Title: Gene-based association studies report four novel genes in the etiology of clinical subtypes of frontotemporal dementia

Authors: Aniket Mishra[1], Raffaele Ferrari[2], IFGC°, Peter Heutink[3], John Hardy[2], Yolande Pijnenburg[4], Danielle Posthuma[1,5]

Affiliations: [1]Department of Complex Trait Genetics, VU University, Center for Neurogenomics and Cognitive Research, Amsterdam, 1081 HV, the Netherlands, [2]Department of Molecular Neuroscience, UCL, Russell Square House, 9-12 Russell Square House London, WC1B 5EH, UK

[3]Department of Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen, 72076, Tübingen, Germany

[4]Alzheimer Center and Department of Neurology, VU University Medical Center (VUMC), Neursocience Campus Amsterdam, Amsterdam, 1081 HV, the Netherlands

[5] Department of Clinical Genetics, VU University Medical Center (VUMC), Neursocience Campus Amsterdam, Amsterdam, 1081 HV, the Netherlands

Corresponding Author: Dr. Aniket Mishra a.mishra@vu.nl and/or Prof. Danille Posthuma d.posthuma@vu.nl

°International FTD-Genomics Consortium (IFGC)

Abstract

INTRODUCTION: Genome-wide association studies (GWASs) in frontotemporal dementia (FTD) showed limited success in identifying associated loci. This is possibly due to small sample size, allelic heterogeneity, small effect sizes of single genetic variants, and the necessity to statistically correct for testing millions of genetic variants.

METHODS: To overcome these issues, we performed gene-based association studies on 3348 clinically identified FTD cases and 9390 controls (discovery, replication and joint-cohort analyses).

RESULTS: We report association of *APOE* and *TOMM40* with behavioural FTD (bvFTD), and *ARHGAP35* and *SERPINA1* with progressive non-fluent aphasia (PNFA). Further, we found the ε2 and ε4 alleles of *APOE* harbouring protective and risk increasing effects, respectively, in clinical subtypes of FTD.

DISCUSSION: The *APOE*-locus association with bvFTD indicates its potential risk-increasing role across different neurodegenerative diseases, whilst the novel genetic association of *ARHGAP35* and *SERPINA1* with PNFA points towards a potential role of the stress-signalling pathway in its pathophysiology.

Keywords:

Gene-based association study, GWAS, FTD, MAGMA, bvFTD, APOE, TOMM40, epsilon alleles, PNFA, stress-signalling pathway, ARHGAP35, GRLF1, SERPINA1, FTD-MND, C9orf72

Research in context (< 150 words)

Frontotemporal dementia (FTD) is a clinically heterogeneous disorder, presenting mainly with behavioural or language symptoms and co-occurring with motor neuron disease in a minority of cases. Recent single marker genome-wide association studies (GWAS) on FTD showed limited success in identifying associated loci possibly due to small sample size, allelic heterogeneity and the stringency of the Bonferroni correction. In this study, we performed an alternative joint-SNP gene-based analyses using the GWAS data on 3348 clinically identified FTD cases and 9390 controls. We identified association of *APOE* and *TOMM40* genes with behavioural FTD, and *ARHGAP35* and *SERPINA1* genes with progressive non-fluent aphasia. The *APOE* gene association with behavioural FTD points towards its potential role across different neurodegenerative diseases. This is the first work reporting a significant genetic association with progressive non-fluent aphasia: *SERPINA1* and *ARHGAP35* point to a potential role of a stress-pathway in the pathogenesis of progressive non-fluent aphasia.

Highlights

- Gene-based association study reports association of *APOE* and *TOMM40* genes with bvFTD, and *ARHGAP35* and *SERPINA1* with PNFA.
- ε2 and ε4 alleles of *APOE*, respectively, show protective and disease increasing effects for clinical subtypes of FTD.
- The *TOMM40* gene harbours genetic variations associated with bvFTD that is independent of the epsilon alleles.
- The *ARHGAP35* and *SERPINA1* gene associations with PNFA guides towards the role of stress-signalling pathway in its pathophysiology.

Abbreveations:

