OXFORD

ASSOCIATION STUDIES ARTICLE

Differential burden of rare protein truncating variants in Alzheimer's disease patients compared to centenarians

Yun Freudenberg-Hua[1,2,*], Wentian Li[3], Avinash Abhyankar[4], Vladimir Vacic[4], Vanessa Cortes[1], Danny Ben-Avraham[5], Jeremy Koppel[1], Blaine Greenwald[2], Soren Germer[4], T2D-GENES Consortium, Robert B. Darnell[4], Nir Barzilai[5], Jan Freudenberg[3,6], Gil Atzmon[5,7] and Peter Davies[1]

[1]Litwin-Zucker Research Center for the Study of Alzheimer's Disease, The Feinstein Institute for Medical Research, Northwell Health, Manhasset, NY 11030, USA, [2]Division of Geriatric Psychiatry, Zucker Hillside Hospital, Northwell Health, Glen Oaks, NY 11004, USA, [3]Robert S Boas Center for Genomics and Human Genetics, The Feinstein Institute for Medical Research, Northwell Health, Manhasset, NY 11030, USA, [4]New York Genome Center, New York, NY 10013, USA, [5]Institute for Aging Research and the Diabetes Research Center, Department of Medicine and Genetics, Albert Einstein College of Medicine, Bronx, NY 10461, USA [6]The Regeneron Genetics Center, Tarrytown, NY 10591, USA and [7]Faculty of Natural Sciences, University of Haifa, Haifa 31905, Israel

*To whom correspondence should be addressed at: Tel: 1.516.5623492; Fax:1.516.5620401; Email: yfreuden@northwell.edu

Abstract

We compared coding region variants of 53 cognitively healthy centenarians and 45 patients with Alzheimer's disease (AD), all of Ashkenazi Jewish (AJ) ancestry. Despite the small sample size, the known AD risk variant APOE4 reached genome-wide significance, indicating the advantage of utilizing 'super-controls'. We restricted our subsequent analysis to rare variants observed at most once in the 1000 Genomes database and having a minor allele frequency below 2% in our AJ sample. We compared the burden of predicted protein altering variants between cases and controls as normalized by the level of rare synonymous variants. We observed an increased burden among AD subjects for predicted loss-of-function (LoFs) variants defined as stop-gain, frame shift, initiation codon (INIT) and splice site mutations ($n = 930$, OR $= 1.3$, $P = 1.5 \times E-5$). There was no enrichment across all rare protein altering variants defined as missense plus LoFs, in frame indels and stop-loss variants ($n = 13\,014$, OR $= 0.97$, $P = 0.47$). Among LoFs, the strongest burden was observed for INIT (OR $= 2.16$, $P = 0.0097$) and premature stop variants predicted to cause non-sense-mediated decay in the majority of transcripts (NMD) (OR $= 1.98$, $P = 0.02$). Notably, this increased burden of NMD, INIT and splice variants was more pronounced in a set of 1397 innate immune genes (OR $= 4.55$, $P = 0.0043$). Further comparison to additional exomes indicates that the difference in LoF burden originated both from the AD and centenarian sample. In summary, we observed an overall increased burden of rare LoFs in AD subjects as compared to centenarians, and this enrichment is more pronounced for innate immune genes.

Introduction

Alzheimer's disease (AD) affects an estimated 35 million people worldwide (1). It has been reported that 24–33% of total phenotypic variance for late onset AD could be explained by common variants (2,3). This leaves room for considerable proportion of genetic risk factors yet to be identified that cannot be attributed to additive effects of common variants. Accordingly, there has been an increasing focus on the potential role of infrequent (minor allele frequency [(MAF) 1–5%] and rare variants (MAF < 1%) in the etiology of AD and other common diseases (4). The recent findings that rare variants such as *TREM2* R47H, *APP* A673T, *PLD3* V232M and *UNC5C* T835M (5–9) influence the risk for AD support this hypothesis. Similar sequencing approaches have also identified genes harboring rare disease variants in other neurodegenerative diseases such as amyotrophic lateral sclerosis (10) and Parkinson's disease with dementia (11), as well as other complex diseases such as myocardial infarction (12). Thus, it is reasonable to hypothesize that many unidentified rare risk variants exist for AD, but their identification will require very large samples to be sequenced. One way of increasing the power to detect association signals of rare variants is to study extreme phenotypes (13–15). Given the high prevalence of AD among people over the age of 85 years (32%) (16), cognitively healthy centenarians (CEN) may be viewed as both an extreme phenotype for longevity and as cognitive 'super controls'. The advantage in terms of statistical power when using super controls may become even more pronounced in a homogenous study population such as the Ashkenazi Jewish (AJ) population, where risk variants may have drifted to higher frequencies (17).

Due to the limited power of rare variants to achieve statistical significance on their own, a variety of burden tests (allelic collapsing or allelic sum tests) has been developed that aggregate sets of variants into a joint test (18). These burden tests are generally performed as gene-based tests because genes are intuitively meaningful elements for aggregation of rare variants. However, even such gene-based tests are expected to require sample sizes far larger than available here to discover novel associations. On the other hand, there is no a priori reason not to aggregate rare variants across higher functional units than genes, such as pathways, or across the whole genome. One such functional unit of particular interest is immune genes that have recently gained much traction for a role in the pathogenesis of AD. A possible link between inflammation and AD had been suggested by pathophysiological evidences long before the genomic era (19,20) and is underscored by the protective effect of long-term use of non-steroidal anti-inflammatory drugs (21). As a possible mechanism, a large body of evidence has provided more details on how microglial activation might play a crucial role in mediating amyloid-β aggregate-induced toxicity (22). More recently, genome-wide association studies (GWAS) have identified common genetic risk variants in genes involved in the immune system function and inflammation (23,24) including *CLU, CR1, CD33, CD2AP, MS4A6A, MS4A4E, ABCA7, EPHA1, HLA-DRB5, HLA-DRB1, INPP5D* and *MEF2C* (24,25). In parallel, the identification of the risk variant *TREM2* R47H with an effect size comparable to that of *APOE4* allele highlights the potential impact of low-frequency variants in immune genes (5,6). The gene *TREM2* encodes a microglial immunoreceptor that functions as a key regulator of microglial activation in response to tissue damage (26). The role of immunity in AD was also implicated in a recent pathway analysis based on GWAS data (27) and differential gene expression data from both human (28) and mice (29). Thus, instead of aggregating rare variants in a gene-based manner, it appears to be promising to aggregate them over larger sets of immune system genes.

In this study, we investigated the burden of rare loss-of-function (LoF) variants in a whole genome sequencing (WGS) study of 45 subjects with AD and 53 cognitively healthy CEN of AJ ancestry. We subsequently evaluated the strength of the observed enrichment signals in genes with an annotated function in immunity and inflammation. In addition, we also tested for an increased burden in genes that are highly expressed in the human brain as a positive control and in genes that have no known phenotype associations as a negative control.

