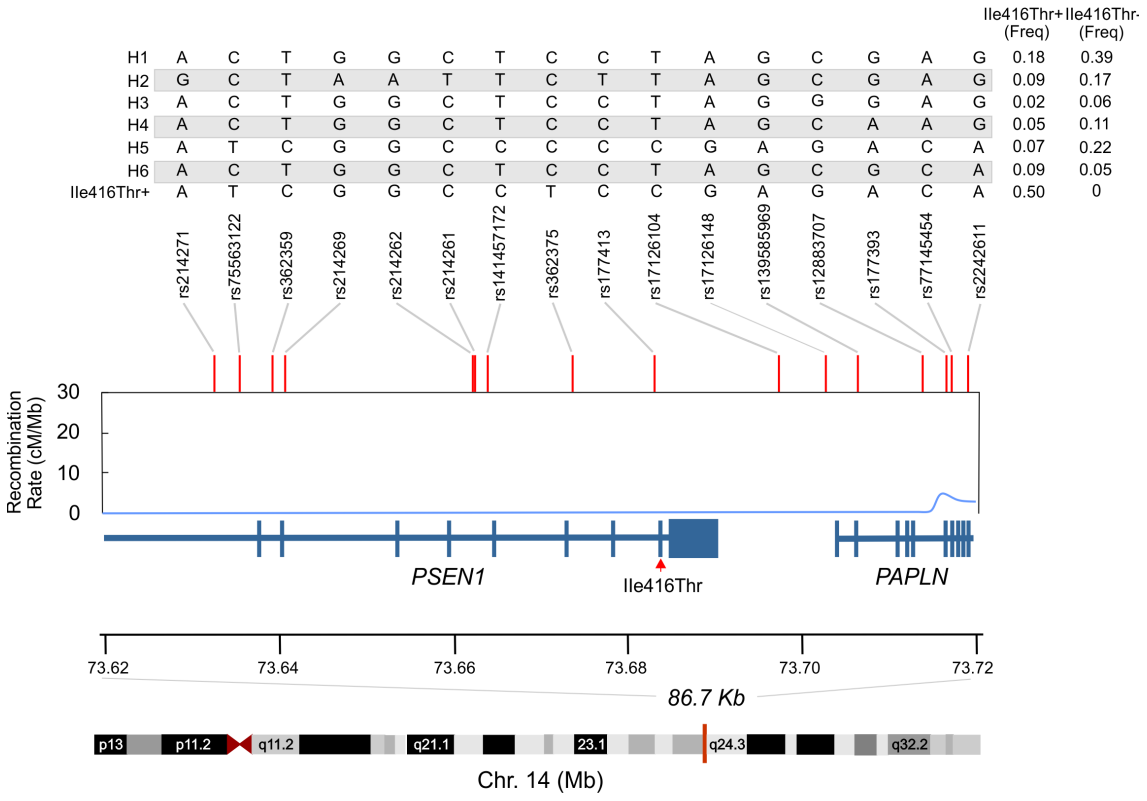


Figure 3.7: Haplotype analysis of phased genotypes from 22 Ile416Thr variant carriers and 9 non-carriers (n = 31) identified seven haplotypes composed of 16 single nucleotide variants (SNVs) spanning 86,664 base pairs on chromosome 14. This haplotype spanned exons 4 – 12 of *PSEN1*, including the Ile416Thr variant (last amino acid of exon 11, and exons 1 - 10 of *PAPLN*. The combination of alleles for all 16 SNVs in each of the seven identified haplotypes are provided in the table in the top panel. A single copy of one haplotype was found exclusively in all variant carriers and was absent from all non-variant carriers (I416T+, last row of top panel table). Recombination rates are represented in centimorgans per megabase, and were obtained from the hg19 genetic map provided as part of the Eagle v2.4 package. The red arrowheads point to the site of the Ile 416Thr variant. Freq, frequency

Chapter 4

Genetic Associations with Age at Dementia Onset in the *PSEN1* E280A Kindred

4.1 Introduction

Complex genetic, environmental, and lifestyle risk factors confounded by the aging process underlie risk for late onset Alzheimer's disease (LOAD). Autosomal dominant Alzheimer's disease (ADAD) closely resembles the clinical and neuropathological features of LOAD, but without the confound of aging, and thus provides a less heterogeneous view of underlying AD-associated processes. ADAD accounts for less than 1% of all cases of AD and mutations in *PSEN1* account for 80% of this monogenic group[169].

The age of onset of all other individuals with a given ADAD mutation correlates with an individual's age of onset (r2 = 0.52)[156], but there remains substantial unexplained

variability in age at onset. Large ADAD families such as the kindred harboring the Colombian *PSEN1* NM_000021:c.839A>C, p.(Glu280Ala) (canonically known as *PSEN1* E280A) mutation, the world's largest ADAD founder population with a comprehensive family tree of thousands of individuals [25],provide an opportunity to assess the contribution of genetic variation to unexplained variability in age of dementia onset. *PSEN1* E280A mutation carriers typically develop MCI at a median age of 44 (95% CI, 43–45) and dementia at age of 49 (95% CI, 49–50)[28]. The value of this family for the nomination of genetic variants that delay the onset of Alzheimer's disease was recently affirmed by the report of a *PSEN1* E280A carrier who developed mild cognitive impairment nearly three decades after the kindred's median age at clinical onset[32]. This individual was homozygous for the rare *APOE* E3 Christchurch variant (*APOE* NM_000041:c.460C>A, p.(Arg154Ser)) and had an exceptionally high amyloid-β plaque burden, but limited neurofibrillary tau burden. In addition to this case report, several studies have explored genetic associations with age at onset in this population of patients, but all with substantially lower numbers of cases (at most 72 individuals)[26]. To expand on the valuable insights gained from these previous studies, we conducted the most comprehensive search to date for genetic variants associated with age of dementia onset in this founder population by assessing 344 individuals, which is the current snapshot of all individuals from this cohort that currently have high quality genotypic and phenotypic information available.

4.2 Subjects and methods

4.2.1 Participants

A cohort of 368 patients was selected from the Neuroscience Group of Antioquia (GNA) database of the *PSEN1* E280A family. Selection criteria included being a *PSEN1*

E280A carrier with diagnosis of dementia, having adequate medical and neuropsychological evaluations and follow-up for a confident age determination of clinical age of dementia onset, and having a DNA sample. Participants were evaluated following a standard protocol including physical and neurological examination, as well as population-validated neuropsychological assessment. The clinician who performed these evaluations was blinded to the genetic status of the patient. Dementia was diagnosed according to most recent DSM criteria at the time of diagnosis. Collected data were stored in medical records software (SISNE v2.0). Family history was obtained from the patients and their relatives, and genealogical data from baptism and death certificates was gathered from local parishes and was incorporated into the pedigree reconstruction. Blood samples from each individual were obtained through standard phlebotomy and collected in EDTA tubes. Genomic DNA was purified from peripheral blood leukocytes using a modified salting-out technique (Gentra Puregene Blood Kit, Qiagen). All individuals were genotyped for *PSEN1* NM_000021:c.839A>C, p.(Glu280Ala) using a restriction length fragment polymorphism assay.

4.2.2 Genome sequencing

Genotyping arrays

We used the Illumina Multi-Ethnic Genotyping Array plus Neuro consortium content (catalog WG-316-1014, beadchip 20028352) to sequence 373 genomes over two batches (299 in batch one and 74 in batch two) representing 368 presumed unique individuals. The five duplicated individuals were from the following: four individuals from batch one with high missingness were re-run, and a new aliquot of one sample from batch one that was identified as a duplicate of another sample was re-run. Of the 373 array measurements, three were excluded because they were identified as mislabeled duplicates

of other samples, and 23 were excluded because they had a missingness rate of greater than 5%. 347 unique individuals with high quality array data remained for analysis. A total of 1,923,394 variants were genotyped. Array data were annotated with build hg38.

Whole Genome Sequencing (WGS)

Twenty-six WGS of *PSEN1* E280A carriers had been sequenced at HudsonAlpha in the TANGL cohort (Chapter 2)[additional data on the sequencing methods can be found in section 2.2.2]. These WGS included seven out of eight of the oldest age of onset individuals along with their affected family members with more typical ages of onset. 55 genomes from additional E280A carriers were sequenced previously with CompleteGenomics technology, and were also aligned to the hg19 reference [26]. This data was lifted over from hg19 to hg38 using CrossMap 0.2.7[60]

Variant measurement, imputation, and quality control

Imputation was conducted using the TOPMed Imputation Panel and Server (version 1.3.3), which includes 97,256 references samples and 308,107,085 variants and uses Minimac4 for imputation. Pre-imputation scripts (version 4.3.0 from William Rayner at the University of Oxford) were run using default settings, which filtered out palindromic single nucleotide variants (SNVs) with minor allele frequency (MAF) >0.4 or variants with >0.2 MAF difference from the TOPMed reference panel. For samples with both array and genome sequencing data, imputed genotypes were checked for concordance with genome sequencing data using SnpEff 4.3s. 540,753 high quality variants from array measurements for all individuals that overlapped with measured genome sequencing variants for 80 individuals were used for imputation using the TOPMed imputation server, and imputed variants were filtered for presence in genome sequencing data resulting in a final set of 9,451,360 variants. Median concordance between imputed genotypes with these

filters and genome sequencing genotypes was 99.4% for the 80 samples with both array and genome sequencing data available, indicating high quality imputation.

***APOE* sequencing**

Because of the importance of the *APOE* locus, we inspected the imputation results for rs429358 and rs7412, which together define *APOE* E alleles. All genotypes matched between imputed and separately genotyped values for rs7412, and all but one genotype matched between imputed and separately genotyped values for rs429358. For the discordant genotype, the directly measured genotype was used. In addition, although it was not imputed because of rarity, we manually inserted directly measured genotypes for rs121918393 because of its implicated role in a previous study[32].

4.2.3 Association analyses

Of the 346 individuals, 344 had age at dementia onset data available and were used for all analysis. Association analyses was performed using PLINK v1.9 , PLINK v2 [170, 68], and GEMMA[171].

4.2.4 Replication sets

Three cohorts were selected for replication: an AD age of onset study (9,162 cases)[172]), a study of AD age at onset survival (14,406 cases and 25,849 controls)[173], and a recent GWAS meta-analysis for AD (21,982 AD vs. 41,944 controls)[174].

4.3 Results

4.3.1 Cohort description

The final cohort had a mean age of dementia onset of 49.2 years [range: 37–75, 10th–90th percentile: 43–56]. 200 of the patients were genetically female (58.1%). Alcohol abuse was reported for 46 patients. The patients had extensive follow up data; the mean number of medical evaluations was 6.7 [1–27], and 4.8 [1–18] for neuropsychological evaluations. A partial pedigree of enrolled individuals annotated with age of dementia onset is presented in Figure 4.1.

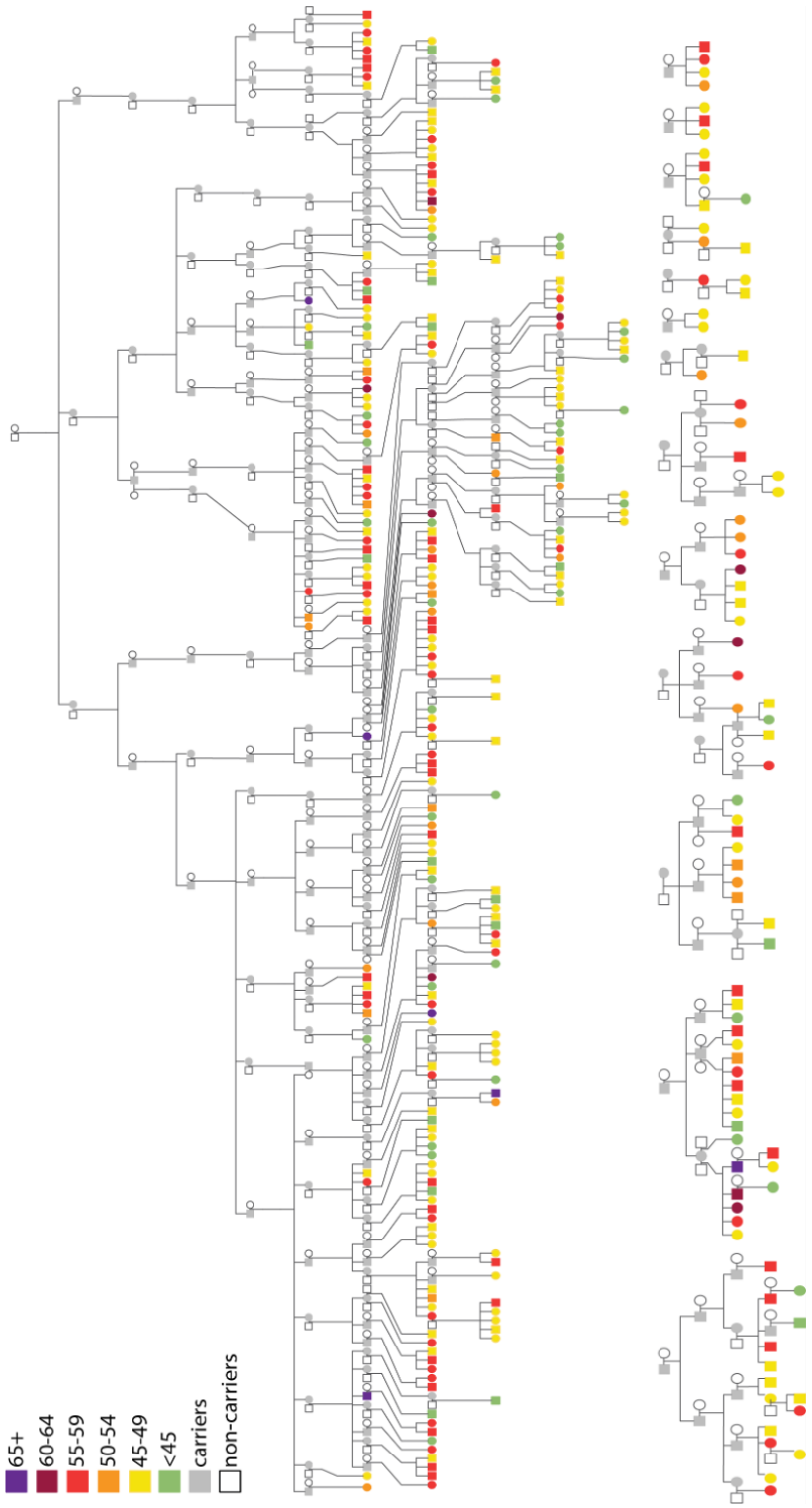


Figure 4.1: Pedigrees of included *PSEN1* E280A patients colored according to age at dementia onset. The E280A variant carriers (known and presumptive) that were not included in the present study are colored in gray. The large pedigree includes 252 of the sequenced participants, and their relatedness until the most recent common ancestors (which were born in 1743–1750). The small pedigrees in the lower row represent 87 additional participants. Healthy siblings and descendants of the participants were excluded for simplicity of the pedigree.

