Rare Functional Variant in *TM2D3* is Associated with Late-Onset Alzheimer's Disease

Johanna Jakobsdottir[1☺], Sven J. van der Lee[2☺], Joshua C. Bis[3☺], Vincent Chouraki[4,5☺], David Li-Kroeger[6,7☺], Shinya Yamamoto[6,7,8☺], Megan L. Grove[9], Adam Naj[10], Maria Vronskaya[11], Jose L. Salazar[6], Anita L. DeStefano[5,12], Jennifer A. Brody[3], Albert V. Smith[1,13], Najaf Amin[2], Rebecca Sims[11], Carla A. Ibrahim-Verbaas[2,14], Seung-Hoan Choi[5,12], Claudia L. Satizabal[4,5], Oscar L. Lopez[15], Alexa Beiser[4,5,12], M. Arfan Ikram[2,14,16], Melissa E. Garcia[17], Caroline Hayward[18,19], Tibor V. Varga[20], Samuli Ripatti[21,22], Paul W. Franks[20,23,24], Göran Hallmans[25], Olov Rolandsson[26], Jan-Håkon Jansson[23,27], David J. Porteous[19,28], Veikko Salomaa[29], Gudny Eiriksdottir[1], Kenneth M. Rice[30], Hugo J. Bellen[6,7,8,31], Daniel Levy[5,32,4], Andre G. Uitterlinden[2,33], Valur Emilsson[1,34], Jerome I. Rotter[35], Thor Aspelund[1,36], Cohorts for Heart and Aging Research in Genomic Epidemiology consortium[¶], Alzheimer's Disease Genetic Consortium[¶], Genetic and Environmental Risk in Alzheimer's Disease consortium[¶], Christopher J. O'Donnell[5,32], Annette L. Fitzpatrick[37,38], Lenore J. Launer[17], Albert Hofman[2], Li-San Wang[39], Julie Williams[11], Gerard D. Schellenberg[39], Eric Boerwinkle[9,40‡], Bruce M. Psaty[3,37,41,42‡], Sudha Seshadri[4,5‡], Joshua M. Shulman[6,7,8‡*], Vilmundur Gudnason[1,13‡*], Cornelia M. van Duijn[2‡*]

1 Icelandic Heart Association, Kopavogur, Iceland, **2** Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands, **3** Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, Washington, United States of America, **4** Boston University School of Medicine, Boston, Massachusetts, United States of America, **5** Framingham Heart Study, Framingham, Massachusetts, United States of America, **6** Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, United States of America, **7** Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital Houston, Texas, United States of America, **8** Program in Developmental Biology, Baylor College of Medicine, Houston, Texas, United States of America, **9** School of Public Health, Human Genetics Center, The University of Texas Health Science Center at Houston, Houston, Texas, United States of America, **10** Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America, **11** Institute of Psychological Medicine and Clinical Neurosciences, Medical Research Council (MRC) Centre for Neuropsychiatric Genetics & Genomics, Cardiff University, Cardiff, United Kingdom, **12** Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, United States of America, **13** Faculty of Medicine, University of Iceland, Reykjavik, Iceland, United States of America, **14** Department of Neurology, Erasmus University Medical Center, CA Rotterdam, The Netherlands, **15** Department of Neurology, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania, United States of America, **16** Departments of Radiology, Erasmus University Medical Center, CA Rotterdam, The Netherlands, **17** Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Bethesda, Maryland, United States of America, **18** MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom, **19** Generation Scotland, Centre for Genomic and Experimental Medicine, University of Edinburgh, Edinburgh, United Kingdom, **20** Department of Clinical Sciences, Genetic and Molecular Epidemiology Unit, Lund University, Malmö, Sweden, **21** Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland, **22** Department of Public Health, University of Helsinki, Helsinki, Finland, **23** Department of Public Health & Clinical Medicine, Umeå University Hospital, Umeå, Sweden, **24** Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, United States of America, **25** Department of Biobank Research, Umeå University, Umeå, Sweden, **26** Department of Public Health & Clinical Medicine, Section for Family Medicine, Umeå University, Umeå, Sweden, **27** Research Unit, Skellefteå Hospital, Skellefteå, Sweden, **28** Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom, **29** National Institute for Health and Welfare, Helsinki, Finland, **30** Department of Biostatistics, University of Washington, Seattle, Washington, United States of America, **31** Howard Hughes Medical Institute, Durham, North Carolina, United States of America, **32** Division of Intramural

Research, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland, United states of America, **33** Department of Internal Medicine, Erasmus University Medical Center, CA Rotterdam, The Netherlands, **34** Faculty of Pharmaceutical Sciences, University of Iceland, Reykjavik, Iceland, **35** Institute for Translational Genomics and Population Sciences, Los Angeles BioMedical Research Institute and Departments of Medicine and Pediatrics, Harbor-UCLA Medical Center, Torrance, California, United States of America, **36** Centre for Public Health, University of Iceland, Reykjavik, Iceland, **37** Department of Epidemiology, University of Washington, Seattle, Washington, United States of America, **38** Collaborative Health Studies Coordinating Center, Seattle, Washington, United States of America, **39** Department of Pathology and Laboratory of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America, **40** Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, United States of America, **41** Department of Health Services, University of Washington, Seattle, Washington, United States of America, **42** Group Health Research Institute, Group Health Cooperative, Seattle, Washington, United States of America

☯ These authors contributed equally to this work.
‡ EB and BMP also contributed equally to this work. SS and JMS also contributed equally to this work. VG and CMvD also contributed equally to this work.
¶ Membership of Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium, Alzheimer's Disease Genetic Consortium, and Genetic and Environmental Risk in Alzheimer's Disease Consortium are listed in S1 Text.
* Joshua.Shulman@bcm.edu (JMS); villi@hjarta.is (VG); c.vanduijn@erasmusmc.nl (CMvD)

Abstract

We performed an exome-wide association analysis in 1393 late-onset Alzheimer's disease (LOAD) cases and 8141 controls from the CHARGE consortium. We found that a rare variant (P155L) in *TM2D3* was enriched in Icelanders (~0.5% versus <0.05% in other European populations). In 433 LOAD cases and 3903 controls from the Icelandic AGES substudy, P155L was associated with increased risk and earlier onset of LOAD [odds ratio (95% CI) = 7.5 (3.5–15.9), p = 6.6x10^{-9}]. Mutation in the *Drosophila TM2D3* homolog, *almondex*, causes a phenotype similar to loss of Notch/Presenilin signaling. Human *TM2D3* is capable of rescuing these phenotypes, but this activity is abolished by P155L, establishing it as a functionally damaging allele. Our results establish a rare *TM2D3* variant in association with LOAD susceptibility, and together with prior work suggests possible links to the β-amyloid cascade.

Author Summary

Alzheimer's disease (AD) is the most common cause of dementia in the older adult population. There is substantial evidence for an important genetic contribution to AD risk. While prior work has comprehensively evaluated the contribution of common genetic variants in large population-based cohorts, the role of rare variants remains to be defined. Here, we have used a newer genotyping array to characterize less common variants, including those likely to impact the function of encoded proteins, in a combined cohort of 1393 AD cases and 8141 control subjects without AD. Our results implicate a novel, amino acid-changing variant, P155L, in the *TM2D3* gene. This variant was discovered to be more common in the Icelandic population, where it was significantly associated with both increased risk and earlier age of onset of AD. Lastly, in order to examine the potential functional impact of the implicated variant, we performed additional studies in the fruit fly. Our results suggest that P155L causes a loss-of-function in TM2D3, in the context of

Notch-Presenilin signal transduction. In sum, we identify a novel, rare *TM2D3* variant in association with AD risk and highlight functional connections with AD-relevant biology.