Results

Our relatively small sample consisting of 45 AD subjects and 53 CEN is underpowered for detecting single variants with weak effects. We nevertheless observed an association signal of the known AD risk allele APOE4 (rs429358) that reaches genome-wide significance with a Fisher's exact $P = 1.02 \times E-08$ and an OR of 10.6 (95% CI: 4.20–26.8), which tends to be on the higher end compared to the reported ORs of around three in case–control studies (ranging from 1.8 to 9) (30–32). This observation may be attributed to the fact that APOE is also the major gene that has been identified for longevity (33–35), highlighting that aging may not be entirely separated from late-onset AD. We also replicated the protective APOE2 allele (rs7412) with an OR of 0.25 (95% CI: 0.068–0.895; $P = 0.034$), which was associated with longevity (35,36). We did not observe any new risk variants that reach genome-wide significance threshold of $P \leq 5E-08$. We performed a systematic survey of all single-nucleotide polymorphisms (SNPs) associated with AD ($P \leq 5E-08$) from the National Human Genome Research Institute (NHGRI) GWAS Catalog (37) as observed in our dataset (Supplementary Table S1). The replication of the APOE association confirms validity of the phenotypes of our sample and supports that our design may magnify the power to detect further potentially causal variants that distinguish individuals with AD from those with unusual long life span.

Among variants showing an association signal of Fisher's exact test P < 0.05, 16 non-synonymous variants with MAF ≤ 5% in the 1000 Genomes Project (1KG) dataset are predicted as being potentially functional by meeting at least four out of the following five criteria implemented in the dbNSFP database (38): SIFT (Damaging), Polyphen2 HDIV (Probably damaging or Possibly damaging), Polyphen2 HVAR (Probably damaging or Possibly damaging), LRT (Deleterious) and MutationTaster (Disease causing automatic or Disease causing) (Supplementary Table S2). Nine of these variants are found only in AD cases, whereas only two are found exclusively in unaffected CEN.

In order to test whether rare variants with potentially high functional impact are enriched among AD subjects, we focused our subsequent analysis on the 19 885 rare coding variants with very low frequency in 1KG (MAF ≤ 0.025%) and low frequency (MAF ≤ 2%) in our AJ sample. To test for relative enrichment, we compare sum of minor allele counts of rare non-synonymous variants with that of rare synonymous variants between AD and CEN subjects. Because we have no reason to expect sequencing errors to affect rare synonymous and rare non-synonymous variants differentially, this comparison would correct for potential sequencing errors still remaining after the QC steps, as well as possible undetected batch effects. We group variants according to predicted functional impact on protein function into predicted LoF variants defined as non-sense,

Table 1. Enrichment of rare PAVs and LoFs among all genes

Variant class	Counts, AD	Counts, CEN	OR	95% CI	Fisher's exact P-value	Empirical P-value
Rare_syn ($n = 6538$)	4240	5148	NA	NA	NA	NA
PAVs ($n = 13\,014$)	8411	10 022	1.02	0.97–1.07	0.468	NA
LoFs ($n = 930$)	654	612	1.30	1.15–1.46	1.50E−05	1.10E−05
LoFs_non-indel ($n = 474$)	331	322	1.25	1.06–1.47	0.0065	0.003
LoFs_indel ($n = 456$)	323	290	1.35	1.14–1.60	0.00031	0.00038
Stop-gain ($n = 313$)	208	225	1.12	0.92- 1.37	0.256	NA
Stop-gain, non-indel ($n = 301$)	199	216	1.12	0.91–1.37	0.268	NA
Stop-gain_NMD ($n = 38$)	31	19	1.98	1.08–3.7	0.022	0.012
FS ($n = 419$)	290	264	1.33	1.12–1.59	0.0011	0.0014
INIT ($n = 31$)	32	18	2.16	1.17- 4.09	0.0097	0.0071
Splicing ($n = 168$)	124	105	1.43	1.09–1.88	0.0072	0.0018

Enrichment of rare PAVs and LoFs among all genes, comparing allele counts of variants with a certain functional classification with rare synonymous variants between AD patients and CEN. One-sided empirical P-values are calculated by 1 million permutations of case–control status. In parenthesis are total numbers of variants in each variant class. Rare_syn: rare synonymous variants, Stop-gain_NMD: stop-gain variants that are predicted to affect > 50% of known transcripts. LoFs defined as non-sense, frame shift, initiation codon and splice site variants. PAVs defined as LoFs plus missense, in frame indels, stop-loss variants.

frame shift, initiation codon (INIT) and canonical splice site mutations (Supplementary Table S3a) and predicted protein altering variants (PAVs) are defined as LoFs plus missense, in frame indels, stop-loss variants. We test enrichment for PAVs and LoFs as well as for individual functional classes as listed above.

Compared to 6538 rare synonymous variants, we did not find any differential enrichment of alternative allele counts for the 13 014 PAVs among AD or CEN subjects (OR = 1.02, Fisher's exact test $P = 0.47$) (Table 1). However, we found LoFs ($n = 930$) to be enriched among AD subjects (OR = 1.30, Fisher's exact test $P = 1.5 \times E{-}5$) (Table 1 and Supplementary Table S3a). The strongest enrichments were observed among INIT variants (OR = 2.16, $P = 0.0097$) and splicing variants (OR = 1.43, $P = 0.0072$) (Table 1). We found higher counts for both non-indel- and indel-type LoFs among AD subjects, indicating that the enrichment signal was not driven by indel variants. For the 313 stop-gain variants, we found no enrichment in either group. However, when we looked at the 38 stop-gain variants that are predicted to cause non-sense-mediated decay (NMD) in >50% of known transcripts (stop-gain_NMD) of the respective gene, we observed a higher mutation burden of this type of variants in AD subjects (OR = 1.98, $P = 0.022$) (Table 1 and Supplementary Table S3b). To ensure that the enrichment signals were not driven by variants in one or a few individuals, we also performed 1 million permutations by randomly shuffling the case–control labels and counting the percentage of ORs that are larger than the observed value for those variant categories showing significant enrichment. This resulted in empirical P-values that are similar to the Fisher's exact test (Table 1).

McCarthy et al. (39) reported that choice of transcripts and software may affect annotations for putative LoF variants. To investigate whether an alternative annotation tool would yield similar results, we used Variant Effect Predictor (VEP) from Ensembl to predict LoF variants. Despite small differences in the total counts of variants (996 LoF and 6552 synonymous variants), we observed the same magnitude of LoF enrichment among AD patients compared to CEN (OR = 1.29, Fisher's exact test $P = 9.5 \times E{-}6$).

Next, we tested whether the burden of rare LoFs is higher in immune genes, paying specific attention to genes annotated to play a role in the innate immune system. To this end, we obtained a publicly available dataset consisting of 4677 genes annotated with immunological function and 1397 genes annotated to play a role in the innate immune system [Gene Ontology (GO) ID: 0045087, innate immune response] from innateDB database (40). We found LoFs enriched in immune system genes, with an increased cumulative burden especially for INIT and splicing variants (Table 2). Interestingly, among innate immune genes, the set of LoFs with higher functional impact consisting of INIT, splicing and stop-gain mutations predicted to cause NMD showed a particularly strong enrichment signal (OR = 4.55, 95% CI: 1.45–18.8, $P = 0.0043$) (Table 2). The list of innate immune genes affected by stop-gain_NMD, INIT and splicing variants is presented in the Supplementary Table S4a. We observed similar enrichment when we normalized the counts for rare LoFs with the sum of rare synonymous alleles found only in the subset of innate immune genes (OR = 4.61, 95% CI: 1.45–19.3, $P = 0.0041$; Supplementary Table S4b).