4.3.2 Assessment of biological relevance of identified signals

Association analysis was conducted using age at dementia onset as a quantitative outcome for 344 individuals passing QC. We employed PLINK with allelic, dominant, genotypic, and recessive models as well as GEMMA, a package that performs a likelihood ratio test using a linear mixed model to adjust for relatedness between individuals. For both packages, we adjusted for genetic sex, the first two principal components (calculated from the set of 540,753 high quality variants used as imputation input using PLINK), reported alcohol abuse, and batch.

We first assessed the dataset as a whole. To assess if the number of hits observed was greater than random chance, we generated QQ plots (Figure 4.4). QQ plots for the age of dementia onset did not deviate detectably from randomly scrambled (10 scrambles) ages of dementia onset across individuals, but this was not surprising given the small size of the cohort. We therefore pursued alternative approaches to assess the biological relevance of the observed associations. First, we compared the number variants with p<1x10–5 that exhibit nominal replication in one of the three replication cohorts (p<0.05) for the observed age of dementia onset compared to scrambles (Figure 4.2A), which revealed a definitive enrichment for replication with the observed ages of onset vs. scrambles. Second, we identified the regions surrounding (+/-500 kb) genome-wide significant hits (p<5x10–8), where multiple hits within 5 MB of one another were considered a part of the same continuous block. Concurrently, we identified these regions in scrambled conditions. We then searched (blinded to whether the source was from the observed ages or scrambled ages) the NHGRI-EBI GWAS catalog for GWAS hits within these regions with either a permissive relationship to AD (Figure 4.2B), or a strict relationship to AD (Figure 4.2C) (i.e. AD itself or related endophenotypes such as CSF tau and Aβ). This revealed an enrichment (observed >95% CI of scramble) of GWAS hits related to AD or other related

phenotypes with the observed age distribution for allelic, dominant, and LRT models, and enrichment of GWAS hits with a strict relationship to AD in allelic, dominant, LRT, and genotypic models. A final test for relevance comes from the proportion of variance explained ("chip heritability") calculated by GEMMA in the observed model at 56.4%, well above the median of 2.9% for scrambled conditions (Figure 4.2D).

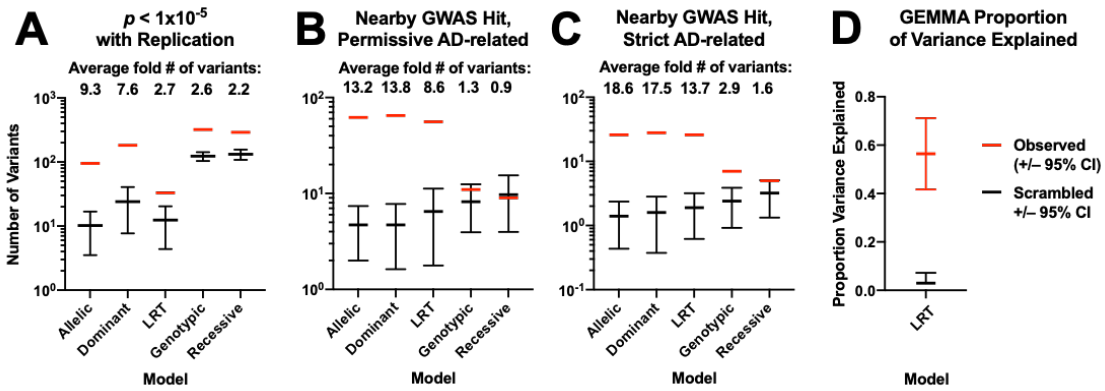


Figure 4.2: Comparisons to scrambled age distributions to assess biological relevance of signals observed. (A) Number of variants with p <1x10–5 in this cohort exhibiting nominal replication (p <0.05) in at least 1 cohort (of three tested) compared to the same cohort with age at dementia onset scrambled (10 scrambles) as a background control. (B) Number of variants +/–500 kb of p <5x10–8 variants (with variants considered as a continuous region from the first observed to last observed variant if multiple variants with p <5x10–8 were less than 5 MB from one another) in this cohort with a signal from the NHGRI-EBI GWAS catalog for permissive AD-related phenotypes (i.e. any aging, cognition, or neurodegeneration-associated phenotypes) compared to the same cohort with age at dementia onset scrambled (10 scrambles) as a background control.(C) Number of variants +/–500 kb of p <5x10–8 variants (with variants considered as a continuous region from the first observed to last observed variant if multiple variants with p <5x10–8 were less than 5 MB from one another) in this cohort with a signal from the NHGRI-EBI GWAS catalog for strict AD-related phenotypes (i.e. AD-only phenotypes or endophenotypes) compared to the same cohort with age at dementia onset scrambled (10 scrambles) as a background control. (D) Proportion of variance explained (PVE) from GEMMA. The 95% confidence interval provided from GEMMA is plotted for the Observed condition, while the 95% confidence interval was calculated based on the PVE values from all scrambles (10 scrambles) for clarity instead of plotting the provided 95% confidence interval outputted with each individual scramble condition. For all panels, errors bars indicated 95% confidence interval of the mean. For panels A–C, no error bars are available for the observed condition because it is a single condition. All models are from PLINK except LRT (likelihood ratio rest) which is calculated using GEMMA.

4.3.3 Genome-wide significant loci

Genome-wide significant loci (p<5x10–8) are shown in Table 4.1 and included two variants at different loci associated with clusterin biology, rs35980966 and rs138295139. In addition, the age of each individual harboring each variant in Table 4.1 is illustrated in Figure 4.3 by variant. Manhattan plots for all models are shown in Figure 4.5 and correlation plots for all models tested are shown in Figure 4.6. All models were correlated (average Pearson's r=0.54). Allelic, dominant, and LRT models were highly correlated (average Pearson's r=0.76) and genotypic and recessive models were highly correlated (Pearson's r=0.83). Because this is a founder population, we present p values from the LRT model (which accounts for family structure most stringently) for all variants listed, although we note that the LRT model p value may be overly conservative when compared to genotypic and recessive models because the LRT model does not assign additional weight to multiple alleles present in the same individual.

4.3.4 Results at key *APOE* variants

Effects of previously established *APOE* variants important for AD association in LOAD are in the expected direction based on previous studies, but modest in magnitude (Table 4.2). Overall, the observations are consistent with previously reported observations including a protective effect of *APOE* E2 in the Colombian E280A population [175], a deleterious effect of *APOE* E4 in the Colombian E280A population in one study[176] but an inability to detect an effect of *APOE* E4 in three other studies in this population [31, 177, 176], and a non-significant trend towards an *APOE* E2 >*APOE* E3 >*APOE* E4 age-of-onset in dominant AD families with a variety of mutations[156]. A recent case report implicated the *APOE* Christchurch variant (rs121918393)[32]. That individual is also enrolled in this study, and while we do observe a significant effect on age at onset

Chr	Pos (MB)	Variant	Best Model	p	LRT p	β (SE)	MAF	Replication
1	22	rs149331601	Allelic	2.5x10–8	4.2x10–8	21.8 (3.8)	0.003	
1	80.5	rs77107089	Allelic	7.7x10–9	4.7x10–7	6.1 (1)	0.044	Huang, 2017
1	112.8	rs115679813	Dominant	1.9x10–8	2.5x10–6	5.7 (1)	0.049	
1	194.4	rs138295139	Allelic	3.1x10–10	2.3x10–9	17.2 (2.7)	0.006	
1	237.9	rs79319775	Allelic	2.5x10–8	4.2x10–8	21.8 (3.8)	0.003	
3	23.9	rs11718580	Allelic	2.5x10–8	4.2x10–8	21.8 (3.8)	0.003	
4	82.1	rs181746127	Allelic	7.7x10–9	2.5x10–8	18.3 (3.1)	0.004	
5	95.7	rs75966684	Allelic	2.5x10–8	4.2x10–8	21.8 (3.8)	0.003	
5	166.3	rs951483	Allelic	3.4x10–8	9.3x10–8	17.6 (3.1	0.004	
5	176.3	rs56389738	Allelic	8.6x10–10	2.2x10–9	16.9 (2.7)	0.006	
6	51.5	rs34506826	Genotypic	5.0x10–9	0.02	11.4 (1.9)	0.112	
6	151.6	rs62444327	Recessive	1.3x10–9	0.12	19.2 (3.1)	0.105	
7	16	rs17352534	Recessive	3.7x10–8	0.02	9.4 (1.7)	0.151	
7	65.1	rs34848990	Genotypic	4.2x10–10	2.4x10–3	10 (1.5)	0.129	
8	27.6	rs35980966	Allelic	2.6x10–8	4.6x10–8	18 (3.2)	0.004	Kunkle, 2019
9	19.7	rs4977232	Genotypic	3.8x10–8	5.7x10–5	-3.6 (0.6)	0.740	
10	0.5	rs75435570	Allelic	3.4x10–8	9.3x10–8	17.6 (3.1)	0.004	
10	35.6	rs673438	Dominant	5.9x10–9	0.32	-22.7 (3.8)	0.923	
12	71.3	rs77075460	Allelic	3.4x10–8	9.3x10–8	17.6 (3.1)	0.004	
12	120.6	rs79248899	Allelic	2.5x10–8	4.2x10–8	21.8 (3.8)	0.003	
13	40.8	rs71427424	Recessive	5.9x10–9	0.09	22.7 (3.8)	0.125	
13	93.6	rs1887660	Recessive	5.0x10–9	2.0x10–3	7.2 (1.2)	0.224	
13	113	rs138705540	Allelic	2.5x10–8	4.2x10–8	21.8 (3.8)	0.003	
16	15.7	rs7195365	Genotypic	1.3x10–8	0.03	8 (1.4)	0.142	
19	3.2	rs4807426	Recessive	6.4x10–9	1.4x10–3	16 (2.7)	0.180	
20	36.5	rs35609661	Allelic	3.4x10–8	9.3x10–8	17.6 (3.1)	0.004	
21	22.3	rs17592663	LRT	3.0x10–8	3.0x10–8	4.6 (1.1)	0.039	
21	44.3	rs1800525	Allelic	1.7x10–10	2.1x10–8	4.2 (0.6)	0.110	
22	48.5	rs11705431	Allelic	7.7x10–9	2.5x10–8	18.3 (3.1)	0.004	

Table 4.1: Genome-wide significant hits (p<5x10–8). All models are from PLINK except LRT (likelihood ratio rest) which is calculated using GEMMA. The model with the lowest p value is shown. For all variants, the LRT p value is also shown, which accounts for family structure using a linear mixed model that adjusts for kinship. β is shown for each model (with the β from the Allelic model shown if LRT is the best model, as GEMMA does not output β values), and is derived from the basis of age of dementia onset. Replication indicates nominal significance (p<0.05) in the specified study with consistent effect direction

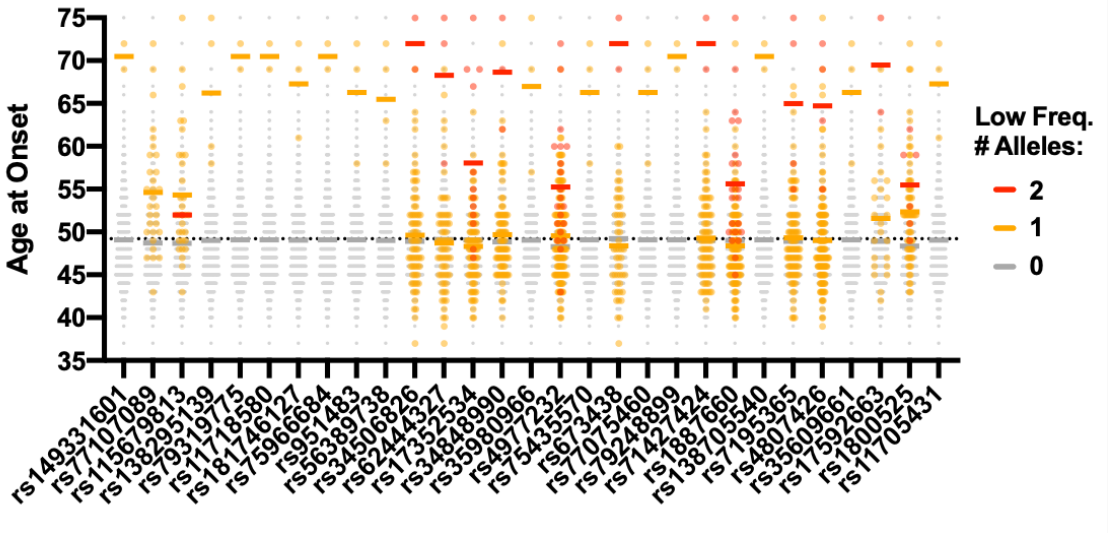


Figure 4.3: Genome-wide significant (p<5x10–8) variants by age at onset. Overall mean age is indicated by the black dashed line (49.2 years). For rs4977232 rs673438, the lower frequency allele is the reference allele. For all other variants, the minor allele is the lower frequency allele. Line indicates mean, points are individuals

of this variant, the effect is driven by this one individual with homozygosity at this variant. In a dominant model, which weights heterozygous and homozygous variants equally, there is not a detectable effect of the Christchurch variant (Table 4.2).

4.3.5 Replication at known AD-associated loci

We evaluated 15 AD genome-wide association studies (GWAS), including the largest case/control studies for AD in European populations, studies in non-Europeans, age of onset modifier studies, and endophenotype studies. These studies identified 55 loci (100 index variants) with the highest confidence associations for AD and endophenotypes. Replication of hits with genome-wide significance for AD-associated phenotypes with nominal significance (p<0.05) in at least one model and with consistent effect direction in this cohort are shown in Table 4.3 and included two variants at the *CLU* locus (rs9331896[12] and rs4236673[178]) and a stop gain variant in *IL34* (rs4985556[178]).