Introduction

Alzheimer's disease (AD, [MIM: 104300]), the most common form of dementia, affects more than 10% of those 65 years and older, increasing to 30% in those 85 years and older [1–4]. Pathologically, AD is characterized by extensive brain neurodegenerative cell loss in association with extra-cellular β-amyloid plaques and intra-neuronal tangles consisting of hyper-phosphorylated tau protein. Multiple mutations in the amyloid-β precursor protein (*APP*, [MIM: 104760]) and the presenilin-1 and -2 (*PSEN1* [MIM: 104311] and *PSEN2* [MIM: 600759]) genes cause familial early-onset (<65 years) AD [5,6]. Genome-wide association studies (GWAS) have also identified numerous common genetic variants of modest effect sizes for late-onset AD (LOAD, [MIM: 104300]) [7–12]. However, *Apolipoprotein E* (*APOE*, [MIM: 107741]) still remains the most important known genetic determinant of LOAD susceptibility [13,14]. Recently, rare variants with effect sizes similar to the *APOE*-ɛ4 allele have also been identified. While the population-attributable risk of such single rare variants is low, their discoveries may have important implications for understanding disease mechanisms and developing novel treatments [15]. Recent examples include a rare protective allele of *APP* [16] as well as variants in the novel genes *TREM2* [MIM: 605086] [17,18], *PLD3* [MIM: 615698] [19], *UNC5C* [MIM: 603610] [20], and *AKAP9* [MIM:604001] [21]. These findings highlight potentially important, new cellular pathways relevant for disease pathophysiology [15]; for example *TREM2* has spurred intense recent interest in the role of microglia in LOAD [22,23]. Importantly, because of population histories and demography, rare variants may be population-specific. The protective *APP* variant, for example, is found predominantly in Iceland and other Scandinavian populations [16]. The two variants in *AKAP9* have only been reported in African-Americans [21], and *PLD3* variants appear to vary greatly in frequency between populations [24].

To date, the identification of rare susceptibility variants in LOAD has been hampered by poor representation on genotyping arrays used for large GWAS [25]; moreover, direct sequencing in large numbers of individuals remains costly. Here, using an exome-wide genotyping array (the Illumina HumanExome Beadchip), we report associations with LOAD in four population-based cohorts from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium [26].

Results

Exome-wide association study of LOAD in CHARGE

The discovery phase of our analysis examined single-variant associations with LOAD. Meta-analysis was performed across four CHARGE studies including 1393 LOAD cases and 8141 controls (**Table 1**), using logistic regression-based scores [27,28] and adjusting for age and sex (see **Methods** and **Supporting methods in S1 Text**). Single-variant results were filtered using quality control filters that included minimum minor allele frequency (MAF ≥ 0.5%) or minor allele count (MAC ≥ 5) in cases (see also **Methods**) resulting in 52026 total variants tested. Given the rarity of many variants captured on the exome array, we additionally performed association analysis, where we considered rare variants from the same gene in aggregate. For this complementary analysis we used the Sequence Kernel Association Test (SKAT) test [29], a

Table 1. Sample characteristics of discovery cohorts.

	AGES		FHS		CHS		RS	
Characteristics	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases
Study Design	Population-based cohort		Population-based cohort		Population-based cohort		Population-based cohort	
No. participants	2374	143	1338	230	2013	557	2416	463
Women, No (%)	1401 (59)	85 (59)	754 (56)	158 (69)	1134 (56)	343 (62)	1227(51)	319(69)
Age, mean (SD), year	78.89 (4.98)	82.50 (4.94)	79.84 (8.57)	85.06 (6.90)	81.18 (5.15)	82.1 (5.32)	78.2(7.71)	83.3(6.59)
APOE ε4+, No. (%)	638 (27)	66 (46)	258 (20)	74 (32)	397 (20)	176 (32)	609(26)	190(43)

AGES: Age, Gene/Environment Susceptibility study, FHS: Framingham Heart Study, CHS: Cardiovascular Health Study, RS: Rotterdam Study.

doi:10.1371/journal.pgen.1006327.t001

variance component score test for the association of a set of multiple variants with a trait. SKAT tests the null hypothesis of no variation of effects and therefore the statistical model of SKAT is applicable to test association of variant sets that may have a combination of variants including both risk and protective alleles or those with no effect. Variants were included in the SKAT analysis based on functional annotation (see **Methods**) and maximum frequency of 5%. Because the SKAT results were filtered by cumulative set-based MAF and MAC, considering all variants in aggregate, and not by minimum MAF of single SNPs (see **Methods**), this complementary analysis included many variants that did not meet the criteria for inclusion in the single-variant tests. The results of both the single-variant and SKAT discovery analysis are presented in **S1 Table** and quantile-quantile plots are in **S1 Fig**. Below, we discuss each of the top three results, our efforts to replicate the novel loci in independent cohorts, and for *TM2D3*, subsequent functional validation of the implicated variant in *Drosophila*.

APOE

As expected based on many prior LOAD GWAS [12], three common intronic variants near *APOE* reached the Bonferroni threshold of exome-wide significance ($p< = 9.6x10^{-7}$) in the single-variant analysis: rs769449 in *APOE* ($p = 5.8x10^{-38}$, $MAF_{case} = 17\%$, $MAF_{control} = 10\%$), rs2075650 in *TOMM40* ($p = 1.2x10^{-24}$, $MAF_{case} = 18\%$, $MAF_{control} = 13\%$), and rs6859 in *PVRL2* ($p = 1.5x10^{-9}$, $MAF_{case} = 46\%$, $MAF_{control} = 40\%$). Although *APOE* is a well-established LOAD risk locus it did not achieve nominal significance in the SKAT analysis, possible because the established common risk variants at this locus exceed our frequency threshold for consideration in the SKAT analysis ($MAF<0.05$) (**S2 Table**).

SKAP2

rs17154402 in *SKAP2* (MIM: 605215) on chromosome 7p15 also showed a significant association with LOAD ($p = 2.1x10^{-7}$, $MAF_{case} = 0.22\%$, $MAF_{control} = 0\%$). This SNP is predicted to introduce a non-synonymous amino acid change (S253T in the longer isoform) in the SKAP2 protein. The *SKAP2* locus also achieved significant association with LOAD ($p = 4.5x10^{-7}$) in the SKAT analysis. Although the SKAT analysis included three distinct *SKAP2* missense alleles, the SKAT result was fully explained by the variant identified in the single-variant analysis (**S2 Table**). rs17154402 was not significantly associated with LOAD risk in the ADGC or GERAD cohorts, which are 2 large European-ancestry case-control datasets (13 carriers in 8256 controls [MAF 0.08%] and 9 carriers in 13333 cases [MAF 0.03%], p = 0.05), and compared to the discovery CHARGE cohort, showed an opposite direction of effect in these samples. We noted that rs17154402 has a higher frequency in persons of African compared to European ancestry (MAF~7% in Yorubans and 3% in all African populations combined in the 1000 Genomes

Project data, and Grove et al [30] reported a ~5% MAF in African populations versus 0.22% in European ancestry subjects). We therefore took advantage of an available African-American subsample of the CHARGE CHS cohort for attempted replication. However, the variant again showed an inconsistent direction of effect and non-significant association (93 [12 carriers] cases, 208 [33 carriers] controls; p = 0.66). Based on the non-replication of *SKAP2* in available independent datasets, it was not pursued further.

TM2D3

Our discovery analysis also identified rs139709573, a missense variant in *TM2D3* (MIM: 610014) on chromosome 15q26. Although the association statistic did not exceed the Bonferroni threshold for exome-wide significance in the CHARGE discovery analysis (single-variant $p = 2.0 \times 10^{-6}$, SKAT $p = 8.3 \times 10^{-6}$), several important observations led us to consider this variant further. First, we noted that while very rare (0–0.06% MAF in European ancestry populations (S3 Table)) the *TM2D3* variant was enriched nearly 10-fold in the Icelandic AGES cohort (0.45% MAF). Second, rs139709573 was significantly associated ($p = 5.9 \times 10^{-8}$, Table 2) with LOAD when the analysis was restricted to the AGES-discovery sample (143 cases, 2374 controls). Third, prior experimental studies potentially link *TM2D3* to AD-relevant biology. Specifically, *TM2D3* shares homology with the β-amyloid peptide binding protein (BBP or *TM2D1*) [31], and as discussed further below, genetic studies of the conserved *Drosophila* ortholog almondex (*amx*) strongly suggest links to γ-secretase function [32].