bvFTD = Behavioural variant of frontotemporal dementia

CBG = Corticosteroid binding globulin

GWAS = Genome-wide association study

IFGC = International FTD-GWAS consortium

FTD = Frontotemporal dementia

FTD-MND = Frontotemporal dementia with motor neuron disease

FTLD = Frontotemporal lobar degeneration

hGR = Human glucocorticoid receptor

LD = Linkage disequilibrium

MAGMA = Multi-marker Analysis of GenoMic Annotation

PNFA = Progressive non fluent aphasia

ORF = Open reading frame

SD = Semantic Dementia

TDP-43 pathology = TAR-DNA binding protein 43 pathology

UTR = Untranslated region

Introduction

Frontotemporal dementia (FTD) is one of the leading causes of dementia in patients younger than 65 years of age[1, 2]. It is characterised by degeneration of the frontal and anterior temporal lobes leading to a decline in behaviour and language. FTD is a heterogeneous condition clinically, pathologically and genetically[2-4]. Clinically, it is broadly categorised into the behavioural variant (bvFTD) and the language variant or primary progressive aphasia (PPA), which is further categorised into semantic dementia (SD) and progressive non-fluent aphasia (PNFA). There is a frequent overlap between FTD and a number of motor diseases such as parkinsonian disorders, corticobasal syndrome, progressive supranuclear palsy and motor neuron disease (FTD-MND)[5]. The underlying pathological spectrum of FTD, termed frontal temporal lobar degeneration (FTLD), is based on neuronal lesions and protein inclusions such as with tau or TAR-DNA binding protein (TDP)-43 pathology. Beside the Mendelian genes *MAPT*, *GRN* and *C9orf72* that are causal in up to ~30-50% of familial FTLD cases, rare variability in few other genes associates with less than 5% of cases[5-7]. To date only a couple of large genome wide association studies (GWAS) have been performed for FTD[8-10] reporting an association with *TMEM106B* for FTLD with TDP-43 pathology[8], and with the locus comprising *RAB38* and *CTSC* as well as the *HLA-DRA/HLA-DRB5* locus for bvFTD and FTD, respectively[9].

In a typical GWAS, an association test on a single variant (SNPs or Indels) is performed to map genes associated with a phenotype; however, many independent risk alleles for a given phenotype can be localised within a gene[11-13]. Hence a classical GWAS approach will be less powered to detect genes containing many independent risk alleles[14]. A joint-variant gene based test that combines independent association

signals within a gene while accounting for the linkage disequilibrium (LD) between variants can overcome this limitation. A number of approaches have been reported to perform joint-SNP gene-based analysis: the permutation test – where empirical evidence of association of the combined test statistics is calculated by shuffling the samples while keeping markers intact – is currently considered the golden standard[15]. However, the requirement of genotype data and computational burden limits its use. Recently, our group developed a new approach called Multi-marker Analysis of GenoMic Annotation (MAGMA) that uses a multiple regression model to perform joint-SNP gene-based analysis using GWAS summary data[16].

In this study, we performed a hypothesis free gene-wide association study on FTD subtypes (bvFTD, SD, PNFA and FTD-MND) using GWAS summary files obtained from the International FTD-Genomics consortium (IFGC)[9]. We used the MAGMA software to perform the gene-based analysis. We report results of discovery, replication and combined cohort analyses for each FTD subtype; we also assessed individual risk variants for associated genes, which can be used for replication in the individual variant genotype setting.

Methods

Samples

The dataset used in the FTD-GWAS was described previously[9]. Briefly, 44 international groups contributed clinical FTD samples. Patients were diagnosed according to the Neary criteria or the revised criteria for bvFTD and language variants of FTD[17, 18]. Approximately 3% of cases were pathologically confirmed. For the current study we used the GWAS summary datasets of the discovery and replication

cohorts of each FTD subtypes, bvFTD (discovery: 1377 cases and 2754 controls; replication: 690 cases and 5092 controls), PNFA (discovery: 269 cases and 538 controls; replication: 221 cases and 5092 controls), SD (discovery: 308 cases and 616 controls; replication: 189 cases and 5092 controls) and FTD-MND (discovery: 200 cases and 400 controls; replication: 94 cases and 5092 controls).

Statistical analysis

We performed the joint-SNP gene-based analysis using MAGMA[16]. The MAGMA approach is based on a multiple linear principal components regression model. By projecting the multivariate LD matrix of SNPs in a gene it first extracts principal components that explain genetic variation. These principal components are further used as predictors of a phenotype under a linear regression framework. MAGMA then uses Fisher's test to compute p-values to test association between a gene and the phenotype. We used 19418 hg19 annotated protein-coding genes to perform the analysis. Since all the samples involved were of European descent, we considered the 1000 genomes phase 1 European reference population to estimate LD between variants[19]. We only considered SNPs in the 5'- and 3'-untranslated region (UTR) and the open reading frame (ORF) for the joint-SNP gene-based tests. This strategy resulted in loss of *cis*-regulatory variants, but was more stringent and ORF-specific.

The schematic representation of the strategy for multi-stage gene-wide association analysis for FTD and subtypes is described in supplementary figure 1. We performed separate gene-based tests using discovery and replication datasets for each FTD subtype reported previously by the IFGC[9]. We performed gene-based tests using only those variants that were either genotyped or imputed with imputation score more than 0.50. Moreover we only considered common variants with minor allele frequency more than

0.01. For individual FTD subtypes, 4303460 and 55375 variants were available for the gene-based analysis in the discovery and replication cohorts respectively. Further, 16313 and 10349 genes that contained at least one variant within the 5'-, 3'-UTR and ORF, were tested for association with a given FTD subtype in the discovery and replication cohorts, respectively. To identify additional genes associated with individual FTD subtypes, we meta-analysed the gene-based p-values obtained in the discovery and replication cohorts using the Stouffer's combination approach for the sample size weighted combination of p-values. For each FTD subtype we tested association of total 16920 genes either in the discovery or replication cohorts. To correct for multiple association tests performed for 16920 genes with one of the four subtypes of FTD, we applied the conservative Bonferroni correction method establishing a gene-wide significance threshold at 7.388×10^{-07} (= 0.05 / (4 × 16920)). To identify genes associated with any FTD subtype, we combined the gene-based test statistics for either subtype (bvFTD, SD, PNFA and FTD-MND) using the sample size weighted Stouffer's combination method.