Variant	Dom. β(SE)	Dom. p	Rec β(SE)	Rec. p	LRT p	MAF	Hom.	PopAF
APOE E4	-2.3 (0.7)	9.7x10–4	-2.2 (2.0)	0.27	1.2x10–3	0.137	8	0.156
APOE Christchurch	1.8 (1.8)	0.31	25.6 (5.4)	2.8x10–6	0.027	0.019	1	8.0x10–6
APOE E2	2.6 (0.9)	3.0x10–3	NA	NA	0.014	0.067	0	0.082

Table 4.2: Assessment of previously implicated variants in *APOE*. Dominant (Dom.) and recessive (Rec.) models are from PLINK, while LRT (likelihood ratio rest) is calculated using GEMMA. The LRT p value accounts for family structure using a linear mixed model that adjusts for kinship. β is shown for dominant and recessive models (but not for LRT, as GEMMA does not output β values), and is derived from the basis of age of dementia onset. NA = not applicable, because no individuals in this cohort were homozygous for rs7412. MAF indicates minor allele frequency in this cohort. Hom. Indicates the number of individuals homozygous for the variant. PopAF indicates the population allele frequency (TOPMed).

Variant	Locus	Nearest Gene	Best Model	p	LRT p	β (SE)	MAF	PopAF	Previous Study	Previous Study p	Effect (Error)	Effect
rs6448453	4p16.1	*CLNK*	Allelic	0.002	0.008	1.6 (0.5)	0.75	0.77	Jansen, 2019	1.9x10−09	0.99 (0.98–0.99)	AD Risk
rs62341097	4q34.1	*GALNT7*	LRT	0.036	0.036	1.2 (1.0)	0.05	0.03	Beecham, 2014	6.0x10−09	-1.147 (0.198)	Neuritic Plaque
rs316341	6p25	*SERPINB1*	Recessive	0.004	0.015	-1.8 (0.6)	0.68	0.71	Deming, 2017	1.8x10−08	-0.025 (0.004)	CSF Aβ42
rs9271058	6p21.32	*HLA-DRB1*	Dominant	0.017	0.125	3.0 (1.3)	0.76	0.75	Kunkle, 2019	1.4x10−11	0.91 (0.88–0.93)	AD Risk
rs9271192	6p21.32	*HLA-DQA1*	Dominant	0.017	0.125	3.0 (1.3)	0.76	0.75	Lambert, 2013	2.9x10−12	0.90 (0.87–0.93)	AD Risk
rs10808026	7q34–q35	*EPHA1*	Dominant	0.035	0.410	1.3 (0.6)	0.24	0.19	Marioni, 2018	1.1x10−14	0.91 (0.89–0.93)	AD Risk
rs4236673	8p21.1	*CLU*	Dominant	0.017	0.138	-2.3 (1.0)	0.69	0.72	Marioni, 2018	1.1x10−28	1.12 (1.09–1.14)	AD Risk
rs9331896	8p21.1	*CLU*	Dominant	0.043	0.174	-1.9 (0.9)	0.64	0.57	Lambert, 2013	2.8x10−25	1.16 (1.13–1.19)	AD Risk
rs983392	11q12.2	*MS4A6A*	Recessive	0.046	0.331	1.6 (0.8)	0.41	0.27	Lambert, 2013	6.1x10−16	0.90 (0.87–0.92)	AD Risk
rs4985556	16q22.1	*IL34*	Allelic	0.049	0.170	-2.9 (1.5)	0.02	0.08	Marioni, 2018	3.7x10−08	1.09 (1.05–1.12)	AD Risk
rs138190086	17q23.3	*CYB561*	LRT	0.015	0.015	-2.7 (1.2)	0.03	0.01	Marioni, 2018	1.9x10−09	1.25 (1.16–1.35)	AD Risk
rs6069736	20q13.2–31	*CSTF1*	Recessive	0.003	0.178	7.5 (2.5)	0.14	0.13	Marioni, 2018	2.0x10−10	0.89 (0.86–0.93)	AD Risk
rs6014724	20q13.2–31	*CASS4*	Recessive	0.003	0.362	7.5 (2.5)	0.15	0.12	Jansen, 2019	6.6x10−10	0.98 (0.97–0.98)	AD Risk

Table 4.3: Replication of hits with genome-wide significance for AD-associated phenotypes from previous studies in this cohort. All models are from PLINK except LRT (likelihood ratio rest) which is calculated using GEMMA. Only variants with nominal significance (p<0.05) in at least one model and with consistent effect direction are shown, and the model with the lowest p value is shown. For all variants, the LRT p value is also shown, which accounts for family structure using a linear mixed model that adjusts for kinship. $\beta$is shown for each model (with the $\beta$from the Allelic model shown if LRT is the best model, as GEMMA does not output βvalues), and is derived from the basis of age of dementia onset. For previous studies, all effects are OR (95% CI) except Beecham et al., 2014 and Deming et al., 2017, which are β(SE). MAF indicates minor allele frequency in this cohort. PopAF indicates the population allele frequency (TOPMed)

4.3.6 Coding variants of interest

We next asked if any coding variants associate with age of dementia onset (Table 4.4). Two missense variants reached genome-wide significance (p<5x10–8) which were in Linkage Disequilibrium (LD) with one another (r2=1 and D'=1), rs34059820 in *SORBS3* and rs143895597 in *CCAR2*, both associated with a later age of onset. To capture hits less stringently, we allowed a more permissive threshold of p<1x10–5, but required replication, which identified two independent signals: a missense variant in *ZBTB3* that replicates in one previous AD AAO study[172] and was associated with a later age of onset, and a pair of missense variants in *TSPAN10* (rs6565617 and rs7210026) in LD with one another (r2=1 and D'=1) that replicated in both previous AD AAO studies ([173, 172]) and were associated with an earlier age of onset. It is possible that rare variants not imputed by TOPMed are associated, so we assessed variants present in genomes but absent in the imputed dataset with p<0.05 and also in perfect LD (r2=1 and D'=1) with a variant (p<1x10–5) in the same model from the imputation set at a population allele frequency >0.01%, and CADD >20. This analysis revealed one canonical loss-of-function and two missense variants including a variant in *STIM2* (rs1457401458) associated with a later age of onset. Next we searched for variants in 765 genes previously implicated in neurodegeneration with p<0.05, a population allele frequency >0.01%, and CADD >20, which uncovered seven missense variants including variants in *HTT* (rs1473464204), *KCNT1* (rs557219607), and the *APOE* Christchurch mutation (rs121918393)[32], all associated with a later age of onset. Finally, we used permissive filters for eight key genes important for AD including *APP*, *PSEN1*, *PSEN2*, *MAPT*, *APOE*, *ABCA7*, *SORL1*, or *TREM2* with p<0.05, a population allele frequency >1%, and CADD >10, which identified one canonical loss-of-function variant in *ABCA7* (rs547447016) and a missense variant in *APP* (rs112263157), both associated with an earlier age of onset.

Category	Source	Chr	Pos	ID	Gene	Protein Δ	Model	p	β (SE)	MAC (Total)	Hom	PopAF	CADD
GW significant (p<5x10–8)	Imputed	8	22.6	rs34059820	*SORBS3*	Arg550His	Allelic	5.6x10–9	23.0 (3.8)	2 (688)	0	2.3x10–3	8.2
	Imputed	8	22.6	rs143895597	*CCAR2*	Ile298Val	Allelic	5.6x10–9	23.0 (3.8)	2 (688)	0	0.01	12.5
p<1x10-5 with replication	Imputed	11	62.8	rs544641	*ZBTB3*	Ile574Met	Genotypic	5.5x10–6	12.7 (2.7)	22 (688)	1	0.09	11.6
	Imputed	17	81.6	rs6565617	*TSPAN10*	Ala118Thr	Genotypic	4.8x10–7	-2.4 (0.5)	418 (688)	123	0.48	0
	Imputed	17	81.6	rs7210026	*TSPAN10*	Val134Ala	Genotypic	4.8x10–7	-2.4 (0.5)	418 (688)	123	0.48	11.1
LD with imputed variant with p<1x10-5, Pop.AF <0.01%, CADD >20	Genome	4	27.0	rs145740145 8	*STIM2*	Thr550Asn	LRT	6.3x10–3	10.1 (5.7)	3 (156)	0	4.1x10–6	24.6
	Genome	10	32.0	10:32031212:T:A	*KIF5B*	Asn481Ile	Allelic	5.6x10–3	13.0 (4.6)	3 (158)	0	Absent	26.3
	Genome	19	45.5	19:45498605:C:CT	*PPM1N*	Leu46fs	Genotypic	1.4x10–3	11.8 (3.5)	6 (160)	1	Absent	22.9
Neurodegen. Association, Pop. AF <0.01%, CADD >20	Imputed	1	11.8	rs149322151	*CLCN6*	Glu779Lys	Allelic	9.8x10–3	10.2 (3.9)	2 (688)	0	1.1x10–5	28.2
	Imputed	4	3.2	rs1473464204	*HTT*	Thr1253Met	Allelic	5.4x10–3	5.9 (2.1)	9 (688)	0	4.0x10–5	22.9
	Imputed	9	77.3	rs572761766	*VPS13A*	Asn2162His	Allelic	0.016	4.2 (1.7)	11 (688)	0	1.1x10–5	23.7
	Imputed	9	135.8	rs557219607	*KCNT1*	Arg573Cys	Allelic	3.7x10–5	16.2 (3.9)	2 (688)	0	4.4x10–5	25.4
	Imputed	16	16.2	rs200800189	*ABCC6*	Val711Met	Allelic	0.01	4.6 (1.8)	10 (688)	0	7.7x10–5	25
	Imputed	19	44.9	rs121918393	*APOE*	Arg154Ser	Genotypic	3.2x10–6	12.8 (2.7)	13 (674)	1	8.0x10–6	31
	Genome	22	39.2	rs190122588	*PDGFB*	Arg66His	Genotypic	0.03	-6.7 (4.1)	4 (160)	0	1.8x10–5	24.4
Key AD gene, Pop. AF <1%, CADD >10	Imputed	19	1	rs547447016	*ABCA7*	Glu709fs	Allelic	0.017	-4.6 (1.9)	9 (688)	0	2.8x10–3	35
	Imputed	21	25.9	rs112263157	*APP*	Ser614Gly	LRT	7.1x10–4	-4.2 (1.5)	15 (688)	0	4.7x10–3	18.5

Table 4.4: Coding variants of interest. Variants were required to meet one of five categories, all of which required observation of alleles in at least two individuals: (1) genome-wide (GW) significant (p<5x10–8), (2) p<1x10–5 with replication in at least one of the aforementioned replication cohorts, (3) a variant observed in the genome only analysis (but not observed in the imputation set) with p<0.05 in perfect Linkage Disequilibrium (LD) (r2=1 and D'=1) with a variant with p<1x10–5 with the same model in the imputation set, a population allele frequency >0.01%, and CADD >20, (4) A variant in a gene previously implicated in neurodegeneration with p<0.05, a population allele frequency >0.01%, and CADD >20, or (5) A variant in APP, *PSEN1*, *PSEN2*, *MAPT*, *APOE*, *ABCA7*, *SORL1*, or *TREM2* with p<0.05, a population allele frequency >1%, and CADD >10. Source indicates if the analysis was on the imputation plus genome call set, or genome only set (implying a rare variant not imputed by TOPMed). Pos indicates hg38 position in megabases. MAC (Total) indicates minor allele count and total allele count. PopAF is the maximum population allele frequency from four databases: ExAC non-psychiatric disease, gnomAD exomes, gnomAD genomes, and TOPMed Bravo genomes.

4.4 Discussion

Genetic association studies for LOAD are limited by heterogeneity of cases and unknown levels of contribution from environmental sources. This study addresses these limitations by employing a well-described phenotype in a geographically isolated population with a monogenic form of AD [25]. While environmental influences will always be present, this population has a relatively homogeneous set of environmental influences. We identified 29 loci reaching genome-wide significance, 27 of which were associated with a later age of onset (median β =17.6 years), implying they may be protective. Two loci directly replicated in previously conducted AD-relevant GWAS studies, and 23 of the 29 loci had a neurodegeneration-associated GWAS hits within 500 kb, indicating the general relevance of identified loci for neurodegeneration. These observations will serve as the foundation for additional studies to nominate particularly important pathways. For example, one variant that exhibited replication, rs35980966, is a rare variant that tags the *CLU* locus on chromosome 8, a locus for which our cohort also provides replication for two common variants identified as genome-wide significant index variants in previous studies, rs9331896[12] and rs4236673[178]. In addition, rs138295139 on chromosome 1 is only 4.4kb from a variant previously associated with plasma clusterin, rs4428865[179]. Taken together, these observations suggest that further investigation on the role of clusterin in EOAD is warranted. Only two coding variants reached genome-wide significance, including a variant in *SORBS3* (rs34059820). A previous study for age of onset modifiers for a different *PSEN1* mutation implicated a related gene, *SORBS2*[180]. In addition to genome-wide significant loci, we investigated previously implicated GWAS loci. We confirmed the modest effect of variants tagging *APOE* E2 and of *APOE* E4 in this population, consistent with either modest or undetectable effects of these alleles for ADAD in other studies[156, 175, 176]. We also observed replication of 13 index variants from