In order to test whether the observed enrichment signals are specific for CNS diseases, we investigated the enrichment for LoFs in a group of genes that may be considered as a positive control. Intuitively, we consider genes that are more highly expressed in the human brain than in the other tissues form a set of candidate genes for any type of brain diseases, including AD. To this end, we retrieved a list of 2472 human genes, which were annotated as orthologs of mouse essential genes (EG) and 3811 human orthologs of genes with known non-lethal phenotypes (NLG) in homozygous mouse mutants (41). Among them 2183 genes were reported to be highly expressed in the human brain tissues as compared to other tissues (BRAIN) by Georgi et al. (41). We performed rare variants burden testing using the same set of rare synonymous variants described above as normalization. As expected from the above results for PAVs, no overall difference was observed for rare PAVs in BRAIN genes or in the 398 genes that are both EG and BRAIN (Supplementary Table S5a). However, an enrichment of rare LoFs was seen in AD subjects and the burden was most pronounced in genes that are classified as being both EG and BRAIN genes (OR = 3.26, $P = 0.030$) (Supplementary Table S5a and S5b). None of the three stop-gain variants found among these 12 genes were predicted to

Table 2. Enrichment of rare LoFs among genes involved in immune function and innate immune system

Gene Groups	Variant class	AD	CEN	OR	95% CI	Fisher's exact P-value
Immune genes (n = 4677)	PAVs (n = 2915)	1965	2151	1.11	1.03–1.19	0.0058
	LoFs (n = 186)	137	120	1.39	1.07–1.79	0.011
	Stop-gain (n = 62)	35	49	0.87	0.54–1.37	0.58
	Stop-gain_NMD + INIT + splicing (n = 46)	40	26	1.87	1.11–3.19	0.013
	INIT + splicing (n = 41)	38	23	2.01	1.16–3.53	0.0094
	INIT (n = 6)	7	3	2.83	0.65–16.99	0.20
	Splicing (n = 35)	31	20	1.88	1.04–3.49	0.033
Innate immune genes (n = 1397)	PAVs (n = 815)	526	602	1.06	0.94–1.20	0.36
	LoFs (n = 54)	39	33	1.43	0.88–2.36	0.15
	Stop-gain_NMD + INIT + splicing (n = 16)	15	4	4.55	1.45–18.85	0.0043

Enrichment of rare PAVS and LoFs among genes involved in immune function and specifically in innate immune function by comparing allele counts of PAVS and LoFs to allele counts of rare synonymous variants in all genes between AD patients and CEN. In parenthesis are total numbers of variants in each variant class. Rare_syn: rare synonymous variants, Stop-gain_NMD: stop-gain variants that are predicted to affect > 50% of known transcripts. LoFs defined as non-sense, frame shift, initiation codon and splice site variants. PAVs defined as LoFs plus missense, in frame indels, stop-loss variants.

Table 3. Enrichment of rare LoFs among all genes in AD, CEN and AJ exomes

Variant class	AD	AJWES	OR	95% CI	Fisher's exact P-value
Rare_syn	4240	22 563	NA	NA	NA
Rare_LoFs	654	3057	1.14	1.04–1.25	0.0056
	CEN	AJWES	OR	95% CI	Fisher's exact P-value
Rare_syn	5148	22 563	NA	NA	NA
Rare_LoFs	612	3057	0.88	0.80–0.96	0.0052

We test for enrichment of aggregated allele counts of rare LoFs among all genes normalized by aggregated allele counts of rare synonymous variants between AD patients, CEN and non-centenarian control AJ exomes from T2D GENES (AJWES). Rare_syn: rare synonymous variants. LoFs defined as non-sense, frame shift, initiation codon and splice site variants, PAVs defined as LoFs plus missense, in frame indels, stop-loss variants.

cause NMD. As a negative control, we performed burden test in the set of 'low-impact' genes that neither belong to any of the three functional categories (i.e. EG, NLG and BRAIN) nor being associated with any human disease traits according to Online Mendelian Inheritance in Man (OMIM) (42) and found no significant enrichment of the 285 LoFs in AD or CEN subjects (Supplementary Table S5a).

To find out whether LoFs are enriched among AD subjects, or whether they are depleted among CEN, in relation to a broader population sample, we looked at the proportion of LoFs that are also found in the non-Finnish European and non-psychiatric dataset from the Exome Aggregation Consortium (ExAC-NFE) (http://biorxiv.org/content/early/2015/10/30/030338, last accessed March 2016). Out of all 930 rare LoFs identified in our WGS data, 354 SNVs (74.7% of all SNV LoFs) and 333 indels (73.0% of all indel LoFs) were found in ExAC-NFE. We compared the respective proportions of LoFs ascertained exclusively in AD (AD-LoFs) and those ascertained exclusively in CEN (CEN-LoFs) being present in the ExAC-NFE. A total of 278 out of 407 AD-LoFs (68.3%) and 281 out of 372 CEN-LoFs (75.5%) were present in ExAC-NFE (OR = 0.70, CI: 0. 50–0.97, P = 0.026). Thus, AD-LoFs are depleted in the ExAC dataset as compared to CEN-LoFs (Supplementary Tables S6a and S6b).

Next, we evaluated AD-LoFs (n = 129) and CEN-LoFs (n = 125) in a whole exome sequencing (WES) dataset of 243 younger AJ subjects (age range 50–89, mean 71.4 ± 7.7) (43) that are ancestrally well matched to our AJ WGS sample (Supplementary Figure S1). We applied the same frequency filter as we used in the WGS dataset and identified 1646 rare LoFs and 11 834 rare synonymous sites. We observed no difference in the ratio of rare LoFs to synonymous variant sites between the AJ WGS and the AJ WES datasets (OR = 1.02, P = 0.61). However, when we evaluated the sum of rare LoF allele counts separately for AD and CEN normalized by synonymous allele counts, we observed that LoFs are enriched in AD compared to younger AJ controls (OR = 1.14, P = 0.0056) and that LoFs are depleted in CEN (OR = 0.88, P = 0.0052) compared to the controls (Table 3).

Finally, we analyzed WGS data of 24 centenarians with dementia (CEND) who did not meet the inclusion criterion of MMSE ≥ 24 for controls in our AD versus CEN analysis. These CEND subjects may have had non-specific dementia or other health problems impairing their ability to participate in the cognitive assessment. We performed enrichment test for LoFs again using synonymous variants for normalization. We found similar magnitude of enrichment of rare LoFs in AD compared to CEND and observed no differences between CEND and cognitively healthy CEN (Supplementary Table S7). This further supports the notion that CEN contribute to the observed differential LoF burden between AD and CEN. However, it also needs to be pointed out that the frequency of APOE4 allele among CEND (6.3%) was close the frequency among CEN (5.7%) as opposed to AD (38.9%), which highlights the difficulty to clearly separate the genetics of healthy aging from unspecified cognitive impairment in the very old population.