In the SKAT analysis, the association of *TM2D3* with LOAD was fully accounted for by rs139709573; although, 3 other variants were considered in this analysis (S2 Table). In the subsequent analyses and functional follow-up, we therefore restricted our focus on rs139709573. First we evaluated potential inflation in the test statistic of the *TM2D3* rs139709573 result in the AGES-discovery cohort (**Supporting results in S1 Text**) and we further confirmed the genotypes of risk allele carriers by re-genotyping them with a TaqMan assay. We next directly genotyped rs139709573 in an independent Icelandic AGES-followup cohort (290 cases, 1529 controls), consisting of individuals who had not been genotyped previously with the exome array. The rs139709573 association was also detected in this sample ($p = 3.1 \times 10^{-3}$, Table 2). As

Table 2. Sample characteristics and association results for P155L (rs139709573) in *TM2D3* in the two Icelandic AGES cohorts.

Cohort	Group	No(% women)	Age, mean (SD), year	*APOE* ε4+ No (%)	*TM2D3* carriers, No (%)	p (OR, 95% CI)[a]	p_{Fisher's Exact}[b]	p_{conditional}[c]
AGES-discovery	Cases	143 (59)	82.5 (4.9)	66 (46)	7 (4.9)	5.9×10^{-8} (8.62, 3.43–21.68)	5.6×10^{-4}	8.4×10^{-8}
	Controls	2374 (59)	78.9 (5.0)	638 (27)	20 (0.8)	. . .	. . .	. . .
AGES-followup	Cases	290 (59)	84.5 (5.1)	127 (44)	6 (2.1)	3.0×10^{-3} (5.42, 1.60–18.32)	6.2×10^{-3}	1.1×10^{-2}
	Controls	1529 (57)	79.4 (5.3)	396 (26)	6 (0.4)	. . .	. . .	. . .
Pooled[d]	Cases	433 (59)	83.9 (5.2)	193 (45)	13 (3.0)	6.6×10^{-9} (7.45, 3.49–15.90)	5.9×10^{-5}	6.8×10^{-8}
	Controls	3903 (58)	79.1 (5.1)	1034 (26)	26 (0.7)	. . .	. . .	. . .

All p-values are based on the 2-sided alternative.

a Score tests adjusted for age and sex based on a logistic regression model. The unconditional MLE of the OR is reported based on fitting the full model including the SNP. Note, that those point estimates of OR could be inflated (**Supporting results in S1 Text**)

b P-value from the Fisher's exact test for carrier status. For the meta-analysis the data were pooled.

c P-value conditional on *APOE* ε4 carrier status. Based on a score test after adjusting for age, sex, and *APOE* ε4.

d The two cohorts are pooled in a stratified analysis (stratified by cohort).

doi:10.1371/journal.pgen.1006327.t002

expected, an analysis combining all of the AGES data on *TM2D3* demonstrated an enhanced association of the *TM2D3* variant with LOAD (OR = 7.45; 95% CI 3.49–15.90; p = 6.6x10^{-9}). We also observed an association with age-at-onset in the AGES cohorts (hazard ratio = 5.3; 95% CI 2.7–10.5; p = 1.1x10^{-6}, see also non-parametric Kaplan-Meier curves in **Fig 1**). Using the PHASE program [33–35] we estimated that rs139709573 resides on a single haplotype in the Icelandic population (**Supporting results in S1 Text**), and we further excluded potential confounding due to cryptic relatedness (**Supporting results in S1 Text**). Similarly, the association between rs139709573 and LOAD was robust to adjustment for *APOE* genotype (**Table 2**). Given the low frequency of rs139709573 in individuals of European ancestry, there were limited numbers of carriers in GERAD and ADGC (11 carriers in 13333 cases [0.041% MAF]; 4 carriers in 8256 controls [0.024% MAF]). While, the association was not statistically significant (p = 0.4), the direction of effect was consistent with that observed in AGES (OR = 1.7, 95% CI 0.5–7.3).

Analysis of *TM2D3* transcripts and expression

The *TM2D3* gene has several alternatively spliced transcripts, including six protein-coding transcripts [36]. rs139709573 falls within an exon common to all six isoforms (**S2 Fig**) and causes a Proline to Leucine amino acid change (P155L in the longest isoform). Gene and transcript expression estimates were extracted from publically available Genotype-Tissue Expression consortium (GTEx) data [37], estimated as "transcripts per million" (TPM), using the Toil workflow [38] in the UCSC Xena browser (http://xena.ucsc.edu/). *TM2D3* is expressed in all GTEx tissues, including from many brain regions (**S3 Fig**). The GTEx expression data further
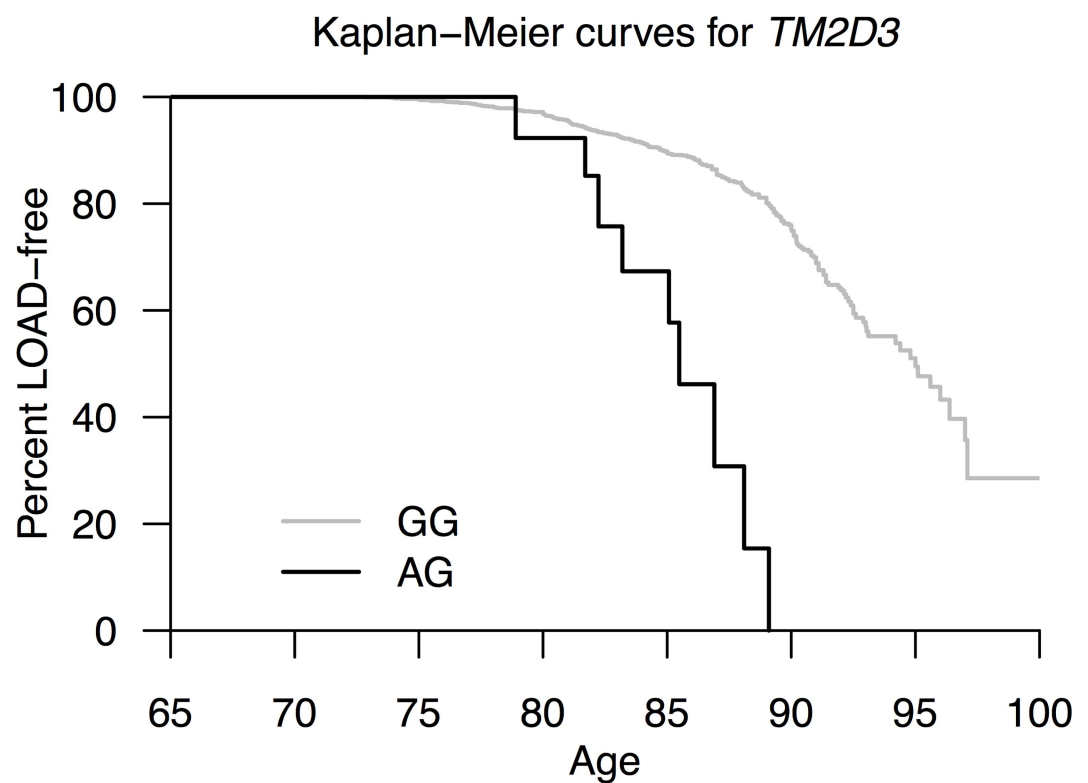

Fig 1. Kaplan-Meier survival curves for *TM2D3*. Curves are based on incident data only. There was no evidence of bias due to competing risk of death (**Supporting results in S1 Text**).

doi:10.1371/journal.pgen.1006327.g001

show that the alternative *TM2D3* transcripts are similarly expressed across diverse human tissues (**S4 Fig**).