10 previously identified loci linked to AD in this cohort. In addition to the variants at the *CLU* locus that we previously discussed, another previous GWAS hit of interest that replicates in this study was a stop gain variant in *IL34* [178] associated with earlier age of dementia onset. IL-34 has protective roles such as protection from excitotoxicity [181]) and promoting clearance of oligomeric Aβ[182], so it is consistent that a stop gain variant in *IL34* would associate with an earlier age of onset by preventing these protective effects. Beyond currently or previously implicated genome-wide significant loci, we conducted genome-wide searches for coding variation using more permissive criteria. Two variants in LD in *TSPAN10* were nominally (p=4.8x10–7) associated with an earlier age of dementia onset in this cohort, and this association replicated in both AD AAO studies queried[173, 172]. Tetraspanin-10 is a part of the TspanC8 subgroup of tetraspanins which promote *ADAM10* maturation[183]. Given *ADAM10*'s established role as an α-secretase promoting non-amyloidogenic processing of Aβ[184] and the recent association of genetic variation in or near *ADAM10* with AD risk by GWAS [174, 178], the basis of the observed association between earlier age of dementia onset and these variants in *TSPAN10* may result from disruption of a protective role of *ADAM10*. An additional coding variant of interest was a rare missense variant in *STIM2* was nominally associated with later age of dementia onset (LRT model p=6.3x10–3 in genome-only analysis and in perfect LD (r2=1 and D'=1) with a variant with p=9.0x10-6 in the imputation analysis). *STIM2* is protective in AD models[185], implying that this variant may be gain of function (or could block *STIM2* loss of function). In addition to variants identified through more permissive genome-wide criteria, we also explored variation in genes previously associated with neurodegeneration. One example was a variant in *HTT*. Repeat expansions in *HTT* are associated with Huntington's disease[186], but a recent study has implicated huntingtin in regulation of presynaptic APP levels [187], which may explain the observation of a rare missense variant in *HTT* associated with later age

of dementia onset. A second example was a variant in *KCNT1*. Variants in *KCNT1* can cause seizures[188] and were also identified in the discovery phase of an ALS rare variant study[189]. Therefore, the observation of a missense variant in *KCNT1* with an association with later age of dementia onset (p=3.7x10–5) in the genome-only analysis is interesting in light of the growing evidence for excitability abnormalities in AD and AD models[190, 191], including abnormalities in other potassium channels[192]. As the largest age of onset modifier study in dominant AD to date (to our knowledge), this study has nominated several new candidate genetic associations with age of dementia onset in dominant AD. However, we acknowledge that this study is limited by the small size of the dominant AD population. Recruitment of more patients with early onset dementias from South American countries will help to overcome this limitation in future studies. An important overarching theme from this analysis is that while age at dementia onset in ADAD has a strong heritable component, it is likely that, as with LOAD, there are many different genetic contributors that sum to determine an individual's age at dementia onset for ADAD. Based on the unique demography of this population as a tri-continental admixture that passed through a narrow bottleneck [193], we conducted this study with the hypothesis that rare variants with a large effect size, i.e., the *APOE* Christchurch mutation[32], could account for much of the difference in age at dementia onset. Indeed, we identified many genetic variants of a similar rarity and damage level and with similar effect sizes and association significance levels to *APOE* Christchurch in this study (Table 4.4). For example, *SORBS3* (rs34059820) exhibited a similar effect size even in the heterozygous state, and several others such as variants discussed in *STIM2* (rs1457401458) and *KCNT1* (rs557219607) occurred in the heterozygous state but could have a similar effect size to *APOE* Christchurch if they occurred in the homozygous state, which was observed for *ZBTB3* (rs544641) and *PPM1N* (hg38 19:45498605:C>CT). We note that due to the nature of the models employed, it is possible for the presence of

alleles in a small number of individuals with a particularly late age of onset can result in a low p value. However, as a class these variants could confer a larger effect. Nevertheless, observations driven by a small number of alleles should be considered with caution. Importantly, we also detected more common and/or lower effect size variation associated with age of dementia onset. Some of the variants identified are associated with the amyloid processing pathway, others with pathways that include immune response and neuronal signaling. Because many of these variants replicate or were identified in non-admixed European populations, it suggests that the associations for many of these variants are robust to ancestral background. The identified variants in this study occur in the presence of a very strong causative mutation for ADAD, emphasizing the importance of the association signals observed for these variants and the need for more investigation of these variants in future studies.

4.5 Permissions and Attributions

The content of Chapter 4 is the result of a collaboration with Dr. J. Nicholas Cochran at HudsonAlpha Institute of Biotechnology.

Cochran JN∗, Acosta-Uribe J∗ , Madrigal L, Aguillón D, García GP, Giraldo MM, Acosta-Baena N, Piedrahita F, Alzate D, López HE, Roberts K, Absher D, Myers RM, Lopera F, Kosik KS. Genetic Associations with Age at Dementia Onset in the PSEN1 E280A Colombian Kindred. Pre-print medRxiv 2020 September 25
doi:https://doi.org/10.1101/2020.09.23.20198424
∗Co-first authors

4.6 Supplementary Material

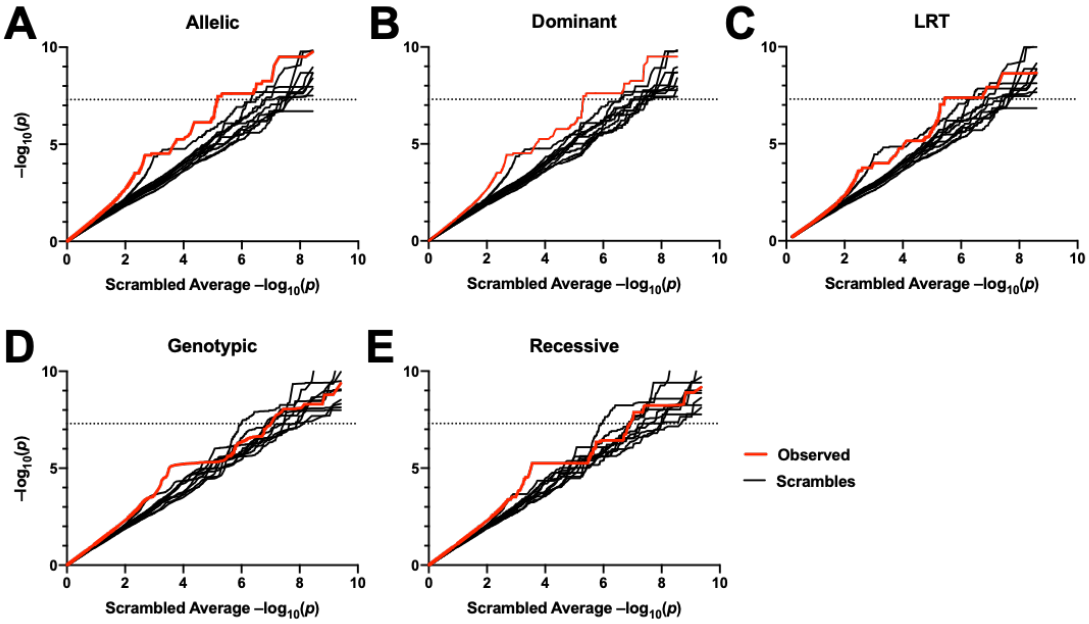


Figure 4.4: QQ plots. Observed (actual age of dementia onset) –log10(p) values are plotted vs. the average of scrambled age of dementia onset –log10(p) values. All models are from PLINK except LRT (likelihood ratio rest) which is calculated using GEMMA.

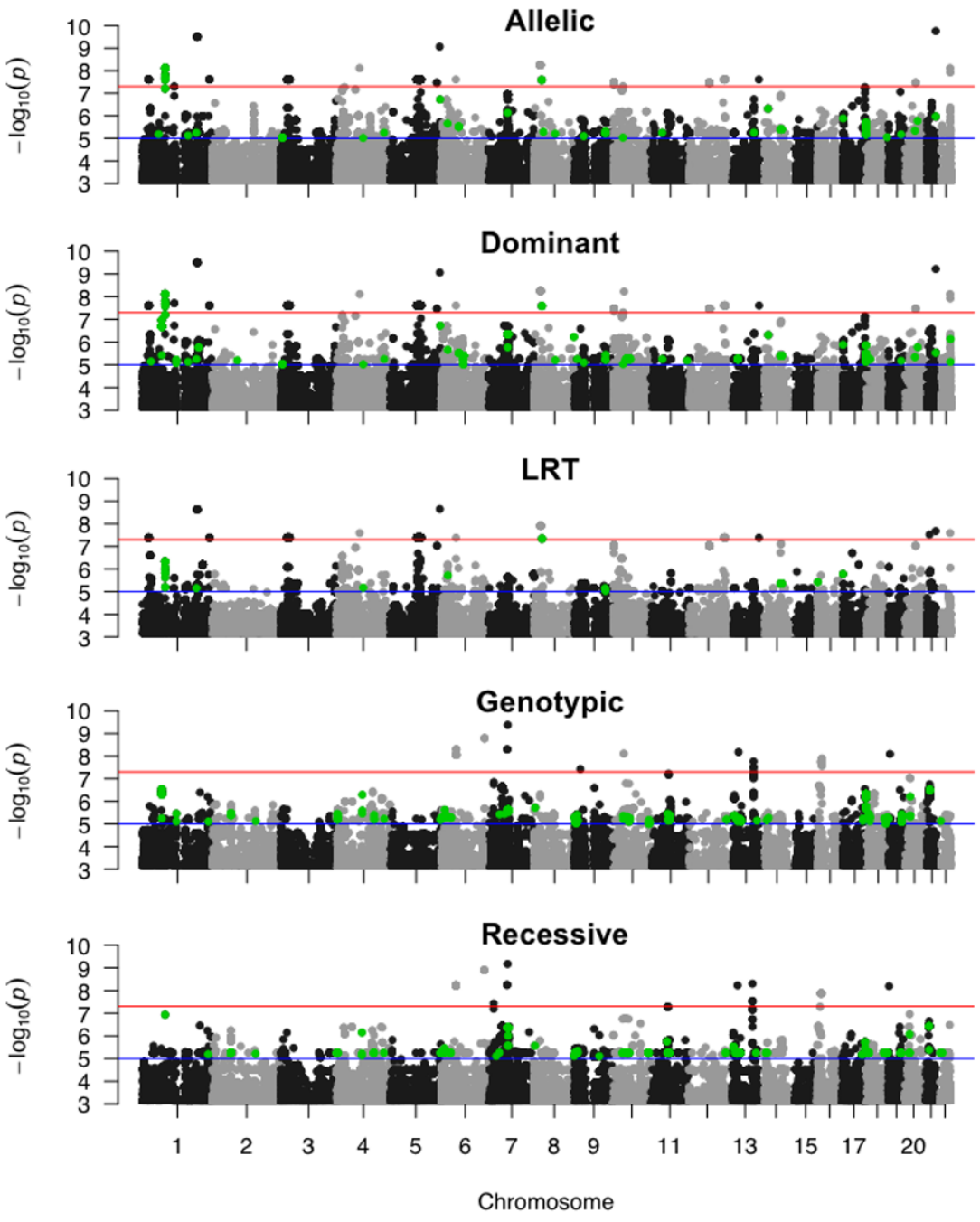


Figure 4.5: Manhattan plots for all models. Any variant with a discovery p¡1x10–5 that exhibited nominal replication (p¡0.05) is highlighted in green. Blue line indicates p¡1x10–5, red line indicates p¡5x10–8.

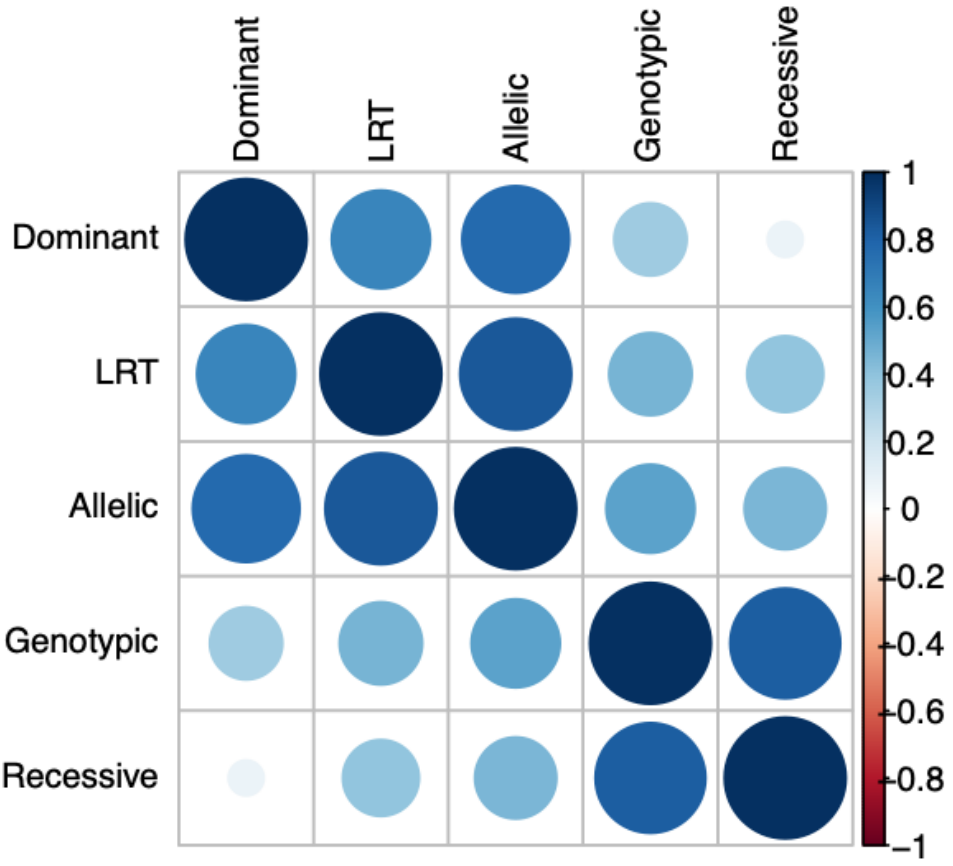


Figure 4.6: Pearson cross-correlations of –log10(p) values for the five models employed. All models are from PLINK except LRT (likelihood ratio rest) which is calculated using GEMMA.

Chapter 5

Concluding Remarks

5.1 Summary

The work presented in this dissertation demonstrates how studying a population with a complex demographic history expands our knowledge of the genetic basis of neurodegenerative disorders.

In Chapter 2 we use different bioinformatic methods to analyze the ancestral background and admixture patterns of a large cohort of patients with neurodegenerative illnesses and healthy individuals (TANGL cohort). This cohort exhibited the patterns of tri-continental ancestry and sex biased admixture that had been previously described in the country. We identified multiple families with mendelian forms of dementia, most of them associated with missense *PSEN1* variants. All of the carriers of disease-causing alleles were identical by descent, suggesting that these families likely arose as consequences of founder effects. Additionally, we observed the confluence of disease-causing and risk-conferring variants from all three ancestral populations. *TREM2* heterozygous

and homozygous variants were the most common among Alzheimer risk genes, a point of interest because the disease risk conferred by these variants differed according to ancestry. Furthermore, several gene variants that have a known association with MND in European populations had FTLD phenotypes on a Native American haplotype. Our results suggest that the serial bottlenecks of the Native American populations prior to the Spanish conquest, followed by the admixture process and additional bottlenecks created a unique setting for founder effects and large drift to alter the frequency of deleterious variants from multiple ancestral origins and create a unique genetic burden in the Colombian population.

In Chapter 3 we describe a large family with a novel pathogenic *PSEN1* variant of African origin. Genetic analysis revealed the missense variant *PSEN1* NM_000021 c.1247T>C p.Ile416Thr, which originated on an African haplotype and segregated with Alzheimer's disease [logarithm of the odds score of 6]. The clinical phenotype of the Ile416Thr carriers is similar to sporadic Alzheimer's disease, except for earlier age at onset. The mean age at onset for mild cognitive impairment was 47.6 years (standard deviation 5.83) and for dementia 51.6 years (standard deviation 5.03). PET-amyloid and PET-Tau scans of Ile416Thr carriers showed similar patterns to those seen in other carriers of *PSEN1* pathogenic variants, evident of increased amyloid pathology years before the clinical onset of the illness. The well-defined age at onset and progression of the disease makes this family a great candidate for preventive clinical trials.