Functional characterisation of associated genes

We downloaded the gene expression profiles of the associated genes across 13 human brain tissues (in alphabetically order: amygdala, anterior cingulate cortex, caudate, cerebral hemisphere, cerebellum, cortex, frontal cortex, hippocampus, nucleus accumbens, putamen, spinal cord, substantia nigra) using the GTEx portal[20] (GTEx Analysis V6 dbGaP Accession phs000424.v6.p1). We also investigated the functional annotations of variants that are in LD ($r^2 > 0.8$ in the 1000 genomes phase 1 European panel) with SNPs used in deriving gene-based p-values of the associated genes using software HaploReg[21] (version 4.1) and RegulomeDB[22] (version 1.1).

Results

Associations with FTD and its subtypes

bvFTD

In the discovery cohort, two genes passed the gene-wide significant p-value threshold 7.388×10^{-07}: *TOMM40* ($p = 5.786 \times 10^{-8}$) and *APOE* ($p = 1.367 \times 10^{-7}$). In the replication cohort the p-values were 6.40×10^{-5} for *TOMM40* and 1.688×10^{-3} for *APOE* suggesting consistency of associations across independent bvFTD samples. No other genes passed the significance threshold for the bvFTD subtype (supplementary figure 2a).

Interestingly, in the discovery cohort the SNPs rs7412 and rs429358, which determine three epsilon (ε) alleles ε2, ε3 and ε4 of the *APOE* gene, were among the SNPs driving its association with bvFTD, with p-values 0.023 (rs7412) and 5.04×10^{-6} (rs429358). In the replication cohort rs429358 was not genotyped, whereas information on rs769449, an intronic variant in high LD ($r^2 = 0.82$, 1000 Genomes phase 1 European population) with rs429358, was available; here the p-values were 0.222 for rs7412 and 1.945×10^{-4} for rs769449.

To check whether the association of *TOMM40* with bvFTD was independent of the epsilon variants, we re-performed the gene-based test on *TOMM40* gene using only those variants in negligible LD ($r^2 < 0.2$) with rs7412 and rs429358 in 1000 genomes phase 1 European panel. This analysis showed moderate association of *TOMM40* with bvFTD ($p = 7.513 \times 10^{-6}$; table 1) suggesting that the *TOMM40* gene harbours signals for the risk of bvFTD that are independent of the epsilon alleles of *APOE* gene.

The summary statistics of variants used for deriving gene-based p-values for *TOMM40* and *APOE* are given in supplementary tables 2a and 2b respectively, and the regional plots are shown in figure 1a and 1b. The regional plots show many variants in *TOMM40* with p-values less than 0.05 that are in negligible LD ($r^2 < 0.2$) with rs769449, a proxy of epsilon variant rs429358.

FIGURE 1A and 1B Regional plots at the *TOMM40*/*APOE* locus in bvFTD cohorts

PNFA

The joint-cohort (discovery and replication) analysis revealed association for *ARHGAP35* ($p = 2.950 \times 10^{-7}$) and *SERPINA1* ($p = 3.024 \times 10^{-7}$) with PNFA (table 1 and supplementary table 3). The regional plots for *ARHGAP35* and *SERPINA1* (figure 2a and 2b respectively) show a robust LD block only for *ARHGAP35* for which all variants show association p-values less than 0.05 with PNFA (also refer to supplementary table 3a). In *SERPINA1* many LD independent variants with PNFA association p-values less than 0.05 can be observed.

FIGURE 2 Regional plots at A) *ARHGAP35* and B) *SERPINA1* loci in PNFA cohorts

SD

No gene exceeded the gene-wide significance threshold 7.388×10^{-07}, possibly because of smaller sample size, thus reduced power. The top gene identified in the combined analysis was *WDR66* ($p = 9.50 \times 10^{-6}$; supplementary table 1).

FTD-MND

No gene reached gene-wide significant association in the FTD-MND subtype. However, the top genes for FTD-MND were *C9orf72* and *IFNK* with gene-based association p-value in joint-cohort analysis 1.232×10^{-6} and 1.77×10^{-6}, respectively. Neither gene showed associations with any other subtypes of FTD (refer to supplementary table 1).

FTD meta-analysis

The meta-analysis across all subtypes (bvFTD, SD, PNFA and FTD-MND) identified association of *TOMM40* and *APOE*. It is worth noting that the bvFTD samples make more than $1/3^{rd}$ of the total sample; hence, p-values for association with bvFTD dominated the meta-analysis of FTD subtypes.