To evaluate whether cardiovascular risk genes contributed to the overall enrichment signals of LoFs in AD subjects, we retrieved the set of 111 genes mapped by significant GWAS SNPs (P ≤ 5E − 8) from the GWAS Catalog for cardiovascular disease, stroke or hypertension (37). Only four LoFs were observed in AD cases and five in CEN (Supplementary Table S8). This data size is too small for further interpretation beyond its anecdotal value. We also evaluated the possible role of haplosufficient (HS) genes. We have retrieved a list of haploinsufficiency scores across all human autosomal genes (44). We compare the 10% of genes delineated as most HS with the 10% delineated as the least HS, i.e. the most haploinsufficient, which does not show a notable difference (Supplementary Table S9).

Discussions

Relative to rare synonymous variants, we observed an overall higher burden of rare LoFs variants in AD subjects as compared with CEN. This burden was more pronounced for those rare variants that are predicted to cause NMD in >50% of known transcripts, variants affecting INITs and splice sites. Due to limited power, we cannot clearly pinpoint any specific rare mutations that are associated with increased risk of developing AD. However, when we narrowed down the analysis to the subset of innate immune genes, we found a stronger accumulation of LoFs among AD subjects than among healthy CEN. This observation that innate immune genes showed a stronger rare mutation burden is consistent with an important role of immunological dysfunction in AD and longevity (22,29,45).

Out of the 68 genes carrying NMD and INIT variants leading to the strong enrichment signals among AD subjects (Supplementary Table S3b), five genes encode ion channels (*SCN10A, KCNK10, TRPM2, KCNJ13* and *SHROOM3*) based on annotations by GO Consortium (46). In this context, it is noteworthy that transient receptor potential melastatin-2 (TRPM2) channels play a key role in amyloid-β-induced neurovascular dysfunction, and that TRPM2 channels are a potential therapeutic target to counteract cerebrovascular dysfunction in Alzheimer's dementia and related pathologies (47). A possible role of the *TRPM2* p.Gln953* mutation needs to be replicated in larger scale sequencing studies. Another interesting ion channel candidate gene is *KCNK10*, which encodes TREK-2, a member of two-pore domain potassium channels involved in the regulation of neuronal excitability in the entorhinal cortex (48). It is also interesting to find the *LPA* gene among the genes contributing to the enrichment signals, as variants in *LPA* have been shown to be both deleterious and protective in terms of cardiovascular risks (49–51). Cardiovascular diseases had been shown to increase the risks for AD (52). Consistently, the two previously reported risk alleles for coronary artery disease rs10455872 and rs3798220 (51) were observed exclusively in our AD subjects ($n = 5$ and 1, respectively) and not among any CEN. The other genes were linked to a wide range of disease conditions according to OMIM (53) such as Leber congenital amaurosis (*KCNJ13*) and iron-refractory iron deficiency anemia (*TMPRSS6*) as well as biological functions such as catalytic activities according to GO (46).

In line with an etiological role for genes with immune function and especially innate immunity in AD and longevity (22), we observed an enrichment of rare LoF variants among immune genes. Remarkably, we observed the strongest enrichment among AD subjects for NMD, INIT and splice site variants in innate immune genes. The innate immune genes harboring these potentially high-impact variants only in AD patients are *ABCA1, ADCY4, BPI, DHCR24, ITPR3, MMP9, NLRP8, POLR3F, POLR3G, RSAD2, SMAP2* and *TRPM2*. Among them, both *ABCA1* and *ITPR3* had been implicated as potential candidate genes for AD in an epigenomic analysis (29). *ABCA1* is a paralogue of the AD-associated *ABCA7* gene coding for a cholesterol efflux pump in the cellular lipid removal pathway and had been linked to AD in expression experiments and network analysis (56,57). A role of *DHCR24* and *MMP9* in AD has also been previously discussed (56,57). Literature search reveals limited prior evidence for an involvement of the other listed genes in AD. Given that the majority of AD-associated genes were identified through association studies of common variants, rare variants burden testing is expected to identify new previously unreported disease genes. Our analysis does not show the involvement of a specific gene in AD or longevity on its own, but nevertheless these genes have an increased probability to be risk genes.

Considering that EGs that are highly expressed in the human brain may have a higher prior probability to be involved in brain diseases, we observed an enrichment of LoF variants among AD subjects in this gene group as a positive control gene set. In contrast, no significant LoF burden was found in the group of genes with no prior link to phenotypic annotations, which provides a negative control gene set.

We did not find any overall enrichment for rare PAVs among AD subjects. Consistent with the expectation that NMDs and INIT mutations are particularly damaging (58), we observed a stronger enrichment for both types of mutations among AD patients compared to other classes of LoFs. We observed no significant enrichment for stop-gain variants in general, indicating that stop-gain variants that do not cause NMD may not be functionally relevant for AD.

A recent exome sequencing study comparing variant allele frequencies between 3000 Finnish individuals with 3000 non-Finnish Europeans showed a significant excess of rare deleterious variants in Finns, while also observing a deficit in variable sites in the population overall (50). An explanation was that population bottleneck had elevated the allele frequencies of deleterious alleles and selection has had insufficient time to remove them. Similar to the Finns, Ashkenazi Jews are a founder population with a recent bottleneck shaping the allele frequency spectrum (59,60). It is unclear whether our observation of rare LoFs enrichment in AD patients can be observed with a similar sample size in a more admixed population.

Another consideration is that protective variants promoting longevity may exist among CEN (61). We found a novel stop-gain variant Gln4247* (not predicted to cause NMD) in a known disease gene *APOB* in one CEN. Truncating mutations in *APOB* have been reported to cause autosomal recessive familial hypobetalipoproteinemia (62), but a non-synonymous variant E4154K has been associated with a reduction in risk of cerebrovascular disease and ischemic stroke (63).

Because our study applied an overall genome-wide burden test for enrichment of rare LoFs variants as a single hypothesis and subsequently evaluated the primary signal in subsets of genes and variants, we did not conduct multiple independent testing.

This study has several limitations. One major limitation is the small sample size that limits the power of our analysis. Future studies involving larger number of subjects are required to replicate the enrichment findings and narrow down the number of gene sets for burden testing. A second limitation is that we did not include a random AJ control group in the design of this WGS experiment to evaluate the origin of the enrichment signal. When matching our data with the ExAC data, LoFs ascertained exclusively among AD subjects are depleted relative to CEN LoFs. This shows that our AD LoFs are more specific to our dataset than CEN LoFs. It supports the notion that at least a part of the enrichment signal is driven by AD causative variants that are depleted both among CEN and a broader population sample. Furthermore, comparing to a younger AJ WES sample, AD subjects carried more and CEN fewer aggregated rare LoFs, which indicates that both AD and CEN contribute to the differential LoF burden in our WGS dataset. Our analysis of demented CEN also supports that a relevant fraction of the enrichment signals comes from CEN. One may speculate that AD and longevity traits share some risk genes. For instance, both AD and longevity are known to share *APOE* as a risk-modifying factor. The AD risk allele APOE4 is a known cardiovascular risk factor (64),

which is found to be less frequent among CEN compared to controls (34,35). Thus, it is possible that some rare LoFs increasing the risk for AD also work against longevity in other ways. It is also likely that those CEN with unspecific cognitive impairments investigated here did not have any AD, as the proportion of AD diagnosis seems to decline among subject over 90 years of age (64). A third limitation of our current analysis is its focus on coding variants for which functional annotations are more clearly defined comparing to non-coding variants. Future studies are necessary to explore rare variants burden in functional non-coding genomic regions.