Structure of TM2D3 protein

TM2D3 encodes a predicted double-pass transmembrane protein with evidence of evolutionary conservation (**Fig 2A**). The variant falls within the overall well-conserved C-Type Lectin domain and is adjacent to other invariant residues in the first predicted extracellular domain [32] (**Fig 2B**). While P155L is not predicted to be strongly damaging by PolyPhen [39], SIFT [40], or CADD (C-Score = 7.3) [41] and cross-species alignments show that the Proline residue is not conserved (**Fig 2A**), the two amino-acids do have important property differences, consistent with a potentially non-conservative substitution. Proline, the only cyclic amino-acid, frequently resides on the surface of folded proteins, and can underlie structural "kinks" or bends. Proline is unique in its inability to form hydrogen bonds that stabilize alpha-helices and beta-sheets. By contrast with Proline, Leucine favors alpha-helical secondary structure, and is more commonly buried in the interior of folded protein structures [42].

Functional validation of *TM2D3* (P155L) in *Drosophila*

In order to investigate the potential functional consequences of P155L and its possible link to AD, we turned to the fruit fly, *Drosophila melanogaster*. Human *TM2D3* and the homologous fly gene, *amx*, are 51% identical and 64% similar (**Fig 2A**). Mutations in *amx* cause a strong maternal effect neurogenic phenotype, characteristic of defective Notch signaling, and lead to embryonic lethality (**Fig 3A and 3B**) [43,44]. Similar to the cleavage of APP to generate the pathogenic ß-amyloid peptide, Notch receptor signaling requires homologous cleavage by γ-secretase for receptor activation. In *Drosophila*, loss of *Notch* or its regulators (e.g. *presenilin*) cause similar neurogenic phenotypes, due to impaired lateral inhibition and the inappropriate differentiation of ectodermal tissue toward a neural fate [45–47]. Moreover, prior genetic epistasis experiments suggest that *amx* may function at the γ-secretase cleavage step [32].

To determine whether *TM2D3* and *amx* have conserved molecular functions, we "humanized" the *Drosophila amx* gene by replacing its coding sequence with that of human *TM2D3* in the context of a genomic rescue construct (**Fig 3C**). Following establishment of stable transgenic lines, this construct was crossed into an *amx* mutant genetic background (*amx¹/Df(1) Exel9049*). Consistent with prior reports [43,44], when crossed to *amx* mutant males, all embryos laid by *amx* mutant females exhibited a strong neurogenic phenotype, failing to hatch from their eggs (**Fig 3B and 3D**). This embryonic defect and the resulting female sterility can be fully rescued by a genomic rescue construct with the wild-type (+) fly *amx* gene (**Fig 3D and 3E**). Due to the lethality of *Df(1)Exel9049/Y* hemizygous male progeny, complete rescue of the *amx* phenotype is predicted to lead to a maximum of ~75% egg hatching based on Mendelian expectations (**Fig 3D**, **S5 Fig**). Introduction of wild-type human *TM2D3*, under control of endogenous *amx* regulatory sequences, demonstrated significant rescue activity compared to *amx* mutant females (34.2% vs. 0% egg hatching p<0.001). Overall, based on the egg hatching assay (**Fig 3D**), the activity of the human *TM2D3* construct was estimated to be roughly half that observed for fly *amx*. Consistent with this, we observed a range in the severity of the neurogenic phenotype severity (**S6 Fig**), including embryos exhibiting a complete rescue (**Fig 3G**) and producing viable progeny that can develop into adult flies with no obvious morphological phenotypes. By contrast, an otherwise identical *TM2D3^P155L^* genomic construct, but harboring the AD-associated P155L variant, was unable to rescue the neurogenic phenotype (**Fig 3H**) or the associated female sterility (**Fig 3D**). No animals hatched out of more than 900 eggs laid from *TM2D3^P155L^* female flies. Consistent results were also seen in complementary assessments

Fig 2. Sequence alignment and domain structure of TM2D3. (A) Sequence alignment of TM2D3 homologs shows overall strong conservation but lack of conservation for P155L. Sequence alignment of human (TM2D3), mouse (Tm2d3), zebrafish (tm2d3), *Drosophila* (amx) and *C. elegans* (C41D11.9) using Clustal X2.1 is shown. The conserved amino acids are highlighted according to the standard color scheme of Clustal X. (B) Primary sequence and schematic diagram of the domain structure of human TM2D3.

of rescue activity for the *amx* peripheral nervous system neurogenic phenotype (S6 Fig). In sum, our results demonstrate that human *TM2D3* can functionally substitute for fly *amx* in the context of embryonic Notch signaling, and that the P155L variant causes a loss-of-function in this context.

Discussion

We identified a novel LOAD-associated gene, *TM2D3*, harboring a rare missense mutation (P155L) that is associated with increased risk and earlier onset of LOAD diagnosis in an Icelandic population. Our experiments in *Drosophila* further suggest that this LOAD-associated P155L variant leads to a loss-of-function of TM2D3 in the context of Notch signaling during embryogenesis. While we could neither confirm nor negate the *TM2D3* finding in available European ancestry populations–possibly because of the low P155L allele frequency, a consistent association was demonstrated in an Icelandic cohort. Although the *TM2D3* variant failed to achieve the Bonferroni significance threshold in our CHARGE-wide meta-analysis our observation of increased frequency and enhanced association within the Icelandic AGES-discovery cohort highlighted the TM2D3 variant for further consideration. Significant association in the AGES-followup sample coupled with variant functional validation in *Drosophila* suggests a role for TM2D3 in AD susceptibility. Our investigation thus exemplifies the importance not only for statistical rigor and transparency, but also flexibility in study design, when performing rare variant genetic association studies, in particular when statistical results are accompanied by relevant functional experiments.

TM2D3 has not previously been directly linked to AD. The encoded protein is predicted to contain two transmembrane regions, and additional domains with homology to C-Type lectins and G-protein coupled receptors (DRF motif) (Fig 2B). Since the P155L variant resides within the well-conserved C-Type lectin domain, the variant could plausibly disrupt the interaction of TM2D3 with a putative glycosylated binding partner. Alternatively, as suggested above, distinct structural properties of Leucine versus Proline may disrupt protein folding [42]. Interestingly, *TM2D3* shares homology with the β-amyloid peptide binding protein (BBP or *TM2D1*), which avidly binds to and sensitizes cells to toxicity caused by the β-amyloid peptide [31]. However, overexpression of TM2D3 (referred to as BLP2 in Kajkowski et al. [31]) was not associated with similar toxicity in cell culture. In mammals, TM2D3 it is highly expressed in the brain, including in the hippocampus, hypothalamus, and amygdala, potentially consistent with a role in AD pathogenesis (http://biogps.org/#goto=genereport&id=80213 and S3 Fig) [31,48,49]. Additionally, micro-array analysis suggests that the expression of *TM2D3* is down-regulated in the hippocampus of AD cases compared to controls [50].