TABLE 1: Significantly associated genes ($p < 7.388 \times 10^{-07}$) with FTD and its subtypes. Gene-wide significant p-values ($p < 7.388 \times 10^{-07}$) are highlighted in bold.

Phenotype	Symbol	Chr	Start	Stop	Stage1 nSNPs	Stage1 P	Stage2 nSNPs	Stage2 p	Combined p
bvFTD	*TOMM40*	19	45394477	45406946	29	$\mathbf{5.786 \times 10^{-8}}$	13	6.40×10^{-5}	$\mathbf{5.43 \times 10^{-11}}$
	APOE	19	45409039	45412650	5	$\mathbf{1.367 \times 10^{-7}}$	2	1.688×10^{-3}	$\mathbf{1.688 \times 10^{-9}}$
	TOMM40 after removing variants in LD (r^2 >= 0.2) with the epsilon variants	19	45394477	45406946	18	1.949×10^{-5}	5	0.073	7.513×10^{-6}
PNFA	*ARHGAP35*	19	47421933	47508334	25	4.262×10^{-6}	12	7.157×10^{-3}	$\mathbf{2.950 \times 10^{-7}}$
	SERPINA1	14	94843084	94857029	50	3.712×10^{-5}	8	1.193×10^{-3}	$\mathbf{3.024 \times 10^{-7}}$
All FTD	*TOMM40*	19	45394477	45406946	NA	NA	NA	NA	$\mathbf{4.982 \times 10^{-11}}$
	APOE	19	45409039	45412650	NA	NA	NA	NA	$\mathbf{2.179 \times 10^{-10}}$

Risk of *APOE* alleles on FTD subtypes

Based on the gene-based association results with *TOMM40* and *APOE* we extended our analysis to the epsilon alleles and genotypes. We compared each FTD case cohort (discovery, replication and combined) against a total of 9390 ancestry matched controls using Fisher's exact test. For replication cohorts, we used rs769449 as a proxy for rs429358. The distribution of epsilon alleles and genotypes in our cohort is given in supplementary table 4 and 5, respectively. We established the significance threshold for allele associations as 4.167×10^{-3} (0.05/12) correcting for three epsilon alleles and four FTD subtypes. We identified the ε2 allele significantly reduces the risk of bvFTD (odds ratio = 0.772, $p = 3.878 \times 10^{-4}$) and SD (odds ratio = 0.651, $p = 3.642 \times 10^{-3}$). We observed marginal association ($p < 0.05$) of ε2 allele with PNFA (odds ratio = 0.706, p = 0.019) and moderate with FTD-MND (odds ratio = 0.571, $p = 6.008 \times 10^{-3}$). The ε4 allele significantly increased risk of bvFTD (odds ratio = 1.278, $p = 8.14 \times 10^{-6}$) and SD (odds ratio = 1.438, $p = 2.931 \times 10^{-4}$). The association for disease increasing effect of ε4 allele was marginal ($p < 0.05$) for PNFA (odds ratio = 1.29, p = 0.011), but the result was inconclusive for FTD-MND (odds ratio = 1.19, p = 0.20) possibly due to underpowered sample size.

We also quantified the risk of homozygous ε4/ε4 genotype on FTD subtypes. We used 1.25×10^{-2} as a significance threshold for association testing of four subtypes with homozygous ε4/ε4 genotype. The homozygous ε4/ε4 genotype showed significant association with increased risk for bvFTD (odds ratio = 1.62, p = 0.012), PNFA (odds ratio = 2.36, $p = 8.52 \times 10^{-3}$) and SD (odds ratio = 2.33, $p = 9.08 \times 10^{-3}$) with notable odds ratio values for PNFA and SD compared to the effect size of a single copy of ε4

allele for respective FTD subtypes. We did not perform association between homozygous ε2/ε2 genotypes with FTD subtypes due to its low frequency in our cohort.

Table 2: The odds of clinical subtypes of FTD and epsilon alleles, and homozygous ε4/ε4 genotype.