Lastly, in Chapter 4 we performed a genetic association study for dementia age at onset in 344 carriers of the *PSEN1* NM_000021 c.839A¿C p.Glu280Ala variant. The proportion of variance explained (+/- 95% CI) was 56% (+/-15%), and was due to genetic variation at numerous loci. 29 loci reached genome-wide significance, of which 23 had a GWAS hit from the NHGRI-EBI catalog within 500 kb of loci relevant to neurodegeneration. Hits included a new variant in the *CLU* locus, which replicated a

variant in a locus on a separate chromosome previously associated with plasma clusterin, and a missense variant in *SORBS3*. Former GWAS index variants at the *CLU* locus were also replicated in this cohort. Overall, these results suggest that clusterin, which is involved lipid transport and immune modulation, plays a significant role in the AAO for dementia in the Glu280Ala Kindred. *APOE* E4 (rs429358) and APOE E2 (rs7412) associated variants exhibited modest (less than 3 years) associations with earlier and later age of dementia onset, respectively. Further analyses will have to be performed in other families with autosomal dominant Alzheimer's disease to confirm this association.

Overall, these results demonstrate how by studying this unique population we have discovered novel genetic variants that cause mendelian forms of neurodegenerative illnesses and how the genetic background can modify the phenotype of a genetic illness, either changing its clinical manifestations, or the age at onset for the disease.

5.2 Outlook

As we include more diverse populations in large genomic studies, we start unveiling novel genetic variation of the human genome, even to thresholds that ten years ago were not imaginable. The large number of variants that are being identified are far more numerous than the clinically labeled phenotypes with which they are associated, and therefore, they can collectively provide novel insights regarding convergent molecular pathways that define a phenotype. Throughout this dissertation, we attributed the large number of deleterious variants to neutral genetic processes like admixture and drift. However, the interplay between our findings with the ancient and recent demographic history of the country raises the possibility of adaptative genetic processes and positive selection shaping the genomes of the current population. Given the intricate relationship between fitness and infection survival, pathogens have been one of the main drivers

of positive selection. There are many examples of genetic variants that confer protection against certain types of pathogens, but are disease-causing in other conditions. A widely known example is the sickle-cell variant in the beta hemoglobin gene (HbS allele), which in heterozygous states confer protection against Malaria and in homozygous state leads to sickle-cell anemia [194]. The large number of *PSEN1* variants and carriers of these pathogenic alleles leads us to hypothesize that these variants could have undergone positive selection, as well. *PSEN1* variants increase the production of amyloid-β, which has been demonstrated to have anti-microbial activity against bacterial and viral infections[141]. We postulate that after the initial contact between the Native American, European and African populations, infectious diseases created an intense selective pressure that favored carriers of *PSEN1* variants to survive, and demographic events that took place 500 years ago shaped the current epidemiology of neurodegenerative illnesses. As a concluding remark, this dissertation reinforces that the combination of medical and evolutionary approaches to the understanding of genetics underlying disease may shed light into fundamental pathways and identification of potential therapeutic targets.

Bibliography

[1] Alzheimer's Association. 2019 Alzheimer's Disease Facts and Figures. dAlzheimer's Association; 2019.

[2] Sierksma A, Escott-Price V, De Strooper B. Translating genetic risk of Alzheimer's disease into mechanistic insight and drug targets. Science. 2020 oct;370(6512):61-6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/33004512.

[3] Bekris LM, Yu CE, Bird TD, Tsuang DW. Review article: Genetics of Alzheimer disease; 2010. Available from: http://journals.sagepub.com/doi/pdf/10.1177/0891988710383571.

[4] Manly JJ, Mayeux R. Critical Perspectives on Racial and Ethnic Differences in Health in Late Life. Anderson NB, Bulatao RA, Cohen B, editors. Washington, D.C.: National Academies Press; 2004. Available from: https://www.ncbi.nlm.nih.gov/books/NBK25535/http://www.nap.edu/catalog/11086.

[5] Steenland K, Goldstein FC, Levey A, Wharton W. A Meta-Analysis of Alzheimer's Disease Incidence and Prevalence Comparing African-Americans and Caucasians. Journal of Alzheimer's Disease. 2016 jan;50(1):71-6.

[6] Mayeda ER, Glymour MM, Quesenberry CP, Whitmer RA. Inequalities in dementia incidence between six racial and ethnic groups over 14 years. Alzheimer's & Dementia. 2016 mar;12(3):216-24. Available from: http://doi.wiley.com/10.1016/j.jalz.2015.12.007.

[7] Popejoy AB, Fullerton SM. Genomics is failing on diversity. Nature. 2016 oct;538(7624):161-4. Available from: http://www.nature.com/doifinder/10.1038/538161a.

[8] Buniello A, Macarthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Research. 2019 jan;47(D1):D1005-12. Available from: https://academic.oup.com/nar/article/47/D1/D1005/5184712.

[9] Wojcik GL, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, et al. Genetic analyses of diverse populations improves discovery for complex traits. Nature. 2019 jun;570(7762):514-8. Available from: https://doi.org/10.1038/s41586-019-1310-4.

[10] Zheng H, Koo EH. Biology and pathophysiology of the amyloid precursor protein. Molecular Neurodegeneration. 2011 apr;6(1):27. Available from: http://molecularneurodegeneration.biomedcentral.com/articles/10.1186/1750-1326-6-27.

[11] Karch CM, Cruchaga C, Goate AM. Alzheimer's disease genetics: From the bench to the clinic. Cell Press; 2014. Available from: https://www.sciencedirect.com/science/article/pii/S0896627314004851.

[12] 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 [Journal Article]. Nat Genet. 2013 dec;45(12):1452-8. Available from: http://www.nature.com/articles/ng.2802https://www.ncbi.nlm.nih.gov/pubmed/24162737.

[13] Cruchaga C, Chakraverty S, Mayo K, Vallania FLM, Mitra RD, Faber K, et al. Rare variants in APP, PSEN1 and PSEN2 increase risk for AD in late-onset Alzheimer's disease families. PLoS ONE. 2012 feb;7(2):e31039. Available from: http://dx.plos.org/10.1371/journal.pone.0031039.

[14] Deming Y, Li Z, Kapoor M, Harari O, Del-Aguila JL, Black K, et al. Genome-wide association study identifies four novel loci associated with Alzheimer's endophenotypes and disease modifiers [Journal Article]. Acta Neuropathol. 2017;133(5):839-56. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28247064.

[15] Ferrari R, Wang Y, Vandrovcova J, Guelfi S, Witeolar A, Karch CM, et al. Genetic architecture of sporadic frontotemporal dementia and overlap with Alzheimer's and Parkinson's diseases. Journal of Neurology, Neurosurgery and Psychiatry. 2017 feb;88(2):152-64. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27899424http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5237405.

[16] Ahmad L, Zhang SY, Casanova JL, Sancho-Shimizu V. Human TBK1: A Gatekeeper of Neuroinflammation. Elsevier; 2016. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27211305http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4890605.

[17] Gorno-Tempini ML, Hillis AE, Weintraub S, Kertesz A, Mendez M, Cappa SF, et al. Classification of primary progressive aphasia and its variants. Neurology. 2011 mar;76(11):1006-14. Available from:

http://www.ncbi.nlm.nih.gov/pubmed/21325651http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3059138.

[18] Woollacott IOC, Rohrer JD. The clinical spectrum of sporadic and familial forms of frontotemporal dementia. Journal of Neurochemistry. 2016 aug;138:6-31. Available from: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jnc.13654https://onlinelibrary.wiley.com/doi/10.1111/jnc.13654.

[19] Pottier C, Ravenscroft TA, Sanchez-Contreras M, Rademakers R. Genetics of FTLD: overview and what else we can expect from genetic studies. Journal of neurochemistry. 2016;138 Suppl:32-53. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27009575.

[20] Blauwendraat C, Wilke C, Simón-Sánchez J, Jansen IE, Reifschneider A, Capell A, et al. The wide genetic landscape of clinical frontotemporal dementia: Systematic combined sequencing of 121 consecutive subjects. Genetics in Medicine. 2018 feb;20(2):240-9. Available from: http://www.nature.com/doifinder/10.1038/gim.2017.102.

[21] Guerreiro R, Gibbons E, Tábuas-Pereira M, Kun-Rodrigues C, Santo GC, Bras J. Genetic architecture of common non-Alzheimer's disease dementias. Academic Press Inc.; 2020.

[22] Giraldo M, Lopera F, Siniard AL, Corneveaux JJ, Schrauwen I, Carvajal J, et al. Variants in triggering receptor expressed on myeloid cells 2 are associated with both behavioral variant frontotemporal lobar degeneration and Alzheimer's disease. Neurobiology of Aging. 2013 aug;34(8):2077.e11-2077.e18. Available from: http://linkinghub.elsevier.com/retrieve/pii/S019745801300095X.

[23] Ruiz A, Dols-Icardo O, Bullido MJ, Pastor P, Rodríguez-Rodríguez E, López de Munain A, et al. Assessing the role of the TREM2 p.R47H variant as a risk factor for Alzheimer's disease and frontotemporal dementia. Neurobiology of Aging. 2014 feb;35(2):444.e1-444.e4. Available from: http://linkinghub.elsevier.com/retrieve/pii/S019745801300345X.

[24] Thelen M, Razquin C, Hernández I, Gorostidi A, Sánchez-Valle R, Ortega-Cubero S, et al. Investigation of the role of rare TREM2 variants in frontotemporal dementia subtypes. Neurobiology of Aging. 2014 nov;35(11):2657.e13-9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25042114.

[25] Lopera F, Ardilla A, Martinez A, Madrigal L, Arango-Viana J, Lemere C, et al. Clinical features of early-onset Alzheimer disease in a large kindred with an E280A presenilis-1 mutation. American Journal of Ophthalmology. 1997 jul;124(1):137-8. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0002939414716770.

[26] Lalli MA, Cox HC, Arcila ML, Cadavid L, Moreno S, Garcia G, et al. Origin of the PSEN1 E280A mutation causing early-onset Alzheimer's disease. Alzheimer's and Dementia. 2014 oct;10(5):S277-83. Available from: https://www.sciencedirect.com/science/article/pii/S1552526013028409?via{%}3Dihub.

[27] Rios-Romenets S, Lopez H, Lopez L, Hincapie L, Saldarriaga A, Madrigal L, et al. The Colombian Alzheimer's Prevention Initiative (API) Registry. Alzheimer's & Dementia. 2017 may;13(5):602-5. Available from: http://doi.wiley.com/10.1016/j.jalz.2016.09.010.

[28] Acosta-Baena N, Sepulveda-Falla D, Lopera-Gomez CM, Jaramillo-Elorza MC, Moreno S, Aguirre-Acevedo DC, et al. Pre-dementia clinical stages in presenilin 1 E280A familial early-onset Alzheimer's disease: A retrospective cohort study [Journal Article]. The Lancet Neurology. 2011 mar;10(3):213-20. Available from: https://www.sciencedirect.com/science/article/pii/S1474442210703239https://www.ncbi.nlm.nih.gov/pubmed/21296022.

[29] Fleisher AS, Chen K, Quiroz YT, Jakimovich LJ, Gomez MG, Langois CM, et al. Florbetapir PET analysis of amyloid-β deposition in the presenilin 1 E280A autosomal dominant Alzheimer's disease kindred: a cross-sectional study. The Lancet Neurology. 2012 dec;11(12):1057-65. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23137949.

[30] Quiroz YT, Sperling RA, Norton DJ, Baena A, Arboleda-Velasquez JF, Cosio D, et al. Association Between Amyloid and Tau Accumulation in Young Adults With Autosomal Dominant Alzheimer Disease. JAMA Neurology. 2018 feb. Available from: http://archneur.jamanetwork.com/article.aspx?doi=10.1001/jamaneurol.2017.4907.

[31] Velez JI, Lopera F, Sepulveda-Falla D, Patel HR, Johar AS, Chuah A, et al. APOE*E2 allele delays age of onset in PSEN1 E280A Alzheimer's disease [Journal Article]. Mol Psychiatry. 2016;21(7):916-24. Available from: https://www.ncbi.nlm.nih.gov/pubmed/26619808.

[32] Arboleda-Velasquez JF, Lopera F, O'Hare M, Delgado-Tirado S, Marino C, Chmielewska N, et al. Resistance to autosomal dominant Alzheimer's disease in an APOE3 Christchurch homozygote: a case report. Nature Medicine. 2019 nov;25(11):1680-3. Available from: https://doi.org/10.1038/s41591-019-0611-3.

[33] Tovar Pinzon H, Tovar Mora JA, Tovar Mora CE. Convocatoria al poder del numero: Censos y estadisticas de la Nueva Granada, 1750-1830. Tovar Pinzón H, editor. 1. Bogota, Colombia.: Archivo General de la Nación, 1994; 1995. Available from: https://franklin.library.upenn.edu/catalog/FRANKLIN{_}9927593473503681.

[34] Reichel-Dolmatoff G. Arqueología de Colombia: un texto introductorio. Santa Fe de Bogota: Presidencia de la República de Colombia.; 2004. Available from: http://babel.banrepcultural.org/cdm/ref/collection/p17054coll9/id/18.

[35] Langebaek CH. ¿Cuántos eran? ¿Dónde estaban? ¿Qué les pasó? Poblamiento indígena en la Colombia prehispánica y su transformación después de la Conquista. In: Bonnett Vélez D, LaRosa M, Nieto M, editors. Colombia, preguntas y respuestas sobre su pasado y su presente. 1st ed. Universidad de los Andes, Colombia; 2010. p. 27-52.

[36] Wheat D. The first great waves: African provenance zones for the transatlantic slave trade to Cartagena de Indias, 1570-1640. Journal of African History. 2011 apr;52(1):1-22.

[37] Maya Restrepo LA. Demografía histórica de la trata por Cartagena, 1533-1810. In: Geografía humana de Colombia: los afrocolombianos - Tomo VI. Bogota, Colombia: Instituto Colombiano de Antropología e Historia - ICANH; 1998. p. 3-41.

[38] Navarrete MC. Génesis y desarrollo de la esclavitud en Colombia, siglos: XVI y XVII. 1st ed. Cali, Colombia: Programa Editorial, Universidad del Valle; 2005.

[39] Borucki A, Eltis D, Wheat D. Atlantic history and the slave trade to Spanish America. Oxford Academic; 2015.

[40] Wheat D. Atlantic Africa and the Spanish caribbean, 1570-1640. 1st ed. University of North Carolina Press; 2016.