In summary, our study of investigating rare variants in extreme phenotypes in a genetically homogenous founder population yielded findings that are consistent with results from larger sequencing studies of mixed population of other complex phenotypes. Our results indicate an increased burden for rare protein truncating mutations among AD patients as compared to CEN, and this enrichment is more pronounced for innate immune genes.

Materials and Methods

Study population

DNA samples from 80 CEN AJ subjects were collected as a part of a longevity study at the Albert Einstein College of Medicine according to the protocol described previously (66), of which 53 AJ CEN had an MMSE score of ≥ 24 (67) at the time of their enrollment and their mean age was 99.5 years old (range 95–106). This group was classified as cognitively healthy CEN. This MMSE score was the reported mean of people over the age of 85 years, and many CEN in this study suffer from vision and hearing impairments that are expected to affect the MMSE score (68). Twenty-four AJ CEN failed to meet the MMSE cutoff (range 0–23; mean = 6.5) were classified as having non-specific dementia (CEND). The enrollment of the CEN was approved by the institutional review board (IRB) of the Montefiore Medical Center and the Committee on Clinical Investigation at the Albert Einstein College of Medicine. Written informed consent was obtained from all subjects prior to participation.

DNA samples of 2046 subjects with AD were obtained from National Institute on Aging Genetics Initiative for Late-Onset Alzheimer Disease/National Cell Repository for Alzheimer Disease. Local IRB committees approved the study. These samples were genotyped for ancestry informative markers (AIMs) (69,70). Principal component analysis (PCA) was performed with all AD cases, and AJ CEN controls using the smartpca tool included in the EIGENSTRAT software (71). Euclidian distances were calculated for all pairs of samples over the top two principal components. For each CEN sample, the closest AD sample was then retrieved in a stepwise fashion, iteratively removing case–control pairs from the pool of all samples. PCA was then repeated with resulting set of case–control pairs to confirm proper matching. DNA of 45 AD subjects was matched to AJ population and had sufficient quantity for WGS. The mean age of onset for the 45 AJ AD subjects was 74.0 years (SD = 7.4 years).

WGS and quality control

WGS was performed on Illumina HiSeq 2500 platform (Illumina, San Diego, CA). When possible, we processed cases and controls in the same batch in order to limit batch effects. Paired-end 2×100 reads were aligned to the GRCh37 human reference using the Burrows-Wheeler Aligner (72), and processed using the best-practices pipeline that includes marking of duplicate reads by the use of Picard tools (http://picard.sourceforge.net), and realignment around DNA insertions and deletions (INDELs) and base recalibration via Genome Analysis Toolkit (GATK) ver. 3.2-2 (73). Variant discovery was performed as a three-step process. Single-sample genotypes were called using GATK HaplotypeCaller followed by joint genotyping of all subjects. The resulting multi-sample Variant Call Format file was used for variant quality score recalibration (74,75). A 99.8% Truth sensitivity tranche level was used for SNVs and 99.0% for INDEL variants. The concordance rate for SNVs between Illumina 2.5M chip and GATK variant calls was >99.7% for all subjects.

To further remove potentially false-positive genotype calls, we performed downstream genotype quality control (QC) filtering for call rates and removal of variants in regions that are reported to have poor mapping quality. Specifically, we removed variants with call rate < 0.98, number of alleles > 2, Hardy–Weinberg Equilibrium (HWE) $P < 0.001$ or Fisher's exact HWE $P < 0.001$. We then removed variants with GATK tags that are relevant to sequencing data quality: qual by depth (QD) < 3, Variant Quality Score (VQSLOD) < 0, or Mapping Quality Zero Read ≥ 1. In addition, we removed variants located in genomic regions annotated as segmental duplications according to UCSC Genome Browser (Genomic Super Dups 2011-10-25, UCSC) (76) and genomic regions that are likely to produce artifact signals (DAC Blacklisted Regions, UCSC) (77). After QC filtering, we obtained a total of 12 341 379 autosomal variants. PCA using AIMs on Immunochip (78) showed no obvious population substructures between AD subjects and AJ CEN subjects (Supplementary Figure S2).

To further exam the sequencing data for potential batch effects, we looked at genome-wide distribution doubleton pairs between AD, CEN and CEND. Among the 425 366 doubletons, we observed following pairs: 54 100 AD–AD, 80 591 CEN–CEN, 15 782 CEND–CEND, 136 330 AD–CEN, 62 064 AD–CEND and 76 499 CEN–CEND. We assume the number of pairs within a group is expected to be proportional to the square of the group size and the number of pairs between two groups is proportional to 2^*n^*m, where n and m are the two group sizes. When we plot the actual number of doubleton pairs versus 'expectation', it is roughly a straight line (Supplementary Figure S3), indicating no visible batch effects. We also obtained similar distributions of singletons between the three AJ sample groups (Supplementary Figure S4).

Annotation of variants

For the analysis of rare variants, we excluded markers found in dbSNPcommon141 (79) and those with MAF>0.025% in the 1000 Genomes Project (1KG) (i.e. seen more than once in 1KG Phase3) (80) and those variants with MAF >2% which is four copies of a variant allele in our total sample of 98 AJ subjects.

Variant classification was based on Reference Sequence GRCh37 and RefSeq Gene transcripts of NCBI Homo sapiens Annotation Release 105 that was implemented in the Golden Helix SNP & Variation Suite (SVS) (Golden Helix, Bozeman, MT 59718, USA). Details of the variants classification procedures can be found at the SVS manual (http://doc.goldenhelix.com). Because choice of software may affect variant annotation (39), we used VEP from Ensembl (81) as a second tool to verify our LoF variant prediction. We used the software snpEff to annotate for stop-gain variants that cause NMD in >50% of all known transcripts of a gene (82). Functional annotations of genes were

retrieved from the GO Consortium database (46). We used the software R to calculate for significance of rare variant burden between cases and controls [R Core Team (2013). R: A language and environment for statistical computing. Vienna, Austria. URL http://www.R-project.org].

Power analysis

With infrequent or rare variants, the statistical power is low for variant-based association testing, unless the sample size or the effect size is very large (18). We calculated the statistical power for the sample size of 45 cases and 53 controls using the prevalence for AD of 32% for people aged > 85 years (16) and a relative risk of 3 for heterozygous carriers. If single variant association is used, a risk variant with a risk allele frequency of 5% would only have 44% power to be detected at $\alpha = 0.001$ and no power at $\alpha = 5E - 8$, according to estimates of genetic power calculator (83). However, if we aggregate multiple rare variants in gene sets or across the whole genome to reach a cumulative allele frequency of 10%, the power for association detection would increase to 64% ($\alpha = 0.001$) (Supplementary Table S10). Furthermore, we performed a separate power analysis that used the two-sample proportion test for estimating the minimum required variant number instead of the required subject sample number to reach 80% power (Supplementary Figure S5).