	ε2 allele			ε3 allele			ε4 allele			ε4/ε4 genotype		
Case cohort	Odds ratio	Conf. interval	p-value	Odds ratio	Conf. interval	p-value	Odds ratio	Conf. interval	p-value	Odds ratio	Conf. interval	p-value
bvFTD discovery	0.834	0.696-0.994	0.041	0.908	0.817-1.011	0.076	1.341	1.182-1.517	**5.00×10^{-6}**	1.737	1.090-2.682	0.018
bvFTD replication	0.771	0.598-0.980	0.0311	0.943	0.815-1.094	0.431	1.154	0.961-1.38	0.11	1.407	0.679-2.637	0.27
bvFTD combined	0.772	0.664-0.894	**3.878×10^{-4}**	0.92	0.841-1.006	0.066	1.278	1.147-1.421	**8.14×10^{-6}**	1.627	1.089-2.384	0.0123
PNFA discovery	0.792	0.526-1.150	0.261	0.95	0.755-1.205	0.639	1.246	0.938-1.631	0.12	2.320	0.902-5.015	0.04
PNFA replication	0.602	0.358-0.956	0.028	0.971	0.754-1.266	0.796	1.363	1.006-1.816	0.041	2.423	0.860 5.537	0.046
PNFA combined	0.706	0.515-0.947	0.0193	0.96	0.808-1.145	0.628	1.298	1.056-1.584	0.011	2.367	1.211-4.261	8.518×10^{-3}
SD discovery	0.808	0.554-1.143	0.258	0.764	0.625-0.940	0.01	1.649	1.3-2.073	**4.33×10^{-5}**	2.614	1.152-5.218	0.011
SD replication	0.402	0.198-0.730	**9.473×10^{-4}**	1.217	0.906-1.664	0.208	1.112	0.777-1.552	0.537	1.878	0.497-5.031	0.176
SD combined	0.651	0.470-0.880	**3.642×10^{-3}**	0.898	0.759-1.066	0.205	1.438	1.181-1.741	**2.931×10^{-4}**	2.333	1.194-4.199	9.076×10^{-3}
FTD-MND discovery	0.522	0.289-0.876	9.299×10^{-3}	1.042	0.793-1.388	0.838	1.311	0.948-1.78	0.087	1.32355	0.266-4.028	0.500
FTD-MND replication	0.674	0.303-1.313	0.311	1.134	0.754-1.764	0.622	0.934	0.53-1.544	0.901	0	0.000-3.536	0.629
FTD-MND combined	0.571	0.361-0.862	6.008×10^{-3}	1.070	0.852-1.357	0.612	1.188	0.901-1.544	0.202	0.895	0.181-2.714	1

Functional characterisation of associated genes

We extracted the gene expression profiles of *APOE, TOMM40, ARHGAP35* and *SERPINA1* across different human brain tissues from the GTeX database[20]; see supplementary figure 3 (a, b, c and d for respective genes). The *APOE, TOMM40 and ARHGAP35* genes are strongly expressed in different brain tissues. Notably the anterior cingulate cortex (Bordmann area 24) and the frontal cortex (Bordmann area 9) are the top tissues for *ARHGAP35* gene expression. The Anterior cingulate cortex is one of the early affected regions in FTD patients[23, 24]; this area is reported to be involved in language control and resolving nonverbal conflict[25]. The *SERPINA1* gene did not show strong expression in the brain tissues.

We used the HaploReg[21] (version 4.1) software to investigate functionally annotated variants linked with variants used in deriving gene-based p-values of *TOMM40* (those in negligible LD $r^2 < 0.2$ with epsilon variants), *ARHGAP35*, and *SERPINA1* respectively. We found that all SNPs used in deriving gene-based p-values of *TOMM40*, *ARHGAP35* and *SERPINA1* are in strong LD ($r^2 > 0.8$, in 1000 genomes phase 1 European panel) with at least one variant residing in the regulatory regions such as chromatin marks or DNase hypersensitive sites, suggesting a possible regulatory roles (see supplementary figures 6a, 6b and 6c for HaploReg results for variants in *TOMM40*, *ARHGAP35* and *SERPINA1* respectively). Overall we identified 21, 56 and 93 regulatory variants in LD with SNPs deriving gene-based p-values of *TOMM40*, *ARHGAP35*, and *SERPINA1* genes respectively. We further ranked these regulatory variants based on their functional relevance using the RegulomeDB[22] (version 1.1) software (see supplementary tables 7a, 7b and 7c for detailed RegulomeDB results of these variants mapped to *TOMM40*, *ARHGAP35* and *SERPINA1* genes respectively).

Discussion

Here we report novel genetic insight into FTD and its clinical subtypes using a joint-SNP gene-based approach. We identified association of the *TOMM40* and *APOE* genes with bvFTD, and the *ARHGAP35* and *SERPINA1* genes with PNFA.

Our study suggested *TOMM40* as the top gene in bvFTD. The *TOMM40* gene encodes a channel forming subunit of the translocase of the mitochondrial outer membrane (TOM complex), which facilitates translocation of unfolded proteins from the cytosol into the mitochondrial intermembrane space for use in oxidative phosphorylation[26]. Recently Bannwarth *et al*.[27] reported mitochondrial origin in pathogenesis of FTD-ALS diseases through association of variants in *CHCHD10*. There is growing evidence suggesting a role of mitochondria in neurodegenerative disorders, also including Parkinson's[28, 29], Huntington's[30] and Alzheimer's disease[31, 32].