While incompletely studied in mammals, loss-of-function mutants in the well-conserved *Drosophila* homolog of *TM2D3*, *amx*, cause embryonic phenotypes associated with disruptions in Notch signaling [51]. Further, genetic epistasis experiments support a potential role for *amx* in Notch receptor intramembranous proteolysis [32], a process that is mediated by the presenilin/γ-secretase complex. One speculative hypothesis is that TM2D3/Amx participates either directly or indirectly in the intramembranous cleavage of Notch as well as in the homologous proteolytic processing of APP by γ-secretase. In this model, the P155L variant might potentially influence the generation of toxic β-amyloid peptides, similar to well-established disease susceptibility variants in *APP*, *PSEN1*, and *PSEN2* [52].
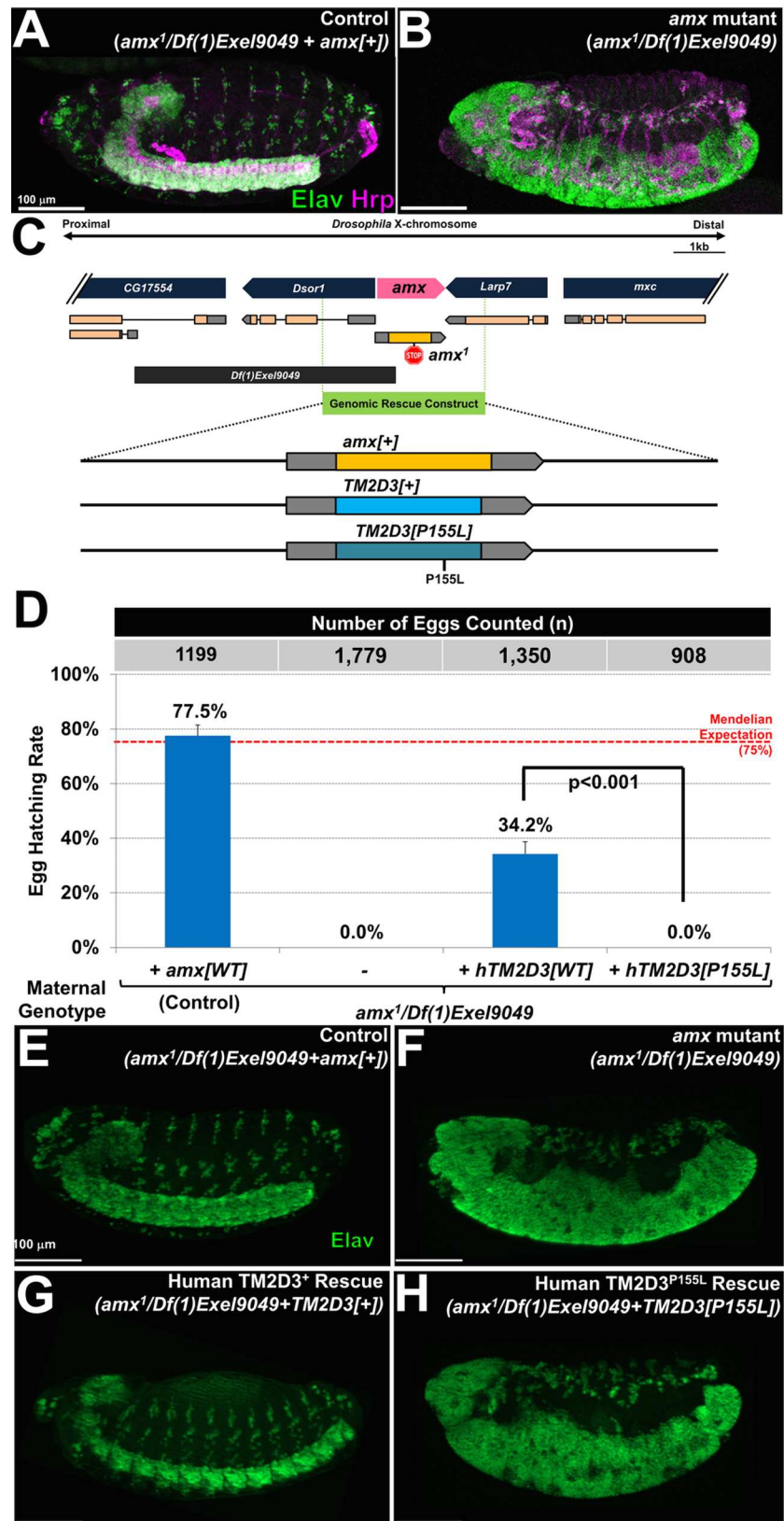
Fig 3. *TM2D3^P155L* is a loss-of-function allele. (A-B) Embryos laid by *amx* mutant females exhibit a strong neurogenic phenotype. (A) Lateral view of a late stage control embryo shows properly patterned and organized central and peripheral nervous system structures. (B) Embryos laid by amx mutant females show dramatic increase in the number of neurons, labeled by Elav (neuronal nucleus, green) and Hrp (neuronal membrane, magenta). (C) Schematic diagram of the *amx* locus and genomic rescue constructs generated for this study. *amx^1* contains an 8 nucleotide deletion that introduces a frameshift followed by a stop codon after residue 184 (Michellod and Randsholt, 2008). *Df(1)Exel9049* is a molecularly defined deletion that covers *amx* and two neighboring genes. The genomic rescue construct contains a ~3.3 kb fragment that can fully rescue the sterility of *amx^1/Df(1)Exel9049* mutant females and the neurogenic defects seen in the progeny (A, E). Coding region of Amx has been replaced by TM2D3 to "humanize" the fly *amx* gene. (D) Egg hatching assay reveals that *hTM2D3[+]* can partially suppress the female sterility of *amx^1/Df(1)Exel9049* mutant females, while *hTM2D3[P155L]* cannot. Due to the lethality of *Df(1)Exel9049/Y* hemizygous male progeny, complete rescue of the *amx* phenotype is expected to lead to a maximum of ~75% egg hatching, as denoted by the Mendelian Expectation line (also see S3 Fig). (E-H) The developing nervous system is shown for embryos laid by *amx* mutant females with and without *amx[+]*, *TM2D3[+]*, and *TM2D3[P155L]* genomic rescue constructs. *TM2D3[+]* is capable of complete rescue of the *amx* neurogenic phenotype in some embryos (G), whereas all embryos with the *hTM2D3[P155L]* construct exhibit strong neurogenic phenotypes (H). Also see S5 Fig for assessment of rescue of the *amx* peripheral nervous system neurogenic phenotype. Scale bars = 100μm.

doi:10.1371/journal.pgen.1006327.g003

Taking advantage of the potential functional conservation, we found that human TM2D3 expressed under the control of endogenous *Drosophila amx* regulatory elements is capable of rescuing the embryonic neurogenic phenotype, indicating that Amx and TM2D3 are orthologs and have conserved molecular functions *in vivo*. Importantly, we found that introduction of the LOAD-associated P155L variant into an otherwise identical *TM2D3* genomic rescue construct abolishes the rescue activity. This result is consistent with either a partial or complete loss-of-function mechanism of the P155L variant, albeit within the heterologous context of Notch signaling in the *Drosophila* embryo. It is notable that the P155L variant was neither well-conserved nor predicted by bioinformatics as damaging; nevertheless, our cross-species approach convincingly demonstrates its functional impact. Our overall strategy of first rescuing the loss-of-function phenotype of the predicted homologous gene in flies, followed by examining the potential consequences of an implicated variant can be a powerful approach for the follow-up of many similar findings emerging from human genomic studies, including others using exome genotyping arrays or next generation sequencing approaches [53,54].

In summary, we have identified a missense mutation in the *TM2D3* gene with a strong impact on LOAD risk. The *TM2D3* variant is enriched ~10-fold and associated with both risk and age-at-onset of LOAD in the Icelandic population. We further show that P155L is associated with a loss-of-function in the heterologous but potentially relevant context of Notch signaling in *Drosophila* embryos. We therefore speculate that TM2D3 may participate in the proteolytic processing of both Notch and APP, linking it to the amyloid cascade like other well-established AD susceptibility variants. Although we have demonstrated an association of the *TM2D3* variant only in the Icelandic population, our findings may thus have broader implications for understanding LOAD.

Methods

Studies and participants

Four studies from the CHARGE consortium genotyped a total of 1393 AD cases and 8141 cognitively intact controls for an ExomeChip (EC) genotyping array. All participants in the discovery phase of the analysis were of European or European American descent (Table 1). In the follow-up analysis we included Icelandic individuals from the AGES-followup cohort, African-American participants of the Cardiovascular Health Study (CHS), and European or European American individuals from the Alzheimer's Disease Genetics Consortium (ADGC) and

Genetic and Environmental Risk in Alzheimer's Disease consortium (GERAD). Further details on phenotyping and other cohort characteristics are in the **Supporting methods in S1 Text**.