[41] Tovar Pinzon H. Las cifras y los métodos en la reconstrucción de la población colombiana. In: Meisel A, Ramírez MT, editors. La economía colonial de la Nueva Granada. primera ed ed. Sección de obras de economía. Bogota: Banco de la República de Colombia; 2015. .

[42] O'Fallon BD, Fehren-Schmitz L. Native Americans experienced a strong population bottleneck coincident with European contact. Proceedings of the National Academy of Sciences of the United States of America. 2011 dec;108(51):20444-8. Available from: https://www.pnas.org/content/108/51/20444https://www.pnas.org/content/108/51/20444.abstract.

[43] Browning SR, Browning BL, Daviglus ML, Durazo-Arvizu RA, Schneiderman N, Kaplan RC, et al. Ancestry-specific recent effective population size in the Americas. PLoS Genetics. 2018 may;14(5):e1007385. Available from: https://dx.plos.org/10.1371/journal.pgen.1007385.

[44] Gravel S. When Is Selection Effective? Genetics. 2016 may;203(1):451-62. Available from: /pmc/articles/PMC4858791//pmc/articles/PMC4858791/?report=

abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4858791/https://academic.oup.com/genetics/article/203/1/451/5930316http://www.ncbi.nlm.nih.gov/pubmed/27010021http://www.pubmedcentral.nih.

[45] Ossa H, Aquino J, Pereira R, Ibarra A, Ossa RH, Pérez LA, et al. Outlining the Ancestry Landscape of Colombian Admixed Populations. PLOSONE. 2016 oct;11(10):e0164414. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27736937http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5063461http://dx.plos.org/10.1371/journal.pone.0164414.

[46] Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, et al. Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. American Journal of Human Genetics. 2017 apr;100(4):635-49. Available from: https://www.sciencedirect.com/science/article/pii/S0002929717301076.

[47] Abel O, Powell JF, Andersen PM, Al-Chalabi A. ALSoD: A user-friendly online bioinformatics tool for amyotrophic lateral sclerosis genetics. Human Mutation. 2012 sep;33(9):1345-51.

[48] Gravel S, Henn BM, Gutenkunst RN, Indap AR, Marth GT, Clark AG, et al. Demographic history and rare allele sharing among human populations. Proceedings of the National Academy of Sciences of the United States of America. 2011 jul;108(29):11983-8. Available from: www.pnas.org/cgi/doi/10.1073/pnas.1019276108.

[49] Clark AG, Slatkin M, Ramos EM, Guttmacher AE, Mackay TFC, McCarthy MI, et al. Finding the missing heritability of complex diseases. Nature. 2009 oct;461(7265):747-53. Available from: http://www.nature.com/doifinder/10.1038/nature08494.

[50] Aguirre-Acevedo DC, Gómez RD, Moreno S, Henao-Arboleda E, Motta M, Muñoz C, et al. Validez y fiabilidad de la batería neuropsicológica CERAD-Col. Revista de Neurologia. 2007;45(11):655-60.

[51] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's and Dementia. 2011 may;7(3):263-9. Available from: https://www.sciencedirect.com/science/article/pii/S1552526011001014?via{%}3Dihub.

[52] Rascovsky K, Hodges JR, Knopman D, Mendez MF, Kramer JH, Neuhaus J, et al. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. Brain. 2011 sep;134(9):2456-77. Available from: https://academic.oup.com/brain/article-lookup/doi/10.1093/brain/awr179.

[53] Höglinger GU, Respondek G, Stamelou M, Kurz C, Josephs KA, Lang AE, et al. Clinical diagnosis of progressive supranuclear palsy: The movement disorder society criteria. Movement Disorders. 2017 jun;32(6):853-64. Available from: http://doi.wiley.com/10.1002/mds.26987.

[54] Strong MJ, Abrahams S, Goldstein LH, Woolley S, Mclaughlin P, Snowden J, et al. Amyotrophic lateral sclerosis - frontotemporal spectrum disorder (ALS-FTSD): Revised diagnostic criteria. Amyotrophic lateral sclerosis & frontotemporal degeneration. 2017 apr;18(3-4):153-74. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28054827.

[55] Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: Fast processing of NGS alignment formats. Bioinformatics. 2015 jun;31(12):2032-4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25697820http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4765878.

[56] McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research. 2010 sep;20(9):1297-303. Available from: https://genome.cshlp.org/content/20/9/1297.fullhttps://genome.cshlp.org/content/20/9/1297https://genome.cshlp.org/content/20/9/1297.abstract.

[57] Cingolani P, Platts A, Wang LLL, Coon M, Nguyen T, Wang LLL, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3 [Journal Article]. Fly. 2012;6(2):80-92.

[58] Kircher M, Witten DM, Jain P, O'roak BJ, Cooper GM, Shendure J, et al. A general framework for estimating the relative pathogenicity of human genetic variants. Nature Genetics. 2014 mar;46(3):310-5. Available from: http://www.nature.com/articles/ng.2892.

[59] Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012 nov;491(7422):56-65. Available from: http://www.nature.com/articles/nature11632.

[60] Zhao H, Sun Z, Wang J, Huang H, Kocher JPP, Wang L. CrossMap: a versatile tool for coordinate conversion between genome assemblies [Journal Article]. Bioinformatics. 2013 apr;30(7):1006-7.

[61] Liu X, White S, Peng B, Johnson AD, Brody JA, Li AH, et al. WGSA: An annotation pipeline for human genome sequencing studies. Journal of Medical

Genetics. 2015 sep;53(2):111-2. Available from: https://www.ncbi.nlm.nih.gov/pubmed/26395054.

[62] Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016 aug;536(7616):285-91. Available from: http://www.nature.com/articles/nature19057.

[63] Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. bioRxiv. 2019:531210. Available from: https://www.biorxiv.org/content/10.1101/531210v3.

[64] Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic acids research. 2001 jan;29(1):308-11. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11125122.

[65] Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011 aug;27(15):2156-8.

[66] Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. Bioinformatics. 2010 nov;26(22):2867-73. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20926424http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3025716.

[67] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics. 2007 sep;81(3):559-75. Available from: https://www.sciencedirect.com/science/article/pii/S0002929707613524http://linkinghub.elsevier.com/retrieve/pii/S0002929707613524.

[68] Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: Rising to the challenge of larger and richer datasets. GigaScience. 2015 feb;4(1).

[69] Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. Nature Protocols. 2010 aug;5(9):1564-73. Available from: https://www.nature.com/articles/nprot.2010.116.

[70] Borrego-Écija S, Antonell A, Puig-Butillé JA, Pericot I, Prat-Bravo C, Abellan-Vidal MT, et al. Novel P397S MAPT variant associated with late onset and slow progressive frontotemporal dementia. Annals of Clinical and Translational Neurology. 2019;6(8):1559-65.

[71] Moreno-Estrada A, Gravel S, Zakharia F, McCauley JL, Byrnes JK, Gignoux CR, et al. Reconstructing the Population Genetic History of the Caribbean. PLoS Genetics. 2013 nov;9(11):e1003925. Available from: http://dx.plos.org/10.1371/journal.pgen.1003925.

[72] Mao X, Bigham AW, Mei R, Gutierrez G, Weiss KM, Brutsaert TD, et al. A genomewide admixture mapping panel for hispanic/latino populations. American Journal of Human Genetics. 2007;80(6):1171-8.

[73] Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Research. 2009 sep;19(9):1655-64. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19648217http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2752134.

[74] Weissensteiner H, Pacher D, Kloss-Brandstätter A, Forer L, Specht G, Bandelt HJ, et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic acids research. 2016 jul;44(W1):W58-63. Available from: http://haplogrep.uibk.ac.at.

[75] Poznik GD. Identifying Y-chromosome haplogroups in arbitrarily large samples of sequenced or genotyped men. bioRxiv. 2016 nov:088716. Available from: https://doi.org/10.1101/088716.

[76] Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. Nature Methods. 2012 feb;9(2):179-81.

[77] O'Connell J, Gurdasani D, Delaneau O, Pirastu N, Ulivi S, Cocca M, et al. A General Approach for Haplotype Phasing across the Full Spectrum of Relatedness. PLoS Genetics. 2014 apr;10(4):e1004234. Available from: https://dx.plos.org/10.1371/journal.pgen.1004234.

[78] Maples BK, Gravel S, Kenny EE, Bustamante CD. RFMix: A discriminative modeling approach for rapid and robust local-ancestry inference. American Journal of Human Genetics. 2013 aug;93(2):278-88. Available from: https://www.sciencedirect.com/science/article/pii/S0002929713002899{#}fig3.

[79] Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics. 2006 aug;38(8):904-9.

[80] Novembre J, Williams R, Hossein P, Wang Y. PCAviz: Visualizing Principal Components Analysis. R package version 0.3-37; 2019. Available from: http://github.com/NovembreLab/PCAviz.

[81] Allot A, Peng Y, Wei CHH, Lee K, Phan L, Lu Z. LitVar: a semantic search engine for linking genomic variant data in PubMed and PMC. Nucleic Acids Research. 2018 may;46(W1):W530-6. Available from: https://pubmed.ncbi.nlm.nih.gov/29762787/https://doi.org/10.1093/nar/gky355.

[82] Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genetics in Medicine. 2015 may;17(5):405-24. Available from: http://www.nature.com/articles/gim201530.

[83] Guerreiro RJ, Baquero M, Blesa R, Boada M, Brás JM, Bullido MJ, et al. Genetic screening of Alzheimer's disease genes in Iberian and African samples yields novel mutations in presenilins and APP. Neurobiology of Aging. 2010 may;31(5):725-31. Available from: https://www.sciencedirect.com/science/article/pii/S0197458008002194{#}fig2.

[84] DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ, et al. Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS. Neuron. 2011 oct;72(2):245-56.

[85] Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics. 2012 sep;28(18):i333-9. Available from: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/bts378.

[86] Zhu M, Need AC, Han Y, Ge D, Maia JM, Zhu Q, et al. Using ERDS to infer copy-number variants in high-coverage genomes. American Journal of Human Genetics. 2012 sep;91(3):408-21.

[87] Abyzov A, Urban AE, Snyder M, Gerstein M. CNVnator: An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Research. 2011 jun;21(6):974-84.

[88] Xi R, Lee S, Xia Y, Kim TM, Park PJ. Copy number analysis of whole-genome data using BIC-seq2 and its application to detection of cancer susceptibility variants. Nucleic Acids Research. 2016 jul;44(13):6274-86.

[89] Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al.. Integrative genomics viewer. Nature Publishing Group; 2011.

[90] Bellenguez C, Charbonnier C, Grenier-Boley B, Quenez O, Le Guennec K, Nicolas G, et al. Contribution to Alzheimer's disease risk of rare variants

in TREM2, SORL1, and ABCA7 in 1779 cases and 1273 controls. Neurobiology of Aging. 2017 nov;59:220.e1-220.e9. Available from: https://ac.els-cdn.com/S0197458017302324/1-s2.0-S0197458017302324-main.pdf?{_}tid=40b88b1a-3e1c-4262-b7cb-33ec8ae2bcfa{&}acdnat=1521235271{_}f9c9f72160c28f3f72f5a3e1e0ed5885https://ac.els-cdn.com/S0197458017302324/1-s2.0-S0197458017302324-main.pdf?{_}tid=b06dd47.

[91] Zhou SL, Tan CC, Hou XH, Cao XP, Tan L, Yu JT. TREM2 Variants and Neurodegenerative Diseases: A Systematic Review and Meta-Analysis. IOS Press; 2019.

[92] Steinberg S, Stefansson H, Jonsson T, Johannsdottir H, Ingason A, Helgason H, et al. Loss-of-function variants in ABCA7 confer risk of Alzheimer's disease. Nature Genetics. 2015 may;47(5):445-7. Available from: http://www.nature.com/http://www.nature.com/articles/ng.3246.

[93] Cuyvers E, De Roeck A, Van den Bossche T, Van Cauwenberghe C, Bettens K, Vermeulen S, et al. Mutations in ABCA7 in a Belgian cohort of Alzheimer's disease patients: A targeted resequencing study. The Lancet Neurology. 2015;14(8):814-22. Available from: http://dx.doi.org/10.1016.

[94] Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. Genome Research. 2001 may;11(5):863-74. Available from: https://pubmed.ncbi.nlm.nih.gov/11337480/https://pubmed.ncbi.nlm.nih.gov/11337480/?dopt=Abstract.

[95] Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al.. A method and server for predicting damaging missense mutations. Nat Methods; 2010. Available from: https://pubmed.ncbi.nlm.nih.gov/20354512/http://www.nature.com/articles/nmeth0410-248.

[96] Schwarz JM, Cooper DN, Schuelke M, Seelow D. Mutationtaster2: Mutation prediction for the deep-sequencing age. Nature Publishing Group; 2014. Available from: http://genetics.

[97] Kim M, Suh J, Romano D, Truong MH, Mullin K, Hooli B, et al. Potential late-onset Alzheimer's disease-associated mutations in the ADAM10 gene attenuate α-secretase activity. Human Molecular Genetics. 2009 oct;18(20):3987-96. Available from: https://academic.oup.com/hmg/article/18/20/3987/645530https://pubmed.ncbi.nlm.nih.gov/19608551/.

[98] Suh J, Choi SH, Romano DM, Gannon MA, Lesinski AN, Kim DY, et al. ADAM10 Missense Mutations Potentiate β-Amyloid Accumulation by Impairing Prodomain Chaperone Function. Neuron. 2013 oct;80(2):385-401.

[99] den Dunnen JT, Dalgleish R, Maglott DR, Hart RK, Greenblatt MS, Mcgowan-Jordan J, et al. HGVS Recommendations for the Description of Sequence Variants: 2016 Update. Human Mutation. 2016 jun;37(6):564-9. Available from: http://doi.wiley.com/10.1002/humu.22981.

[100] The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature. 2015 oct;526(7571):68-74. Available from: http://www.nature.com/articles/nature15393.

[101] Rishishwar L, Conley AB, Wigington CH, Wang L, Valderrama-Aguirre A, Jordan IK, et al. Ancestry, admixture and fitness in Colombian genomes. Scientific reports. 2015 jul;5(1):12376. Available from: http://www.nature.com/articles/srep12376http://www.ncbi.nlm.nih.gov/pubmed/26197429http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4508918.

[102] Carvajal-Carmona LG, Soto ID, Pineda N, Ortíz-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. American journal of human genetics. 2000 nov;67(5):1287-95. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11032790http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1288568.