The association of the *APOE* gene with bvFTD was primarily driven by SNPs rs7412 and rs429358 (or variants in strong LD such as rs769449). The SNPs rs7412 and rs429358 determine the *APOE* epsilon alleles ε2, ε3 and ε4. We quantified the risk of epsilon alleles across clinical subtypes of FTD diagnosed using the Neary's criteria and saw that the ε2 and ε4 alleles showed protective and increased disease risk effects, respectively, for FTD subtypes (strong associations for bvFTD and SD, and marginal associations for PNFA and FTD-MND). Interestingly, individuals carrying homozygous copies of the ε4 allele revealed higher risk for PNFA and SD (odds ratio > 2.3), suggesting dose dependent effect for each copy of a gene. The pattern of association of epsilon alleles with FTD subtypes might reflect on the potential overlap between patients diagnosed with clinical FTD and Alzheimer's disease[33, 34] or a genuine

315 association with FTD and its subtypes given the increasing number of studies arguing in
316 favour of the latter hypothesis[10, 35-38].
317 In the central nervous system (CNS) *APOE* is synthesised in response to neuronal injury
318 or stress to initiate the neuronal repair mechanisms. The ε4 carriers are hypothesised to
319 have reduced neuronal repair capacity compared to the other alleles[37]. The protein
320 products of *APOE* were also reported to modulate neuroinflammation[39]. The
321 hypothesis of enhanced inflammatory response in FTD patients is supported by both
322 neuroimaging and genetic studies[40, 41]. Tau pathology is found in up to ~50% of
323 FTLD cases[42]; interestingly the knock-in study in mice showed association between
324 epsilon alleles and the concentration of hyper-phosphorylated tau in neurons: ε4 knock
325 in mice showed higher concentration of hyper-phosphorylated tau than ε3 knock in
326 mice[43]. It is worth noting that in this scenario the *APOE* locus and the epsilon allelic
327 variability might impact processes such as modulation of neuronal repair mechanisms,
328 neuroinflammation, broad lipid metabolism, synaptic plasticity, neuronal toxicity and
329 tau phosphorylation[44]. It is likely that variability in the genes or isoforms turnover or
330 larger haplotype blocks at this locus, coupled with aging, might influence negative
331 outcomes in brain and thus support our findings from a biological and functional
332 perspective. More work should be directed towards testing these possibilities in the
333 future.
334 Our study is the first to report association of the *ARHGAP35* and *SERPINA1* genes with
335 PNFA. The *ARHGAP35* gene encodes the glucocorticoid receptor DNA-binding factor
336 1, which is a repressor of glucocorticoid receptor (hGR) transcription. At the cellular
337 level, the glucocorticoid receptor mediates the maintenance of basal and stress-related
338 homeostasis. The second gene we found associated with PNFA was *SERPINA1*, which

was previously reported to be associated with cortisol level[45] (top variants: rs11621961, rs12589136, rs2749527) and serum lipid profile[46] (top variant: rs1303). The nonsynonymous variant rs1303 (Glu400Asp) in *SERPINA1*, which is also in moderate LD with morning plasma cortisol level associated variant rs12589136[45], showed PNFA association p-values less than 0.05 in both discovery and replication cohorts (supplementary table 3a). The top *SERPINA1* variant in the PNFA discovery cohort rs11628917 (this variant in moderate LD with rs1303) is an established blood eQTL[47], with the C allele increasing *SERPINA1* expression in blood[46, 47]. We observed that the C allele at rs11628917 increased the risk of PNFA in both discovery and replication cohorts. The *SERPINA1* gene encodes protease inhibitor α1-antitrypsin enzyme, which inhibits cleavage of the reactive centre loop of the corticosteroid binding globulin (CBG) by ceasing neutrophil elastase activity[48]. The reactive centre loop cleavage by neutrophil elastase reduces the CBG binding affinity to cortisol. During stress CBG activity is positively correlated with the glucocorticoid access to the brain[49]. Increased glucocorticoid level in brain can activate glucocorticoid signalling through binding to the low affinity glucocorticoid receptors and result in the reduced neurogenesis and impaired neuroplasticity[50]. This hypothesis suggesting the role of enhanced glucocorticoid signalling leading to neurodegeneration in PNFA patients is based on our current preliminary report on association of *ARHGAP35* and *SERPINA1* with PNFA; this finding will need to be further explored and replicated in an independent cohort.

In conclusion, we report novel genetic associations for frontotemporal dementia and its subtypes – notably of the *TOMM40* and *APOE* genes with bvFTD and the *ARHGAP35* and *SERPINA1* genes with PNFA – using the joint-SNP gene-based approach. This

363 approach improves power of the association test by combining signals across variants in
364 a functional unit such as a gene. Replication and functional characterisation of these
365 findings will further establish their role in pathology of the frontotemporal dementia and
366 help towards a better management of the disease.
367

Acknowledgements

The Netherlands Organization for Scientific Research (NWO VICI 453-14-005) funded this work. RF is supported by the Alzheimer's Society (grant number 284).
We acknowledge the investigators of the original GWAS by IFGC, the acknowledgments as well as the consortia members are listed within the supplementary text (Appendix A and B).
We also acknowledge contributors to the Genotype-Tissue Expression (GTEx) Project, described in the supplementary text (Appendix C).

References

[1] Rabinovici GD, Miller BL. Frontotemporal lobar degeneration: epidemiology, pathophysiology, diagnosis and management. CNS Drugs. 2010;24:375-98.

[2] Seelaar H, Rohrer JD, Pijnenburg YA, Fox NC, van Swieten JC. Clinical, genetic and pathological heterogeneity of frontotemporal dementia: a review. J Neurol Neurosurg Psychiatry. 2011;82:476-86.