Additional sequencing data of broader population sample

We downloaded the non-psychiatric non-Finnish European variants sites dataset from the Exome Aggregation Consortium (ExAC) server (ftp://ftp.broadinstitute.org/pub/ExAC_release/, release 0.3.1). We also obtained genotype data of 243 younger Ashkenazi Jewish controls from the T2D-GENES Consortium (age 50–89 years) sequenced on an Illumina HiSeq 2000 (43). Additional QC filtering included removing variants with call rate < 0.98, number of alleles > 2, Fisher's exact HWE $P < 0.001$, VQSLOD < 0, QD < 1 as well as those in genomic regions annotated as segmental duplications and DAC Blacklisted Regions described above. Ancestry was checked by PCA together with our AJ WGS sample based on 1KG European variants with MAF $> 5\%$ (Supplementary Figure S1).

Supplementary Material

Supplementary Material is available at *HMG* online.

Acknowledgements

Authors thank the centenarians and their families as well as the members of the Longevity Genes Project at the Albert Einstein College of Medicine for their contribution to this study. We also thank the participants and staff members of the National Institute on Aging Genetics Initiative for Late-Onset Alzheimer Disease/National Cell Repository for Alzheimer Disease (NIA-LOAD/NCRAD). The T2D-Genes Consortium is funded by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK).

Conflict of Interest statement. J.F. is now full time employee of Regeneron Pharmaceuticals, Inc. V.V. is now full time employee of 23andMe Inc.

References

1. Querfurth, H.W. and LaFerla, F.M. (2010) Alzheimer's disease. *N. Engl. J. Med.*, **362**, 329–344.
2. Ridge, P.G., Mukherjee, S., Crane, P.K. and Kauwe, J.S. (2013) Alzheimer's disease: analyzing the missing heritability. *PloS One*, **8**, e79771.
3. Lee, S.H., Harold, D., Nyholt, D.R., Goddard, M.E., Zondervan, K.T., Williams, J., Montgomery, G.W., Wray, N.R. and Visscher, P.M. (2013) Estimation and partitioning of polygenic variation captured by common SNPs for Alzheimer's disease, multiple sclerosis and endometriosis. *Hum. Mol. Genet.*, **22**, 832–841.
4. Cirulli, E.T. and Goldstein, D.B. (2010) Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat. Rev.*, **11**, 415–425.
5. Guerreiro, R., Wojtas, A., Bras, J., Carrasquillo, M., Rogaeva, E., Majounie, E., Cruchaga, C., Sassi, C., Kauwe, J.S., Younkin, S., *et al.* (2012) TREM2 variants in Alzheimer's disease. *N. Engl. J. Med.*, **368**, 117–127.
6. Jonsson, T., Stefansson, H., Steinberg, S., Jonsdottir, I., Jonsson, P.V., Snaedal, J., Bjornsson, S., Huttenlocher, J., Levey, A.I., Lah, J.J., *et al.* (2012) Variant of TREM2 associated with the risk of Alzheimer's disease. *N. Engl. J. Med.*, **368**, 107–116.
7. Jonsson, T., Atwal, J.K., Steinberg, S., Snaedal, J., Jonsson, P.V., Bjornsson, S., Stefansson, H., Sulem, P., Gudbjartsson, D., Maloney, J., *et al.* (2012) A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature*, **488**, 96–99.
8. Cruchaga, C., Karch, C.M., Jin, S.C., Benitez, B.A., Cai, Y., Guerreiro, R., Harari, O., Norton, J., Budde, J., Bertelsen, S., *et al.* (2014) Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. *Nature*, **505**, 550–554.
9. Wetzel-Smith, M.K., Hunkapiller, J., Bhangale, T.R., Srinivasan, K., Maloney, J.A., Atwal, J.K., Sa, S.M., Yaylaoglu, M.B., Foreman, O., Ortmann, W., *et al.* (2014) A rare mutation in UNC5C predisposes to late-onset Alzheimer's disease and increases neuronal cell death. *Nat. Med.*, **20**, 1452–1457.
10. Cirulli, E.T., Lasseigne, B.N., Petrovski, S., Sapp, P.C., Dion, P.A., Leblond, C.S., Couthouis, J., Lu, Y.F., Wang, Q., Krueger, B.J., *et al.* (2015) Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. *Science (New York, N.Y.)*
11. Schulte, E.C., Fukumori, A., Mollenhauer, B., Hor, H., Arzberger, T., Perneczky, R., Kurz, A., Diehl-Schmid, J., Hull, M., Lichtner, P., *et al.* (2015) Rare variants in beta-Amyloid precursor protein (APP) and Parkinson's disease. *Eur. J. Hum. Genet.*, **23**, 1328–1333.
12. Do, R., Stitziel, N.O., Won, H.H., Jorgensen, A.B., Duga, S., Angelica Merlini, P., Kiezun, A., Farrall, M., Goel, A., Zuk, O.,

et al. (2015) Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. *Nature*, **518**, 102–106.

13. Price, R.A., Li, W.D. and Zhao, H. (2008) FTO gene SNPs associated with extreme obesity in cases, controls and extremely discordant sister pairs. *BMC Med. Genet.*, **9**, 4.

14. Barnett, I.J., Lee, S. and Lin, X. (2013) Detecting rare variant effects using extreme phenotype sampling in sequencing association studies. *Genet. Epidemiol.*, **37**, 142–151.

15. Garagnani, P., Giuliani, C., Pirazzini, C., Olivieri, F., Bacalini, M.G., Ostan, R., Mari, D., Passarino, G., Monti, D., Bonfigli, A.R., *et al.* (2013) Centenarians as super-controls to assess the biological relevance of genetic risk factors for common age-related diseases: a proof of principle on type 2 diabetes. *Aging (Albany NY)*, **5**, 373–385.

16. Hebert, L.E., Weuve, J., Scherr, P.A. and Evans, D.A. (2013) Alzheimer disease in the United States (2010-2050) estimated using the 2010 census. *Neurology*, **80**, 1778–1783.

17. Guha, S., Rosenfeld, J.A., Malhotra, A.K., Lee, A.T., Gregersen, P.K., Kane, J.M., Pe'er, I., Darvasi, A. and Lencz, T. (2012) Implications for health and disease in the genetic signature of the Ashkenazi Jewish population. *Genome Biol.*, **13**, R2.

18. Lee, S., Abecasis, G.R., Boehnke, M. and Lin, X. (2014) Rare-variant association analysis: study designs and statistical tests. *Am. J. Hum. Genet.*, **95**, 5–23.

19. Aisen, P.S. and Davis, K.L. (1994) Inflammatory mechanisms in Alzheimer's disease: implications for therapy. *Am. J. Psychiatr.*, **151**, 1105–1113.

20. Cagnin, A., Brooks, D.J., Kennedy, A.M., Gunn, R.N., Myers, R., Turkheimer, F.E., Jones, T. and Banati, R.B. (2001) In-vivo measurement of activated microglia in dementia. *Lancet*, **358**, 461–467.

21. in 't Veld, B.A., Ruitenberg, A., Hofman, A., Launer, L.J., van Duijn, C.M., Stijnen, T., Breteler, M.M. and Stricker, B.H. (2001) Nonsteroidal antiinflammatory drugs and the risk of Alzheimer's disease. *N. Engl. J. Med.*, **345**, 1515–1521.