Ethics statement

All participants provided informed consent and all studies were approved by their respective ethics committees.

Genotyping

All four CHARGE cohorts were genotyped for the HumanExome BeadChip v1.0 from Illumina (San Diego, CA, USA). Genotype calling and quality control was performed centrally as described previously [30]. Each study also performed quality control locally (**Supporting methods in S1 Text**). Post-analysis quality control on the top results included visual inspection of cluster plots (**S1 Fig**) and re-genotyping of *TM2D3* carriers using a TaqMan assay (genotype calls were 100% validated). *TM2D3* variant was genotyped in the AGES-followup cohort using the TaqMan assay. In ADGC and GERAD consortia the *SKAP2* and *TM2D3* variants were gentoyped on the Illumina HumanExome BeadChip v1.0 or v1.1 from Illumina (**Supporting methods in S1 Text**).

Statistical analysis

Statistical methods of the discovery phase.   In the discovery phase of our analysis we used score tests [27] for the single-variant analysis and the Sequence Kernel Association Test (SKAT [29]) to test for the aggregate effect of multiple low-frequency variants within a gene on LOAD. Each study adjusted for age, sex, and for those principal components associated with LOAD (see **Supporting methods in S1 Text**). The discovery analysis was not adjusted for *APOE* genotypes; however this important LOAD risk factor was considered in secondary analyses of top findings. Both tests were performed in R with the seqMeta package (http://cran.r-project.org/web/packages/seqMeta/) to meta-analyze study-specific scores and respective variances and covariances (see details in **Supporting methods in S1 Text**) [28]. In the single-variant analysis we included variants present in at least two studies, variants with a minor allele frequency (MAF) $\geq$ 0.5% or assuming that rare damaging alleles could be prevalent predominantly in cases we also included variants with minor allele count (MAC) $\geq$ 5 in cases. This resulted in 52,026 single-variant tests. In the SKAT analysis we grouped the variants by genes using the start and stop positions as annotated by dbNSFP v2.0 [55] to define gene boundaries. We included only variants of MAF < 5% in the combined sample, and further annotated as missense, stop-gain, stop-loss, or splice-site variants. Then we required that genes have at least two such variants as well as at least two studies having a polymorphic variant. This resulted in 11,303 SKAT tests. After filtering, the Bonferroni correction thresholds that accounted for multiple testing were 9.6×10^{-7} for single-variant tests and 4.4×10^{-6} for SKAT. Two analysts independently performed the meta-analysis, leading to identical results.

Statistical methods in follow-up analysis of *TM2D3* in AGES.   We used a score test and Fisher's exact test for the follow-up analysis based on an independent Icelandic AGES sample (AGES-followup). We report (**Table 2**) estimates of the ORs based on fitting the full model (AD ~ age + sex + *TM2D3*) instead of the one-step approximation used in seqMeta (see **Supporting Methods in S1 Text** for details). For the AGES-discovery cohort we similarly update and report estimates of the ORs (**Table 2**) based on fitting the full model (**S1A Table, S2 Table**). In the AGES study, we also used a mixed-model to account for potential confounding effect relatedness might have on the association of P155L in *TM2D3* with LOAD (**Supporting results in S1 Text**). Cox regression was used to test for association of P155L with age-at-onset

in the AGES data after adjusting for sex and *APOE* genotype. Individuals with LOAD diagnosis at the baseline visit were excluded so only at-risk individuals were included in the analysis. As detailed in the **Supporting methods in S1 Text,** the survival analysis accounted for left-truncation (i.e. follow-up begins after 65 years) and right-censoring (i.e. censoring that happens if a participant is lost to follow-up before having an event).

Drosophila

Detailed methods for our *Drosophila* experiments can be found in the **Supporting methods in S1 Text.** Briefly, the rescue activity of *amx*, *TM2D3*, or *TM2D3^P155L^* genomic constructs was examined in female flies of the genotype *amx^1/Df(1)Exel9049;{Rescue Trangene}/+* or *amx^1/Df (1)Exel9049;{Rescue Trangene}* when crossed to *amx^1* males (with or without the *{Rescue Transgene}*)[43]. To visualize the developing nervous system, resulting embryos were stained with anti-Hrp (1:1000) [56], a neuronal membrane marker and anti-ELAV (1:100) [57], a neuronal nuclear marker. For egg hatching, adults were allowed to lay eggs for 5 hours on grape juice agar plates, and larvae were counted 24 hours later.

Supporting Information

S1 Text. Supporting results and methods. Four supporting results sections and seven supporting methods sections, and sections with collaborators, and detailed funding information. (PDF)

S1 Fig. Quantile-quantile and cluster plots. Top row Quantile-quantile plots for the exome-wide discovery meta-analysis. Known genes are in orange. The genomic control coefficient (λ_{GC}) is reported. Middle panel: Variants near *APOE* excluded Bottom row Cluster plots for SNPs in *SKAP2* and *TM2D3* demonstrate appropriate calling of the rare variant genotypes. (PDF)

S2 Fig. Schematic of *TM2D3* transcripts. A schematic of *TM2D3* transcripts. Ensembl identifiers (amino acid change due to rs13970957 in parenthesis) are marked by the schematic of each protein-coding transcripts. Bold text highlights transcripts that are also in RefSeq. The exon that contains rs13970957 is marked with an arrow. The schematic was retrieved 27 May 2016 from the Ensembl browser. (PDF)

S3 Fig. Expression of *TM2D3* in GTEx samples. Overall gene expression in various tissues, brain tissues are in yellow. TPM ("transcripts per million") estimated using RSEM [58] on the GTEx data [37] (retrieved 26 May 2016 from UCSC Xena browser). (PDF)

S4 Fig. Expression of *TM2D3* protein-coding transcripts in GTEx tissues. Expression of protein-coding transcripts in all tissues (excluding blood), all brain tissues, and hippocampus only. TPM ("transcripts per million) estimated using kallisto [59] on the GTEx data [37] (retrieved 26 May 2016 from UCSC Xena browser). Transcript ENST00000559107 was not included in the plots because it is very lowly expressed and visualization was better without it. (PDF)

S5 Fig. Experimental crossing scheme for rescue experiments. The general crossing scheme is shown for the egg hatching rescue, quantified in Fig 2D. Due to the lethality of *Df(1) Exel9049/Y* hemizygous male progeny, complete rescue of the *amx* neurogenic phenotype is expected to lead to a maximum of ~75% egg hatching. (TIFF)

S6 Fig. Neurogenic phenotypes in developing flies. *TM2D3[+]* but not *TM2D3[P155L]* can suppress the neurogenic phenotype in the developing fly embryonic peripheral nervous system (PNS). (A-B) Maternal effect neurogenic defects in *amx* mutants can also be seen in the PNS. Compared to control embryos (A), maternal *amx* mutant embryos (B) show increased number of Sensory Organ Precursor cells (SOPs), labeled by Senseless (Sens, red), due to defective lateral inhibition. Embryos are counterstained with DAPI (blue). (C-F) PNS phenotypes in embryos laid by *amx* mutant females with or without *amx[+]*, *TM2D3[+]*, and *TM2D3[P155L]* genomic rescue constructs. Thoracic segments (T1, T2, T3) of stage 11 embryos are shown. *TM2D3[+]* can rescue the lateral inhibition defects in SOPs in some (Compare C, D and E) but not all embryos (E'). In contrast, all embryos from *amx* mutant females with *TM2D3[P155L]* exhibit neurogenic PNS phenotypes (F, F'). Scale bars = 100μm for A-B, = 50μm for C-F. (TIFF)

S1 Table. Results of exome-wide discovery analysis. Top results of exome-wide single variant (**A**) and SKAT (**B**) discovery analysis. Complete results are available in dbGAP (accession phs000930). A: Direction = AGES-CHS-FHS-RS, Alleles = non-coding/coding, AF displayed as percentage (%). B: cumAF = cumulative allele frequency displayed as percentage (%), N_SNPs = number of SNPs in test. (PDF)

S2 Table. Variant- and study-specific break-down of *TM2D3*, *SKAP2*, and *APOE* results. (PDF)

S3 Table. Allele frequency of P155L in populations of European ancestry. (PDF)

Author Contributions

Conceptualization: CMvD VG JMS SY SS LJL EB.