[3] Cairns NJ, Bigio EH, Mackenzie IR, Neumann M, Lee VM, Hatanpaa KJ, et al. Neuropathologic diagnostic and nosologic criteria for frontotemporal lobar degeneration: consensus of the Consortium for Frontotemporal Lobar Degeneration. Acta Neuropathol. 2007;114:5-22.

[4] Mackenzie IR, Rademakers R. The molecular genetics and neuropathology of frontotemporal lobar degeneration: recent developments. Neurogenetics. 2007;8:237-48.

[5] Rohrer JD, Guerreiro R, Vandrovcova J, Uphill J, Reiman D, Beck J, et al. The heritability and genetics of frontotemporal lobar degeneration. Neurology. 2009;73:1451-6.

[6] Ferrari R, Thumma A, Momeni P. Molecular Genetics of Frontotemporal Dementia. eLS: John Wiley & Sons, Ltd; 2001.

[7] Woollacott IO, Rohrer JD. The clinical spectrum of sporadic and familial forms of frontotemporal dementia. J Neurochem. 2016.

[8] Van Deerlin VM, Sleiman PM, Martinez-Lage M, Chen-Plotkin A, Wang LS, Graff-Radford NR, et al. Common variants at 7p21 are associated with frontotemporal lobar degeneration with TDP-43 inclusions. Nat Genet. 2010;42:234-9.

[9] Ferrari R, Hernandez DG, Nalls MA, Rohrer JD, Ramasamy A, Kwok JB, et al. Frontotemporal dementia and its subtypes: a genome-wide association study. Lancet Neurol. 2014;13:686-99.

[10] Ferrari R, Grassi M, Salvi E, Borroni B, Palluzzi F, Pepe D, et al. A genome-wide screening and SNPs-to-genes approach to identify novel genetic risk factors associated with frontotemporal dementia. Neurobiol Aging. 2015;36:2904 e13-26.

[11] Yang J, Ferreira T, Morris AP, Medland SE, Genetic Investigation of ATC, Replication DIG, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet. 2012;44:369-75, S1-3.

[12] Zhang J, Zhang Y, Yang J, Zhang L, Sun L, Pan HF, et al. Three SNPs in chromosome 11q23.3 are independently associated with systemic lupus erythematosus in Asians. Hum Mol Genet. 2014;23:524-33.

[13] Tada H, Won HH, Melander O, Yang J, Peloso GM, Kathiresan S. Multiple associated variants increase the heritability explained for plasma lipids and coronary artery disease. Circ Cardiovasc Genet. 2014;7:583-7.

[14] Hagg S, Ganna A, Van Der Laan SW, Esko T, Pers TH, Locke AE, et al. Gene-based meta-analysis of genome-wide association studies implicates new loci involved in obesity. Hum Mol Genet. 2015;24:6849-60.

[15] Liu JZ, McRae AF, Nyholt DR, Medland SE, Wray NR, Brown KM, et al. A versatile gene-based test for genome-wide association studies. Am J Hum Genet. 2010;87:139-45.

[16] de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol. 2015;11:e1004219.

[17] Neary D, Snowden JS, Gustafson L, Passant U, Stuss D, Black S, et al. Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. Neurology. 1998;51:1546-54.
[18] Rascovsky K, Hodges JR, Knopman D, Mendez MF, Kramer JH, Neuhaus J, et al. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. Brain. 2011;134:2456-77.
[19] Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491:56-65.
[20] Mele M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, et al. Human genomics. The human transcriptome across tissues and individuals. Science. 2015;348:660-5.
[21] Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 2012;40:D930-4.
[22] Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. 2012;22:1790-7.
[23] Seeley WW, Carlin DA, Allman JM, Macedo MN, Bush C, Miller BL, et al. Early frontotemporal dementia targets neurons unique to apes and humans. Ann Neurol. 2006;60:660-7.
[24] Seeley WW, Crawford R, Rascovsky K, Kramer JH, Weiner M, Miller BL, et al. Frontal paralimbic network atrophy in very mild behavioral variant frontotemporal dementia. Arch Neurol. 2008;65:249-55.
[25] Abutalebi J, Della Rosa PA, Green DW, Hernandez M, Scifo P, Keim R, et al. Bilingualism tunes the anterior cingulate cortex for conflict monitoring. Cereb Cortex. 2012;22:2076-86.
[26] Mager F, Gessmann D, Nussberger S, Zeth K. Functional refolding and characterization of two Tom40 isoforms from human mitochondria. J Membr Biol. 2011;242:11-21.
[27] Bannwarth S, Ait-El-Mkadem S, Chaussenot A, Genin EC, Lacas-Gervais S, Fragaki K, et al. A mitochondrial origin for frontotemporal dementia and amyotrophic lateral sclerosis through CHCHD10 involvement. Brain. 2014;137:2329-45.
[28] Schapira AH, Cooper JM, Dexter D, Clark JB, Jenner P, Marsden CD. Mitochondrial complex I deficiency in Parkinson's disease. J Neurochem. 1990;54:823-7.
[29] Parker WD, Jr., Boyson SJ, Parks JK. Abnormalities of the electron transport chain in idiopathic Parkinson's disease. Ann Neurol. 1989;26:719-23.
[30] Moran M, Moreno-Lastres D, Marin-Buera L, Arenas J, Martin MA, Ugalde C. Mitochondrial respiratory chain dysfunction: implications in neurodegeneration. Free Radic Biol Med. 2012;53:595-609.
[31] Petrozzi L, Ricci G, Giglioli NJ, Siciliano G, Mancuso M. Mitochondria and neurodegeneration. Biosci Rep. 2007;27:87-104.
[32] Hroudova J, Singh N, Fisar Z. Mitochondrial dysfunctions in neurodegenerative diseases: relevance to Alzheimer's disease. Biomed Res Int. 2014;2014:175062.
[33] Bang J, Spina S, Miller BL. Frontotemporal dementia. Lancet. 2015;386:1672-82.
[34] van der Zee J, Sleegers K, Van Broeckhoven C. Invited article: the Alzheimer disease-frontotemporal lobar degeneration spectrum. Neurology. 2008;71:1191-7.