22. Heneka, M.T., Golenbock, D.T. and Latz, E. (2015) Innate immunity in Alzheimer's disease. *Nat. Immunol.*, **16**, 229–236.

23. Lambert, J.C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., DeStafano, A.L., Bis, J.C., Beecham, G.W., Grenier-Boley, B., *et al.* (2013) Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.*, **45**, 1452–1458.

24. Guerreiro, R., Bras, J. and Hardy, J. (2013) SnapShot: genetics of Alzheimer's disease. *Cell*, **155**, 968–968. e1.

25. Rosenthal, S.L. and Kamboh, M.I. (2014) Late-Onset Alzheimer's Disease Genes and the Potentially Implicated Pathways. *Curr. Genet. Med. Rep.*, **2**, 85–101.

26. Cantoni, C., Bollman, B., Licastro, D., Xie, M., Mikesell, R., Schmidt, R., Yuede, C.M., Galimberti, D., Olivecrona, G., Klein, R.S., *et al.* (2015) TREM2 regulates microglial cell activation in response to demyelination in vivo. *Acta Neuropathol.*, **129**, 429–447.

27. International Genomics of Alzheimer's Disease Consortium (IGAP) (2014) Convergent genetic and expression data implicate immunity in Alzheimer's disease. *Alzheimers Dement.*, **11**, 658–671.

28. Zhang, B., Gaiteri, C., Bodea, L.G., Wang, Z., McElwee, J., Podtelezhnikov, A.A., Zhang, C., Xie, T., Tran, L., Dobrin, R., *et al.* (2013) Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell*, **153**, 707–720.

29. Gjoneska, E., Pfenning, A.R., Mathys, H., Quon, G., Kundaje, A., Tsai, L.H. and Kellis, M. (2015) Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease. *Nature*, **518**, 365–369.

30. Seshadri, S., Fitzpatrick, A.L., Ikram, M.A., DeStefano, A.L., Gudnason, V., Boada, M., Bis, J.C., Smith, A.V., Carassquillo, M.M., Lambert, J.C., *et al.* (2010) Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA*, **303**, 1832–1840.

31. Harold, D., Abraham, R., Hollingworth, P., Sims, R., Gerrish, A., Hamshere, M.L., Pahwa, J.S., Moskvina, V., Dowzell, K., Williams, A., *et al.* (2009) Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat. Genet.*, **41**, 1088–1093.

32. Jun, G., Naj, A.C., Beecham, G.W., Wang, L.S., Buros, J., Gallins, P.J., Buxbaum, J.D., Ertekin-Taner, N., Fallin, M.D., Friedland, R., *et al.* (2010) Meta-analysis confirms CR1, CLU, and PICALM as alzheimer disease risk loci and reveals interactions with APOE genotypes. *Arch. Neurol.*, **67**, 1473–1484.

33. Schachter, F., Faure-Delanef, L., Guenot, F., Rouger, H., Froguel, P., Lesueur-Ginot, L. and Cohen, D. (1994) Genetic associations with human longevity at the APOE and ACE loci. *Nat. Genet.*, **6**, 29–32.

34. Freudenberg-Hua, Y., Freudenberg, J., Vacic, V., Abhyankar, A., Emde, A.K., Ben-Avraham, D., Barzilai, N., Oschwald, D., Christen, E., Koppel, J., *et al.* (2014) Disease variants in genomes of 44 centenarians. *Mol. Genet. Genomic Med.*, **2**, 438–450.

35. Ryu, S., Atzmon, G., Barzilai, N., Raghavachari, N. and Suh, Y. (2016) Genetic landscape of APOE in human longevity revealed by high-throughput sequencing. *Mech. Ageing Dev.*, **155**, 7–9.

36. Beekman, M., Blanche, H., Perola, M., Hervonen, A., Bezrukov, V., Sikora, E., Flachsbart, F., Christiansen, L., De Craen, A.J., Kirkwood, T.B., *et al.* (2013) Genome-wide linkage analysis for human longevity: Genetics of Healthy Aging Study. *Aging Cell*, **12**, 184–193.

37. Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., *et al.* (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.*, **42**, D1001–D1006.

38. Liu, X., Jian, X. and Boerwinkle, E. (2013) dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. *Hum. Mut.*, **34**, E2393–E2402.

39. McCarthy, D.J., Humburg, P., Kanapin, A., Rivas, M.A., Gaulton, K., Cazier, J.B. and Donnelly, P. (2014) Choice of transcripts and software has a large effect on variant annotation. *Genome Med.*, **6**, 26.

40. Breuer, K., Foroushani, A.K., Laird, M.R., Chen, C., Sribnaia, A., Lo, R., Winsor, G.L., Hancock, R.E., Brinkman, F.S. and Lynn, D.J. (2013) InnateDB: systems biology of innate immunity and beyond–recent updates and continuing curation. *Nucleic Acids Res.*, **41**, D1228–D1233.

41. Georgi, B., Voight, B.F. and Bucan, M. (2013) From mouse to human: evolutionary genomics analysis of human orthologs of essential genes. *PLoS Genet.*, **9**, e1003484.

42. Hamosh, A., Scott, A.F., Amberger, J.S., Bocchini, C.A. and McKusick, V.A. (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.*, **33**, D514–D517.

43. Flannick, J., Thorleifsson, G., Beer, N.L., Jacobs, S.B., Grarup, N., Burtt, N.P., Mahajan, A., Fuchsberger, C., Atzmon, G., Benediktsson, R., *et al.* (2014) Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. *Nat. Genet.*, **46**, 357–363.

44. Huang, N., Lee, I., Marcotte, E.M. and Hurles, M.E. (2010) Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet.*, **6**, e1001154.

45. Gandy, S. and Heppner, F.L. (2013) Microglia as dynamic and essential components of the amyloid hypothesis. *Neuron*, **78**, 575–577.

46. Gene_Ontology_Consortium. (2015) Gene Ontology Consortium: going forward. *Nucleic Acids Res.*, **43**, D1049–D1056.

47. Park, L., Wang, G., Moore, J., Girouard, H., Zhou, P., Anrather, J. and Iadecola, C. (2014) The key role of transient receptor potential melastatin-2 channels in amyloid-beta-induced neurovascular dysfunction. *Nat. Commun.*, **5**, 5318.

48. Xiao, Z., Deng, P.Y., Rojanathammanee, L., Yang, C., Grisanti, L., Permpoonputtana, K., Weinshenker, D., Doze, V.A., Porter, J.E. and Lei, S. (2009) Noradrenergic depression of neuronal excitability in the entorhinal cortex via activation of TREK-2 K+ channels. *J. Biol. Chem.*, **284**, 10980–10991.

49. Helgadottir, A., Gretarsdottir, S., Thorleifsson, G., Holm, H., Patel, R.S., Gudnason, T., Jones, G.T., van Rij, A.M., Eapen, D.J., Baas, A.F., *et al.* (2012) Apolipoprotein(a) genetic sequence variants associated with systemic atherosclerosis and coronary atherosclerotic burden but not with venous thromboembolism. *J. Am. Coll. Cardiol.*, **60**, 722–729.