[103] Mathias RA, Taub MA, Gignoux CR, Fu W, Musharoff S, O'Connor TD, et al. A continuum of admixture in the Western Hemisphere revealed by the African Diaspora genome. Nature Communications. 2016 dec;7(1):12522. Available from: http://www.nature.com/articles/ncomms12522.

[104] Ramirez Aguilar L, Acosta-Uribe J, Giraldo MM, Moreno S, Baena A, Alzate D, et al. Genetic origin of a large family with a novel PSEN1 mutation (Ile416Thr). Alzheimer's & dementia : the journal of the Alzheimer's Association. 2019 may;15(5):709-19. Available from: https://alz-journals.onlinelibrary.wiley.com/doi/10.1016/j.jalz.2018.12.010.

[105] Arango D, Cruts M, Torres O, Backhovens H, Serrano ML, Villareal E, et al. Systematic genetic study of Alzheimer disease in Latin America: Mutation frequencies of the amyloid β precursor protein and presenilin genes in Colombia. American Journal of Medical Genetics. 2001 oct;103(2):138-43. Available from: http://doi.wiley.com/10.1002/1096-8628{%}2820011001{%}29103{%}3A2{%}3C138{%}3A{%}3AAID-AJMG1529{%}3E3.0.CO{%}3B2-8.

[106] Gómez-Tortosa E, Barquero S, Barón M, Gil-Neciga E, Castellanos F, Zurdo M, et al. Clinical-genetic correlations in familial Alzheimer's disease caused by presenilin 1 mutations. Journal of Alzheimer's disease : JAD. 2010;19(3):873-84. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20157243.

[107] Itzcovich T, Chrem-Méndez P, Vázquez S, Barbieri-Kennedy M, Niikado M, Martinetto H, et al. A novel mutation in PSEN1 (p.T119I) in an Argentine family with early- and late-onset Alzheimer's disease. Neurobiology of Aging. 2020 jan;85:155.e9-155.e12.

[108] Campion D, Flaman JM, Brice A, Hannequin D, Dubois B, Martin C, et al. Mutations of the presenilin I gene in families with early-onset alzheimer's disease. Human Molecular Genetics. 1995 dec;4(12):2373-7. Available from: https://pubmed.ncbi.nlm.nih.gov/8634712/.

[109] Rogaeva EA, Fafel KC, Song YQ, Medeiros H, Sato C, Liang Y, et al. Screening for PS1 mutations in a referral-based series of AD cases: 21 Novel mutations. Neurology. 2001 aug;57(4):621-5.

[110] Miravalle L, Calero M, Takao M, Roher AE, Ghetti B, Vidal R. Amino-Terminally Truncated Aβ Peptide Species Are the Main Component of Cotton Wool Plaques †. Biochemistry. 2005 aug;44(32):10810-21. Available from: https://pubmed.ncbi.nlm.nih.gov/16086583/.

[111] Jiménez Caballero PE, De Diego Boguna C, Martin Correa E, Serviá Candela M, Marsal Alonso C. A novel presenilin 1 mutation (V261L) associated with presenile Alzheimer's disease and spastic paraparesis. European Journal of Neurology. 2008 sep;15(9):991-4. Available from: https://pubmed.ncbi.nlm.nih.gov/18637955/.

[112] Tabira T, De Chui H, Nakayama H, Kuroda S, Shibuya M. Alzheimer's disease with spastic paresis and cotton wool type plaques. Wiley-Blackwell; 2002. Available from: http://doi.wiley.com/10.1002/jnr.10392.

[113] O'Riordan S, McMonagle P, Janssen JC, Fox NC, Farrell M, Collinge J, et al. Presenilin-1 mutation (E280G), spastic paraparesis, and cranial MRI white-matter abnormalities. Neurology. 2002 oct;59(7):1108-10. Available from: https://pubmed.ncbi.nlm.nih.gov/12370477/.

[114] Lanoiselée HMM, Nicolas G, Wallon D, Rovelet-Lecrux A, Lacour M, Rousseau S, et al. APP, PSEN1, and PSEN2 mutations in early-onset Alzheimer disease: A genetic screening study of familial and sporadic cases. PLoS Medicine. 2017 mar;14(3):e1002270. Available from: https://doi.org/10.1371/journal.pmed.1002270http://dx.plos.org/10.1371/journal.pmed.1002270https://dx.plos.org/10.1371/journal.pmed.1002270.

[115] Rutherford NJ, Zhang YJ, Baker M, Gass JM, Finch NCA, Xu YF, et al. Novel mutations in TARDBP(TDP-43) in patients with familial amyotrophic lateral sclerosis. PLoS Genetics. 2008 sep;4(9):e1000193. Available from: http://dx.plos.org/10.1371/journal.pgen.1000193.

[116] Corrado L, Ratti A, Gellera C, Buratti E, Castellotti B, Carlomagno Y, et al. High frequency of TARDBP gene mutations in Italian patients with amyotrophic lateral sclerosis. Human Mutation. 2009 apr;30(4):688-94. Available from: http://doi.wiley.com/10.1002/humu.20950.

[117] Ramos EM, Koros C, Dokuru DR, Van Berlo V, Kroupis C, Wojta K, et al. Frontotemporal dementia spectrum: first genetic screen in a Greek cohort. Neurobiology of Aging. 2019 mar;75:224.e1-224.e8. Available from: https://www-sciencedirect-com.proxy.library.ucsb.edu:9443/science/article/pii/S0197458018303932?via{%}3Dihubhttps://www.sciencedirect.com/science/article/pii/S0197458018303932?via{%}3Dihub.

[118] Cuyvers E, van der Zee J, Bettens K, Engelborghs S, Vandenbulcke M, Robberecht C, et al. Genetic variability in SQSTM1 and risk of early-onset Alzheimer dementia: A European early-onset dementia consortium study. Neurobiology of Aging. 2015 may;36(5):2005.e15-2005.e22. Available from: https://pubmed.ncbi.nlm.nih.gov/25796131/.

[119] Sleegers K, Pastor P, Sarafov S, Mattheijssens M, Nacmias B, Ortega-Cubero S, et al. Rare mutations in SQSTM1 modify susceptibility to frontotemporal lobar degeneration. Acta Neuropathologica. 2014 sep;128(3):397-410. Available from: http://link.springer.com/10.1007/s00401-014-1298-7.

[120] Andersen PM, Sims KB, Xin WW, Kiely R, O'Neill G, Ravits J, et al.. Sixteen novel mutations in the Cu/Zn superoxide dismutase gene in amyotrophic lateral sclerosis: A decade of discoveries, defects and disputes; 2003. Available from: https://www.tandfonline.com/action/journalInformation?journalCode=iafd20.

[121] Lattante S, Conte A, Zollino M, Luigetti M, Del Grande A, Marangi G, et al. Contribution of major amyotrophic lateral sclerosis genes to the etiology of sporadic disease. Neurology. 2012 jul;79(1):66-72. Available from: http://www.neurology.org/cgi/doi/10.1212/WNL.0b013e31825dceca.

[122] Kenna KP, McLaughlin RL, Byrne S, Elamin M, Heverin M, Kenny EM, et al. Delineating the genetic heterogeneity of ALS using targeted high-throughput sequencing. Journal of Medical Genetics. 2013 nov;50(11):776-83.

[123] Lenk GM, Ferguson CJ, Chow CY, Jin N, Jones JM, Grant AE, et al. Pathogenic mechanism of the FIG4 mutation responsible for charcot-marie-tooth disease CMT4J. PLoS Genetics. 2011 jun;7(6). Available from: https://pubmed.ncbi.nlm.nih.gov/21655088/.

[124] Renaud L, Picher-Martel V, Codron P, Julien JP. Key role of UBQLN2 in pathogenesis of amyotrophic lateral sclerosis and frontotemporal dementia. BioMed Central Ltd.; 2019. Available from: https://actaneurocomms.biomedcentral.com/articles/10.1186/s40478-019-0758-7.

[125] Rademakers R, Baker M, Nicholson AM, Rutherford NJ, Finch N, Soto-Ortolaza A, et al. Mutations in the colony stimulating factor 1 receptor (CSF1R) gene cause hereditary diffuse leukoencephalopathy with spheroids. Nature Genetics. 2012 feb;44(2):200-5. Available from: https://pubmed.ncbi.nlm.nih.gov/22197934/.

[126] Pimenova AA, Marcora E, Goate AM. A Tale of Two Genes: Microglial Apoe and Trem2. Cell Press; 2017. Available from: https://www.sciencedirect.com/science/article/pii/S1074761317303734.

[127] Jiang T, Hou JK, Gao Q, Yu JT, Zhou JS, Zhao HD, et al. TREM2 p.H157Y Variant and the Risk of Alzheimer's Disease: A Meta-Analysis Involving 14,510 Subjects. Current Neurovascular Research. 2016 aug;13(4):318-20. Available from: https://pubmed.ncbi.nlm.nih.gov/27501831/.

[128] Jin SC, Carrasquillo MM, Benitez BA, Skorupa T, Carrell D, Patel D, et al. TREM2 is associated with increased risk for Alzheimer's disease in African Americans. Molecular Neurodegeneration. 2015 dec;10(1):19. Available from: http://www.molecularneurodegeneration.com/content/10/1/19.

[129] Guerreiro RJ, Lohmann E, Brás JM, Gibbs JR, Rohrer JD, Gurunlian N, et al. Using exome sequencing to reveal mutations in TREM2 presenting as a frontotemporal dementia-like syndrome without bone involvement. Archives of Neurology. 2013 jan;70(1):78-84. Available from: http://archneur.jamanetwork.com/article.aspx?doi=10.1001/jamaneurol.2013.579.

[130] Rayaprolu S, Mullen B, Baker M, Lynch T, Finger E, Seeley WW, et al. TREM2 in neurodegeneration: Evidence for association of the p.R47H variant with frontotemporal dementia and Parkinson's disease. Molecular Neurodegeneration. 2013 jun;8(1):19. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23800361http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3691612.

[131] Jiao B, Liu X, Tang B, Hou L, Zhou L, Zhang F, et al. Investigation of TREM2, PLD3, and UNC5C variants in patients with Alzheimer's disease from mainland China. Neurobiology of Aging. 2014;35(10):2422.e9-2422.e11. Available from: https://pubmed.ncbi.nlm.nih.gov/24866402/.

[132] Ma JF, Zhou Y, Xu J, Liu XH, Wang Y, Deng YL, et al. Association study of TREM2 polymorphism rs75932628 with late-onset Alzheimer's disease in Chinese Han population. Neurological Research. 2014 oct;36(10):894-6. Available from: https://pubmed.ncbi.nlm.nih.gov/24725293/.

[133] Song W, Hooli B, Mullin K, Jin SC, Cella M, Ulland TK, et al. Alzheimer's disease-associated TREM2 variants exhibit either decreased or increased ligand-dependent activation. Alzheimer's and Dementia. 2017 apr;13(4):381-7. Available from: https://pubmed.ncbi.nlm.nih.gov/27520774/.

[134] Sims R, Van Der Lee SJ, Naj AC, Bellenguez C, Badarinarayan N, Jakobsdottir J, et al. Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. Nature Genetics. 2017 jul;49(9):1373-84. Available from: https://pubmed.ncbi.nlm.nih.gov/28714976/http://www.nature.com/doifinder/10.1038/ng.3916https://www.nature.com/articles/ng.3916.pdf.

[135] Albers PK, McVean G. Human Genome Dating – Allele age of variant rs75932628; 2020. Available from: https://human.genome.dating/snp/rs2234255.jnthttps://human.genome.dating/snp/rs75932628.

[136] McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, et al. Runs of Homozygosity in European Populations. American Journal of Human Genetics. 2008 sep;83(3):359-72. Available from: http://www.cell.com/article/S000292970800445X/fulltexthttp://www.cell.com/article/S000292970800445X/abstracthttps://www.cell.com/ajhg/abstract/S0002-9297(08)00445-X.

[137] Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg NA, Li JZ. Genomic patterns of homozygosity in worldwide human populations. American Journal of Human Genetics. 2012;91(2):275-92. Available from: http://dx.doi.org/10.1016/j.ajhg.2012.06.014.

[138] Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW, Cavalli-Sforza LL. Support from the relationship of genetic and geographic in human populations for a serial founder effect originating in Africa. Proceedings of the National Academy of Sciences of the United States of America. 2005 nov;102(44):15942-7. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.0507611102.

[139] Coop G, Pickrell JK, Novembre J, Kudaravalli S, Li J, Absher D, et al. The role of geography in human adaptation. PLoS Genetics. 2009 jun;5(6):e1000500. Available from: https://dx.plos.org/10.1371/journal.pgen.1000500.

[140] Zhao C, Strobino K, Moon YP, Cheung YK, Sacco RL, Stern Y, et al. APOE $\epsilon 4$ modifies the relationship between infectious burden and poor cognition. Neurology: Genetics. 2020 aug;6(4):e462. Available from: https://ng.neurology.org/content/6/4/e462https://ng.neurology.org/content/6/4/e462.abstract.

[141] Soscia SJ, Kirby JE, Washicosky KJ, Tucker SM, Ingelsson M, Hyman B, et al. The Alzheimer's Disease-Associated Amyloid β-Protein Is an Antimicrobial Peptide. PLoS ONE. 2010 mar;5(3):e9505. Available from: https://dx.plos.org/10.1371/journal.pone.0009505.

[142] Shen J, Kelleher RJ. The presenilin hypothesis of Alzheimer's disease: Evidence for a loss-of-function pathogenic mechanism; 2007. 2. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.0608332104.

[143] Cochran JN, McKinley EC, Cochran M, Amaral MD, Moyers BA, Lasseigne BN, et al. Genome sequencing for early-onset or atypical dementia: High diagnostic yield and frequent observation of multiple contributory alleles. Cold Spring Harbor Molecular Case Studies. 2019;5(6):1-19.

[144] Joutel A, Chabriat H, Vahedi K, Domenga V, Vayssiere C, Ruchoux MM, et al. Splice site mutation causing a seven amino acid Notch3 in-frame deletion in CADASIL. Neurology. 2000 may;54(9):1874-5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10802807.

[145] Arboleda-Velasquez JF, Lopera F, Lopez E, Frosch MP, Sepulveda-Falla D, Gutierrez JE, et al. C455R notch3 mutation in a Colombian CADASIL kindred with early onset of stroke. Neurology. 2002 jul;59(2):277-9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12136071.

[146] Kremeyer B, Lopera F, Cox JJ, Momin A, Rugiero F, Marsh S, et al. A Gain-of-Function Mutation in TRPA1 Causes Familial Episodic Pain Syndrome. Neuron. 2010 jun;66(5):671-80. Available from: https://pubmed.ncbi.nlm.nih.gov/20547126/.