Methodology: JJ SJvdL TA JCB VC JAB KMR CMvD VG SS LJL JMS SY DLK.

Software: JJ SJvdL JCB VC ALD JAB KMR.

Validation: JJ SJvdL MLG SY DLK JLS.

Formal analysis: JJ SJvdL NA JCB JAB VC AN MV RS ALD AVS CAIV SHC CLS TVV CH SR MAI KMR TA LSW.

Investigation: SY DLK JLS HJB JMS MLG JIR BMP EB ALF VG SS CMvD JJ SJvdL LJL VC JCB.

Resources: CMvD SS LJL VG LSW OLL AB MAI MEG GE DJP PWF GH OR JHJ VS DL AGU JIR ALF AH JW GDS BMP SY CJO EB HJB JMS.

Writing original draft: JJ CMvD SJvdL JMS SS LJL VG.

Writing review & editing: JJ SJvdL JCB VC DLK SY MLG AN MV JLS ALD JAB AVS NA RS CAIV SHC CLS OLL AB MAI MEG CH TVV SR PWF GH OR JHJ DJP VS GE KMR HJB DL AGU VE JIR TA CJO ALF LJL AH LSW JW GDS EB BMP SS JMS VG CMvD.

Visualization: JJ SJvdL SY DLK JLS MLG.

Supervision: CMvD VG JMS SS LJL BMP.

Project administration: CMvD VG JMS SS LJL ALF EB JIR BMP.

References

1. Barker WW, Luis CA, Kashuba A, Luis M, Harwood DG, Loewenstein D, et al. Relative frequencies of Alzheimer disease, Lewy body, vascular and frontotemporal dementia, and hippocampal sclerosis in the State of Florida Brain Bank. Alzheimer Dis Assoc Disord. 2002; 16: 203–12. Available: http://www.ncbi.nlm.nih.gov/pubmed/12468894 doi: 10.1097/00002093-200210000-00001 PMID: 12468894

2. Hebert LE, Weuve J, Scherr PA, Evans DA. Alzheimer disease in the United States (2010–2050) estimated using the 2010 census. Neurology. 2013; 80: 1778–83. doi: 10.1212/WNL.0b013e31828726f5 PMID: 23390181

3. 2015 Alzheimer's disease facts and figures. Alzheimers Dement. 2015; 11: 332–84. Available: http://www.ncbi.nlm.nih.gov/pubmed/25984581 doi: 10.1016/j.jalz.2015.02.003 PMID: 25984581

4. James BD, Leurgans SE, Hebert LE, Scherr PA, Yaffe K, Bennett DA. Contribution of Alzheimer disease to mortality in the United States. Neurology. 2014; 82: 1045–50. doi: 10.1212/WNL.0000000000000240 PMID: 24598707

5. Cruts M, Theuns J, Van Broeckhoven C. Locus-specific mutation databases for neurodegenerative brain diseases. Hum Mutat. 2012; 33: 1340–4. doi: 10.1002/humu.22117 PMID: 22581678

6. Bekris LM, Yu C-E, Bird TD, Tsuang DW. Genetics of Alzheimer disease. J Geriatr Psychiatry Neurol. 2010; 23: 213–27. doi: 10.1177/0891988710383571 PMID: 21045163

7. Naj AC, Jun G, Beecham GW, Wang L-S, Vardarajan BN, Buros J, et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. Nat Genet. 2011; 43: 436–41. doi: 10.1038/ng.801 PMID: 21460841

8. Hollingworth P, Harold D, Sims R, Gerrish A, Lambert J-C, Carrasquillo MM, et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. Nat Genet. 2011; 43: 429–35. doi: 10.1038/ng.803 PMID: 21460840

9. Lambert J-C, Heath S, Even G, Campion D, Sleegers K, Hiltunen M, et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. Nat Genet. 2009; 41: 1094–9. doi: 10.1038/ng.439 PMID: 19734903

10. Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nat Genet. 2009; 41: 1088–93. doi: 10.1038/ng.440 PMID: 19734902

11. Seshadri S, Fitzpatrick AL, Ikram MA, DeStefano AL, Gudnason V, Boada M, et al. Genome-wide analysis of genetic loci associated with Alzheimer disease. JAMA. 2010; 303: 1832–40. doi: 10.1001/jama.2010.574 PMID: 20460622

12. Lambert J-C, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2013; 45: 1452–8. doi: 10.1038/ng.2802 PMID: 24162737

13. Saunders AM, Strittmatter WJ, Schmechel D, St. George-Hyslop PH, Pericak-Vance MA, Joo SH, et al. Association of apolipoprotein E allele 4 with late-onset familial and sporadic Alzheimer's disease. Neurology. 1993; 43: 1467–1467. doi: 10.1212/WNL.43.8.1467 PMID: 8350998

14. Strittmatter WJ, Saunders AM, Schmechel D, Pericak-Vance M, Enghild J, Salvesen GS, et al. Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. Proc Natl Acad Sci U S A. 1993; 90: 1977–81. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=46003&tool=pmcentrez&rendertype=abstract doi: 10.1073/pnas.90.5.1977 PMID: 8446617

15. Chouraki V, Seshadri S. Genetics of Alzheimer's disease. Adv Genet. 2014; 87: 245–94. doi: 10.1016/B978-0-12-800149-3.00005–6 PMID: 25311924

16. Jonsson T, Atwal JK, Steinberg S, Snaedal J, Jonsson P V, Bjornsson S, et al. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. Nature. 2012; 488: 96–9. doi: 10.1038/nature11283 PMID: 22801501

17. Jonsson T, Stefansson H, Steinberg S, Jonsdottir I, Jonsson P V, Snaedal J, et al. Variant of TREM2 associated with the risk of Alzheimer's disease. N Engl J Med. 2013; 368: 107–16. doi: 10.1056/NEJMoa1211103 PMID: 23150908

18. Guerreiro R, Wojtas A, Bras J, Carrasquillo M, Rogaeva E, Majounie E, et al. TREM2 variants in Alzheimer's disease. N Engl J Med. 2013; 368: 117–27. doi: 10.1056/NEJMoa1211851 PMID: 23150934

19. Cruchaga C, Karch CM, Jin SC, Benitez BA, Cai Y, Guerreiro R, et al. Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. Nature. 2014; 505: 550–4. doi: 10.1038/nature12825 PMID: 24336208

20. Wetzel-Smith MK, Hunkapiller J, Bhangale TR, Srinivasan K, Maloney JA, Atwal JK, et al. A rare mutation in UNC5C predisposes to late-onset Alzheimer's disease and increases neuronal cell death. Nat Med. 2014; 20: 1452–7. doi: 10.1038/nm.3736 PMID: 25419706

21. Logue MW, Schu M, Vardarajan BN, Farrell J, Bennett DA, Buxbaum JD, et al. Two rare AKAP9 variants are associated with Alzheimer's disease in African Americans. Alzheimers Dement. 2014; 10: 609–618.e11. doi: 10.1016/j.jalz.2014.06.010 PMID: 25172201

22. Zhang B, Gaiteri C, Bodea L-G, Wang Z, McElwee J, Podtelezhnikov AA, et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. Cell. Elsevier; 2013; 153: 707–20. doi: 10.1016/j.cell.2013.03.030 PMID: 23622250