50. Lim, E.T., Wurtz, P., Havulinna, A.S., Palta, P., Tukiainen, T., Rehnstrom, K., Esko, T., Magi, R., Inouye, M., Lappalainen, T., *et al.* (2014) Distribution and medical impact of loss-of-function variants in the Finnish founder population. *PLoS Genet.*, **10**, e1004494.

51. Clarke, R., Peden, J.F., Hopewell, J.C., Kyriakou, T., Goel, A., Heath, S.C., Parish, S., Barlera, S., Franzosi, M.G., Rust, S., *et al.* (2009) Genetic variants associated with Lp(a) lipoprotein level and coronary disease. *N. Engl. J. Med.*, **361**, 2518–2528.

52. de Bruijn, R.F. and Ikram, M.A. (2014) Cardiovascular risk factors and future risk of Alzheimer's disease. *BMC Med.*, **12**, 130.

53. McKusick, V.A. (2007) Mendelian Inheritance in Man and its online version, OMIM. *Am. J. Hum. Genet.*, **80**, 588–604.

54. Kim, W.S., Bhatia, S., Elliott, D.A., Agholme, L., Kagedal, K., McCann, H., Halliday, G.M., Barnham, K.J. and Garner, B. (2010) Increased ATP-binding cassette transporter A1 expression in Alzheimer's disease hippocampal neurons. *J. Alzheimers Dis.*, **21**, 193–205.

55. Talwar, P., Silla, Y., Grover, S., Gupta, M., Agarwal, R., Kushwaha, S. and Kukreti, R. (2014) Genomic convergence and network analysis approach to identify candidate genes in Alzheimer's disease. *BMC Genomics*, **15**, 199.

56. Crameri, A., Biondi, E., Kuehnle, K., Lutjohann, D., Thelen, K.M., Perga, S., Dotti, C.G., Nitsch, R.M., Ledesma, M.D. and Mohajeri, M.H. (2006) The role of seladin-1/DHCR24 in cholesterol biosynthesis, APP processing and Abeta generation in vivo. *Embo J.*, **25**, 432–443.

57. Flex, A., Giovannini, S., Biscetti, F., Liperoti, R., Spalletta, G., Straface, G., Landi, F., Angelini, F., Caltagirone, C., Ghirlanda, G., *et al.* (2014) Effect of proinflammatory gene polymorphisms on the risk of Alzheimer's disease. *Neurodegener. Dis.*, **13**, 230–236.

58. MacArthur, D.G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., Jostins, L., Habegger, L., Pickrell, J.K., Montgomery, S.B., *et al.* (2012) A systematic survey of loss-of-function variants in human protein-coding genes. *Science (New York, N.Y.)*, **335**, 823–828.

59. Atzmon, G., Hao, L., Pe'er, I., Velez, C., Pearlman, A., Palamara, P.F., Morrow, B., Friedman, E., Oddoux, C., Burns, E., *et al.* (2010) Abraham's children in the genome era: major Jewish diaspora populations comprise distinct genetic clusters with shared Middle Eastern Ancestry. *Am. J. Hum. Genet.*, **86**, 850–859.

60. Carmi, S., Hui, K.Y., Kochav, E., Liu, X., Xue, J., Grady, F., Guha, S., Upadhyay, K., Ben-Avraham, D., Mukherjee, S., *et al.* (2014) Sequencing an Ashkenazi reference panel supports population-targeted personal genomics and illuminates Jewish and European origins. *Nat. Commun.*, **5**, 4835.

61. Bergman, A., Atzmon, G., Ye, K., MacCarthy, T. and Barzilai, N. (2007) Buffering mechanisms in aging: a systems approach toward uncovering the genetic component of aging. *PLoS Comput. Biol.*, **3**, e170.

62. Tybjaerg-Hansen, A., Steffensen, R., Meinertz, H., Schnohr, P. and Nordestgaard, B.G. (1998) Association of mutations in the apolipoprotein B gene with hypercholesterolemia and the risk of ischemic heart disease. *N. Engl. J. Med.*, **338**, 1577–1584.

63. Benn, M., Nordestgaard, B.G., Jensen, J.S. and Tybjaerg-Hansen, A. (2007) Polymorphisms in apolipoprotein B and risk of ischemic stroke. *J. Clin. Endocrinol. Metab.*, **92**, 3611–3617.

64. Bennet, A.M., Di Angelantonio, E., Ye, Z., Wensley, F., Dahlin, A., Ahlbom, A., Keavney, B., Collins, R., Wiman, B., de Faire, U., *et al.* (2007) Association of apolipoprotein E genotypes with lipid levels and coronary risk. *JAMA*, **298**, 1300–1311.

65. Crystal, H.A., Dickson, D., Davies, P., Masur, D., Grober, E. and Lipton, R.B. (2000) The relative frequency of "dementia of unknown etiology" increases with age and is nearly 50% in nonagenarians. *Arch. Neurol.*, **57**, 713–719.

66. Barzilai, N., Atzmon, G., Schechter, C., Schaefer, E.J., Cupples, A.L., Lipton, R., Cheng, S. and Shuldiner, A.R. (2003) Unique lipoprotein phenotype and genotype associated with exceptional longevity. *JAMA*, **290**, 2030–2040.

67. Folstein, M.F., Folstein, S.E. and McHugh, P.R. (1975) "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.*, **12**, 189–198.

68. Crum, R.M., Anthony, J.C., Bassett, S.S. and Folstein, M.F. (1993) Population-based norms for the Mini-Mental State Examination by age and educational level. *JAMA*, **269**, 2386–2391.

69. Tian, C., Plenge, R.M., Ransom, M., Lee, A., Villoslada, P., Selmi, C., Klareskog, L., Pulver, A.E., Qi, L., Gregersen, P.K., *et al.* (2008) Analysis and application of European genetic substructure using 300 K SNP information. *PLoS Genet.*, **4**, e4.

70. Shriver, M.D. and Kittles, R.A. (2004) Genetic ancestry and the search for personalized genetic histories. *Nat. Rev.*, **5**, 611–618.

71. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.*, **38**, 904–909.

72. Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, **25**, 1754–1760.

73. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.*, **20**, 1297–1303.

74. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.*, **43**, 491–498.

75. Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., *et al.* (2013) From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics*, **11**, 11 10 1-11 10 33.

76. Bailey, J.A., Gu, Z., Clark, R.A., Reinert, K., Samonte, R.V., Schwartz, S., Adams, M.D., Myers, E.W., Li, P.W. and Eichler, E.E. (2002) Recent segmental duplications in the human genome. *Science (New York, N.Y*, **297**, 1003–1007.

77. ENCODE(2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.

78. Eyre, S., Bowes, J., Diogo, D., Lee, A., Barton, A., Martin, P., Zhernakova, A., Stahl, E., Viatte, S., McAllister, K., *et al.* (2012) High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nat. Genet.*, **44**, 1336–1340.

79. Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M. and Sirotkin, K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29**, 308–311.

80. Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T. and McVean, G.A. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, **491**, 56–65.

81. McLaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P. and Cunningham, F. (2010) Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics (Oxford, England)*, **26**, 2069–2070.

82. Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X. and Ruden, D.M. (2012) 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. *Fly (Austin)*, **6**, 80–92.

83. Purcell, S., Cherny, S.S. and Sham, P.C. (2003) Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics (Oxford, England)*, **19**, 149–150.