[147] Arango-Lasprilla JC, Iglesias-Dorado J, Moreno S, Lopera F. A neuropsychological study of Huntington's disease in families in Antioquia, Colombia. Revista de Neurologia. 2003 jul;37(1):7-13. Available from: https://europepmc.org/article/med/12861500.

[148] Cornejo-Olivas M, Torres L, Velit-Salazar MR, Inca-Martinez M, Mazzetti P, Cosentino C, et al. Variable frequency of LRRK2 variants in the Latin American research consortium on the genetics of Parkinson's disease (LARGE-PD), a case of ancestry /692/617/375/1718 /631/208/1516 article. npj Parkinson's Disease. 2017 dec;3(1).

[149] Ramírez-Castro JL, Pineda-Trujillo N, Valencia AV, Muñetón CM, Botero O, Trujillo O, et al. Mutations in FOXL2 underlying BPES (types 1 and 2) in Colombian families. American Journal of Medical Genetics. 2002 nov;113(1):47-51. Available from: https://pubmed.ncbi.nlm.nih.gov/12400065/.

[150] Pineda-Trujillo N, Cornejo W, Carrizosa J, Wheeler RB, Múnera S, Valencia A, et al. A CLN5 mutation causing an atypical neuronal ceroid lipofuscinosis of juvenile onset. Neurology. 2005 feb;64(4):740-2. Available from: https://n.neurology.org/content/64/4/740https://n.neurology.org/content/64/4/740.abstract.

[151] Bouzaher MH, Worden CP, Jeyakumar A. Systematic Review of Pathogenic GJB2 Variants in the Latino Population. Lippincott Williams and Wilkins; 2020. Available from: http://journals.lww.com/10.1097/MAO.0000000000002505.

[152] Torres D, Lorenzo Bermejo J, Rashid MU, Bricenõ I, Gil F, Beltran A, et al. Prevalence and Penetrance of BRCA1 and BRCA2 Germline Mutations in Colombian Breast Cancer Patients. Scientific Reports. 2017 dec;7(1):1-9.

[153] Naj AC, Schellenberg GD. Genomic variants, genes, and pathways of Alzheimer's disease: An overview. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics. 2017 jan;174(1):5-26. Available from: https://onlinelibrary.wiley.com/doi/10.1002/ajmg.b.32499.

[154] Cruts M, Theuns J, Van Broeckhoven C. Locus-specific mutation databases for neurodegenerative brain diseases. Human Mutation. 2012 sep;33(9):1340-4. Available from: http://doi.wiley.com/10.1002/humu.22117.

[155] Ryan NS, Nicholas JM, Weston PSJ, Liang Y, Lashley T, Guerreiro R, et al. Clinical phenotype and genetic associations in autosomal dominant familial Alzheimer's disease: a case series. The Lancet Neurology. 2016 dec;15(13):1326-35. Available from: https://www.sciencedirect.com/science/article/pii/S1474442216301934http://linkinghub.elsevier.com/retrieve/pii/S1474442216301934http://www.ncbi.nlm.nih.gov/pubmed/27777022.

[156] Ryman DC, Acosta-Baena N, Aisen PS, Bird T, Danek A, Fox NC, et al. Symptom onset in autosomal dominant Alzheimer disease: a systematic review and meta-analysis [Journal Article]. Neurology. 2014 jul;83(3):253-60. Available from: https://www.ncbi.nlm.nih.gov/pubmed/24928124.

[157] Fleisher AS, Chen K, Quiroz YT, Jakimovich LJ, Gutierrez Gomez M, Langois CM, et al. Associations between biomarkers and age in the presenilin 1 E280A autosomal dominant Alzheimer disease kindred: a cross-sectional study. JAMA neurology. 2015 mar;72(3):316-24. Available from: https://jamanetwork.com/journals/jamaneurology/fullarticle/2089217.

[158] Moulder KL, Snider BJ, Mills SL, Buckles VD, Santacruz AM, Bateman RJ, et al.. Dominantly inherited alzheimer network: Facilitating research and clinical trials. BioMed Central; 2013. Available from: http://alzres.biomedcentral.com/articles/10.1186/alzrt213.

[159] Ardila A, Lopera F, Rosselli M, Moreno S, Madrigal L, Arango-Lasprilla JC, et al. Neuropsychological profile of a large kindred with familial Alzheimer's disease caused by the E280A single presenilin-1 mutation [Journal Article]. Arch Clin Neuropsychol. 2000 aug;15(6):515-28. Available from: https://www.sciencedirect.com/science/article/pii/S0887617799000414https://www.ncbi.nlm.nih.gov/pubmed/14590205.

[160] Petersen RC. Mild Cognitive Impairment. New England Journal of Medicine. 2011 jun;364(23):2227-34. Available from: http://www.nejm.org/doi/abs/10.1056/NEJMcp0910237http://www.nejm.org/doi/abs/10.1056/NEJMc1108238.

[161] Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. Current Protocols in Bioinformatics. 2013 oct;43(SUPL.43):11.10.1-33. Available from: http://doi.wiley.com/10.1002/0471250953.bi1110s43http://www.ncbi.nlm.nih.gov/pubmed/25431634http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4243306.

[162] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009 jul;25(14):1754-60. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19451168http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2705234https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btp324http://massgenomics.org/2009/12/making-the-leap-maq-t.

[163] Browning SR, Browning BL. Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. The American Journal of Human Genetics. 2007 nov;81(5):1084-97. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17924348http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2265661http://linkinghub.elsevier.com/retrieve/pii/S0002929707638828.

[164] Machiela MJ, Chanock SJ. LDlink: A web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics. 2015 nov;31(21):3555-7. Available from: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btv402http://www.ncbi.nlm.nih.gov/pubmed/26139635http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4626747.

[165] Loh PR, Danecek P, Palamara PF, Fuchsberger C, Reshef YA, Finucane HK, et al. Reference-based phasing using the Haplotype Reference Consortium panel. Nature Genetics. 2016;48(11):1443-8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27694958http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5096458.

[166] Cummings JL, Mega M, Gray K, Rosenberg-Thompson S, Carusi DA, Gornbein J. The Neuropsychiatric Inventory: Comprehensive assessment of psychopathology in dementia. Neurology. 1994 dec;44(12):2308-8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7991117http://www.neurology.org/cgi/doi/10.1212/WNL.44.12.2308.

[167] Correa Bustamante CM. De Hatogrande a Girardota [Monografia]. Universidad de Antioquia; 2002. Available from: http://bibliotecadigital.udea.edu.co/bitstream/10495/228/1/HatograndeGirardota.pdf.

[168] Aguirre-Acevedo DC, Lopera F, Henao E, Tirado V, Munoz C, Giraldo M, et al. Cognitive decline in a colombian kindred with autosomal dominant Alzheimer disease a retrospective cohort study [Journal Article]. JAMA Neurology. 2016 apr;73(4):431-8. Available from: http://archneur.jamanetwork.com/article.aspx?doi=10.1001/jamaneurol.2015.4851https://www.ncbi.nlm.nih.gov/pubmed/26902171.

[169] Neuner SM, TCW J, Goate AM. Genetic architecture of Alzheimer's disease [Journal Article]. Neurobiology of Disease. 2020 sep;143:104976. Available from: https://www.ncbi.nlm.nih.gov/pubmed/32565066.

[170] Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009 jul;460(7256):748-52. Available from: http://www.nature.com/doifinder/10.1038/nature08185.

[171] Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies [Journal Article]. Nature Genetics. 2012 jul;44(7):821-4. Available from: http://www.nature.com/articles/ng.2310.

[172] Naj AC, Jun G, Reitz C, Kunkle BW, Perry W, Park YS, et al. Effects of multiple genetic loci on age at onset in late-onset Alzheimer disease: a genome-wide association study [Journal Article]. JAMA Neurol. 2014;71(11):1394-404. Available from: https://www.ncbi.nlm.nih.gov/pubmed/25199842.

[173] Huang KL, Marcora E, Pimenova AA, Di Narzo AF, Kapoor M, Jin SC, et al. A common haplotype lowers PU.1 expression in myeloid cells and delays onset of Alzheimer's disease [Journal Article]. Nat Neurosci. 2017;20(8):1052-61. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28628103.

[174] Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. Nature Genetics. 2019 mar;51(3):414-30. Available from: http://www.nature.com/articles/s41588-019-0358-2.

[175] Velez JI, Rivera D, Mastronardi CA, Patel HR, Tobon C, Villegas A, et al. A Mutation in DAOA Modifies the Age of Onset in PSEN1 E280A Alzheimer's Disease [Journal Article]. Neural Plast. 2016;2016:9760314. Available from: https://www.ncbi.nlm.nih.gov/pubmed/26949549.

[176] Pastor P, Roe CM, Villegas A, Bedoya G, Chakraverty S, García G, et al. Apolipoprotein Eϵ4 modifies Alzheimer's disease onset in an E280A PS1 kindred [Journal Article]. Annals of Neurology. 2003 aug;54(2):163-9. Available from: https://www.ncbi.nlm.nih.gov/pubmed/12891668https://onlinelibrary.wiley.com/doi/10.1002/ana.10636.

[177] Lendon CL, Martinez A, Behrens IM, Kosik KS, Madrigal L, Norton J, et al. E280A PS-1 mutation causes Alzheimer's disease but age of onset is not modified by ApoE alleles [Journal Article]. Hum Mutat. 1997;10(3):186-95. Available from: https://www.ncbi.nlm.nih.gov/pubmed/9298817.

[178] Marioni RE, Harris SE, Zhang Q, McRae AF, Hagenaars SP, Hill WD, et al. GWAS on family history of Alzheimer's disease [Journal Article]. Transl Psychiatry. 2018;8(1):99. Available from: https://www.ncbi.nlm.nih.gov/pubmed/29777097.

[179] Deming Y, Xia J, Cai Y, Lord J, Holmans P, Bertelsen S, et al. A potential endophenotype for Alzheimer's disease: cerebrospinal fluid clusterin [Journal Article]. Neurobiol Aging. 2016;37:208 e1-208 e9. Available from: https://www.ncbi.nlm.nih.gov/pubmed/26545630.

[180] Lee JH, Cheng R, Vardarajan B, Lantigua R, Reyes-Dumeyer D, Ortmann W, et al. Genetic Modifiers of Age at Onset in Carriers of the G206A Mutation in PSEN1 With Familial Alzheimer Disease Among Caribbean Hispanics [Journal Article]. JAMA Neurology. 2015 sep;72(9):1043. Available from: https://www.ncbi.nlm.nih.gov/pubmed/26214276.

[181] Luo J, Elwood F, Britschgi M, Villeda S, Zhang H, Ding Z, et al. Colony-stimulating factor 1 receptor (CSF1R) signaling in injured neurons facilitates protection and survival [Journal Article]. J Exp Med. 2013;210(1):157-72. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23296467.

[182] Mizuno T, Doi Y, Mizoguchi H, Jin S, Noda M, Sonobe Y, et al. Interleukin-34 selectively enhances the neuroprotective effects of microglia to attenuate oligomeric amyloid-beta neurotoxicity [Journal Article]. Am J Pathol. 2011;179(4):2016-27. Available from: https://www.ncbi.nlm.nih.gov/pubmed/21872563.

[183] Haining EJ, Yang J, Bailey RL, Khan K, Collier R, Tsai S, et al. The TspanC8 subgroup of tetraspanins interacts with A disintegrin and metalloprotease 10 (ADAM10) and regulates its maturation and cell surface expression [Journal Article]. J Biol Chem. 2012;287(47):39753-65. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23035126.

[184] Peron R, Vatanabe IP, Manzine PR, Camins A, Cominetti MR. Alpha-Secretase ADAM10 Regulation: Insights into Alzheimer's Disease Treatment [Journal Article]. Pharmaceuticals (Basel). 2018;11(1). Available from: https://www.ncbi.nlm.nih.gov/pubmed/29382156.

[185] Deng M, Zhang Q, Wu Z, Ma T, He A, Zhang T, et al. Mossy cell synaptic dysfunction causes memory imprecision via miR-128 inhibition of STIM2 in Alzheimer's disease mouse model [Journal Article]. Aging Cell. 2020;19(5):e13144. Available from: https://www.ncbi.nlm.nih.gov/pubmed/32222058.

[186] Group THDCR. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. [Journal Article]. Cell. 1993;72(6):971-83. Available from: https://www.ncbi.nlm.nih.gov/pubmed/8458085.

[187] Bruyere J, Abada YS, Vitet H, Fontaine G, Deloulme JC, Ces A, et al. Presynaptic APP levels and synaptic homeostasis are regulated by Akt phosphorylation of huntingtin [Journal Article]. Elife. 2020;9. Available from: https://www.ncbi.nlm.nih.gov/pubmed/32452382.

[188] Barcia G, Fleming MR, Deligniere A, Gazula VR, Brown MR, Langouet M, et al. De novo gain-of-function KCNT1 channel mutations cause malignant migrating partial seizures of infancy [Journal Article]. Nat Genet. 2012;44(11):1255-9. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23086397.

[189] Cirulli ET, Lasseigne BN, Petrovski S, Sapp PC, Dion PA, Leblond CS, et al. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways [Journal Article]. Science. 2015;347(6229):1436-41. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25700176.

[190] Vossel KA, Tartaglia MC, Nygaard HB, Zeman AZ, Miller BL. Epileptic activity in Alzheimer's disease: causes and clinical relevance [Journal Article]. Lancet Neurol. 2017;16(4):311-22. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28327340.

[191] Dunn AR, Kaczorowski CC. Regulation of intrinsic excitability: Roles for learning and memory, aging and Alzheimer's disease, and genetic diversity [Journal Article]. Neurobiol Learn Mem. 2019;164:107069. Available from: https://www.ncbi.nlm.nih.gov/pubmed/31442579.

[192] Hall AM, Throesch BT, Buckingham SC, Markwardt SJ, Peng Y, Wang Q, et al. Tau-dependent Kv4.2 depletion and dendritic hyperexcitability in a mouse model of Alzheimer's disease [Journal Article]. J Neurosci. 2015;35(15):6221-30. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25878292.

[193] Mooney JA, Huber CD, Service S, Sul JH, Marsden CD, Zhang Z, et al. Understanding the Hidden Complexity of Latin American Population Isolates. [Journal Article]. American journal of human genetics. 2018 nov;103(5):707-26. Available from: https://doi.org/10.1016/j.ajhg.2018.09.013.

[194] Allison AC. Protection afforded by sickle-cell trait against subtertian malarial infection. British Medical Journal. 1954;1(4857):290-4. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/pmc2093356/.