23. Ulrich JD, Finn MB, Wang Y, Shen A, Mahan TE, Jiang H, et al. Altered microglial response to Aβ plaques in APPPS1-21 mice heterozygous for TREM2. Mol Neurodegener. 2014; 9: 20. doi: 10.1186/1750-1326-9-20 PMID: 24893973

24. van der Lee SJ, Holstege H, Wong TH, Jakobsdottir J, Bis JC, Chouraki V, et al. PLD3 variants in population studies. Nature. 2015; 520: E2–3. doi: 10.1038/nature14038 PMID: 25832410

25. Kiezun A, Garimella K, Do R, Stitziel NO, Neale BM, McLaren PJ, et al. Exome sequencing and the genetic basis of complex traits. Nat Genet. 2012; 44: 623–30. doi: 10.1038/ng.2303 PMID: 22641211

26. Psaty BM, O'Donnell CJ, Gudnason V, Lunetta KL, Folsom AR, Rotter JI, et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. Circ Cardiovasc Genet. 2009; 2: 73–80. doi: 10.1161/CIRCGENETICS.108.829747 PMID: 20031568

27. Radhakrishna Rao C, Bartlett MS. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. Math Proc Cambridge Philos Soc. Cambridge University Press; 2008; 44: 50. doi: 10.1017/S0305004100023987

28. Lumley T, Brody J, Dupuis J, Cupples A. Meta-analysis of a rare-variant association test. [Internet]. 2012. Available: http://stattech.wordpress.fos.auckland.ac.nz/files/2012/11/skat-meta-paper.pdf

29. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet. 2011; 89: 82–93. doi: 10.1016/j.ajhg.2011.05.029 PMID: 21737059

30. Grove ML, Yu B, Cochran BJ, Haritunians T, Bis JC, Taylor KD, et al. Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. PLoS One. 2013; 8: e68095. doi: 10.1371/journal.pone.0068095 PMID: 23874508

31. Kajkowski EM, Lo CF, Ning X, Walker S, Sofia HJ, Wang W, et al. beta -Amyloid peptide-induced apoptosis regulated by a novel protein containing a g protein activation module. J Biol Chem. 2001; 276: 18748–56. doi: 10.1074/jbc.M011161200 PMID: 11278849

32. Michellod M-A, Randsholt NB. Implication of the Drosophila beta-amyloid peptide binding-like protein AMX in Notch signaling during early neurogenesis. Brain Res Bull. 2008; 75: 305–9. doi: 10.1016/j.brainresbull.2007.10.060 PMID: 18331889

33. Stephens M, Scheet P. Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. Am J Hum Genet. 2005; 76: 449–62. doi: 10.1086/428594 PMID: 15700229

34. Stephens M, Donnelly P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet. 2003; 73: 1162–9. doi: 10.1086/379378 PMID: 14574645

35. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet. 2001; 68: 978–89. doi: 10.1086/319501 PMID: 11254454

36. Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, et al. Ensembl 2016. Nucleic Acids Res. Oxford University Press; 2016; 44: D710–D716. doi: 10.1093/nar/gkv1157 PMID: 26687719

37. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. Nat Genet. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2013; 45: 580–5. doi: 10.1038/ng.2653 PMID: 23715323

38. Vivian J, Rao A, Nothaft FA, Ketchum C, Armstrong J, Novak A, et al. Rapid and efficient analysis of 20,000 RNA-seq samples with Toil [Internet]. bioRxiv. Cold Spring Harbor Labs Journals; 2016 Jul. doi: 10.1101/062497

39. 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; 7: 248–9. doi: 10.1038/nmeth0410-248 PMID: 20354512

40. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009; 4: 1073–81. doi: 10.1038/nprot.2009.86 PMID: 19561590

41. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014; 46: 310–5. doi: 10.1038/ng.2892 PMID: 24487276

42. Barnes MB, Gray IC. Bioinformatics for Geneticists [Internet]. 2003 [cited 11 May 2016]. Available: http://www.amazon.com/exec/obidos/tg/detail/-/0470843934/qid=1049297099/sr=8-7/ref=sr_8_7/103-1037454-5494226?v=glance&s=books&n=507846

43. Shannon MP. Characterization of the female-sterile mutant Almondex of Drosophila melanogaster. Genetica. 1972; 43: 244–256. doi: 10.1007/BF00123632 PMID: 4625164

44. Lehmann R, Dietrich U, Jimènez F, Campos-Ortega JA. Mutations of early neurogenesis inDrosophila. Wilhelm Roux's Arch Dev Biol. 1981; 190: 226–229. doi: 10.1007/BF00848307

45. Poulson DF. Chromosomal Deficiencies and the Embryonic Development of Drosophila Melanogaster. Proc Natl Acad Sci U S A. 1937; 23: 133–7. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1076884&tool=pmcentrez&rendertype=abstract doi: 10.1073/pnas.23.3.133 PMID: 16588136

46. Struhl G, Greenwald I. Presenilin is required for activity and nuclear access of Notch in Drosophila. Nature. 1999; 398: 522–5. doi: 10.1038/19091 PMID: 10206646

47. Goriely A, Dumont N, Dambly-Chaudière C, Ghysen A. The determination of sense organs in Drosophila: effect of the neurogenic mutations in the embryo. Development. 1991; 113: 1395–404. Available: http://www.ncbi.nlm.nih.gov/pubmed/1811951 PMID: 1811951

48. Wu C, Orozco C, Boyer J, Leglise M, Goodale J, Batalov S, et al. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. Genome Biol. 2009; 10: R130. doi: 10.1186/gb-2009-10-11-r130 PMID: 19919682

49. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci U S A. 2004; 101: 6062–7. doi: 10.1073/pnas.0400782101 PMID: 15075390

50. Zhang L, Guo XQ, Chu JF, Zhang X, Yan ZR, Li YZ. Potential hippocampal genes and pathways involved in Alzheimer's disease: a bioinformatic analysis. Genet Mol Res. 2015; 14: 7218–32. doi: 10.4238/2015.June.29.15 PMID: 26125932

51. Michellod M-A, Forquignon F, Santamaria P, Randsholt NB. Differential requirements for the neurogenic gene almondex during Drosophila melanogaster development. Genesis. 2003; 37: 113–22. doi: 10.1002/gene.10233 PMID: 14595834

52. Karch CM, Cruchaga C, Goate AM. Alzheimer's disease genetics: from the bench to the clinic. Neuron. 2014; 83: 11–26. doi: 10.1016/j.neuron.2014.05.041 PMID: 24991952

53. Shulman JM. Drosophila and experimental neurology in the post-genomic era. Exp Neurol. 2015; doi: 10.1016/j.expneurol.2015.03.016 PMID: 25814441

54. Bellen HJ, Yamamoto S. Morgan's legacy: fruit flies and the functional annotation of conserved genes. Cell. 2015;In press. https://www.ncbi.nlm.nih.gov/pubmed/26406362

55. Liu X, Jian X, Boerwinkle E. dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. Hum Mutat. 2013; 34: E2393–402. doi: 10.1002/humu.22376 PMID: 23843252

56. Snow PM, Patel NH, Harrelson AL, Goodman CS. Neural-specific carbohydrate moiety shared by many surface glycoproteins in Drosophila and grasshopper embryos. J Neurosci. 1987; 7: 4137–44. Available: http://www.ncbi.nlm.nih.gov/pubmed/3320283 PMID: 3320283

57. Robinow S, White K. Characterization and spatial distribution of the ELAV protein during Drosophila melanogaster development. J Neurobiol. 1991; 22: 443–61. doi: 10.1002/neu.480220503 PMID: 1716300

58. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. BioMed Central; 2011; 12: 323. doi: 10.1186/1471-2105-12-323 PMID: 21816040

59. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2016; 34: 525–527. doi: 10.1038/nbt.3519 PMID: 27043002