[35] Engelborghs S, Dermaut B, Marien P, Symons A, Vloeberghs E, Maertens K, et al. Dose dependent effect of APOE epsilon4 on behavioral symptoms in frontal lobe dementia. Neurobiol Aging. 2006;27:285-92.
[36] Agosta F, Vossel KA, Miller BL, Migliaccio R, Bonasera SJ, Filippi M, et al. Apolipoprotein E epsilon4 is associated with disease-specific effects on brain atrophy in Alzheimer's disease and frontotemporal dementia. Proc Natl Acad Sci U S A. 2009;106:2018-22.
[37] Rubino E, Vacca A, Govone F, De Martino P, Pinessi L, Rainero I. Apolipoprotein E polymorphisms in frontotemporal lobar degeneration: a meta-analysis. Alzheimers Dement. 2013;9:706-13.
[38] Seripa D, Bizzarro A, Panza F, Acciarri A, Pellegrini F, Pilotto A, et al. The APOE gene locus in frontotemporal dementia and primary progressive aphasia. Arch Neurol. 2011;68:622-8.
[39] Tai LM, Ghura S, Koster KP, Liakaite V, Maienschein-Cline M, Kanabar P, et al. APOE-modulated Abeta-induced neuroinflammation in Alzheimer's disease: current landscape, novel data, and future perspective. J Neurochem. 2015;133:465-88.
[40] Cagnin A, Rossor M, Sampson EL, Mackinnon T, Banati RB. In vivo detection of microglial activation in frontotemporal dementia. Ann Neurol. 2004;56:894-7.
[41] Rainero I, Rubino E, Cappa G, Rota E, Valfre W, Ferrero P, et al. Pro-inflammatory cytokine genes influence the clinical features of frontotemporal lobar degeneration. Dement Geriatr Cogn Disord. 2009;27:543-7.
[42] Halliday G, Bigio EH, Cairns NJ, Neumann M, Mackenzie IR, Mann DM. Mechanisms of disease in frontotemporal lobar degeneration: gain of function versus loss of function effects. Acta Neuropathol. 2012;124:373-82.
[43] Inbar D, Belinson H, Rosenman H, Michaelson DM. Possible role of tau in mediating pathological effects of apoE4 in vivo prior to and following activation of the amyloid cascade. Neurodegener Dis. 2010;7:16-23.
[44] Verghese PB, Castellano JM, Holtzman DM. Apolipoprotein E in Alzheimer's disease and other neurological disorders. Lancet Neurol. 2011;10:241-52.
[45] Bolton JL, Hayward C, Direk N, Lewis JG, Hammond GL, Hill LA, et al. Genome wide association identifies common variants at the SERPINA6/SERPINA1 locus influencing plasma cortisol and corticosteroid binding globulin. PLoS Genet. 2014;10:e1004474.
[46] Inouye M, Ripatti S, Kettunen J, Lyytikainen LP, Oksala N, Laurila PP, et al. Novel Loci for metabolic networks and multi-tissue expression studies reveal genes for atherosclerosis. PLoS Genet. 2012;8:e1002907.
[47] Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet. 2013;45:1238-43.
[48] Lewis JG, Elder PA. The reactive centre loop of corticosteroid-binding globulin (CBG) is a protease target for cortisol release. Mol Cell Endocrinol. 2014;384:96-101.
[49] Moisan MP, Minni AM, Dominguez G, Helbling JC, Foury A, Henkous N, et al. Role of corticosteroid binding globulin in the fast actions of glucocorticoids on the brain. Steroids. 2014;81:109-15.
[50] Anacker C, Zunszain PA, Carvalho LA, Pariante CM. The glucocorticoid receptor: pivot of depression and of antidepressant treatment? Psychoneuroendocrinology. 2011;36:415-25.