Genetics of dementia in a Finnish cohort

Petra Pasanen[1,2], Liisa Myllykangas[3], Minna Pöyhönen[4], Anna Kiviharju[8], Maija Siitonen[1], John Hardy[5], Jose Bras[5,6,7], Anders Paetau[3], Pentti J. Tienari[8,9], Rita Guerreiro[5,6,7], and Auli Verkkoniemi-Ahola[8]

Running title: Genetics of dementia in Finnish families

Author affiliations:

[1] Department of Medical Biochemistry and Genetics, Institute of Biomedicine, University of Turku, Turku, Finland

[2] Tyks Microbiology and Genetics, Department of Medical Genetics, Turku University Hospital, Turku, Finland

[3] Department of Pathology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

[4] Department of Clinical Genetics, Helsinki University Central Hospital and Department of Medical Genetics, University of Helsinki, Helsinki

[5] Department of Molecular Neuroscience, Institute of Neurology, University College London, London, UK

[6] UK Dementia Research Institute at UCL

[7] Department of Medical Sciences and Institute of Biomedicine – iBiMED, University of Aveiro, 3810-193 Aveiro, Portugal

[8] Clinical Neurosciences, Neurology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

[9] Research Programs Unit, Molecular Neurology, University of Helsinki, Helsinki, Finland

Corresponding author: Petra Pasanen

Address: Kiinamyllynkatu 10 FIN-20520 Turku

Phone: +358 2 333 7233

e-mail: petra.pasanen@utu.fi

Conflict of interest

The authors declare that they have no conflict of interest.

Abstract

Alzheimer's disease (AD) and frontotemporal dementia (FTD) are the two most common neurodegenerative dementias. Variants in *APP*, *PSEN1* and *PSEN2* are typically linked to early-onset AD, and several genetic risk loci are associated with late-onset AD. Inherited FTD can be caused by hexanucleotide expansions in *C9orf72*, or variants in *GRN*, *MAPT* or *CHMP2B*. Several other genes have also been linked to FTD or FTD with motor neuron disease.

Here we describe a cohort of 60 Finnish families with possible inherited dementia. Our aim was to clarify the genetic background of dementia in this cohort by analysing both known dementia-associated genes and searching for rare or novel segregating variants with exome sequencing. *C9orf72* repeat expansions were detected in 12 (20%) of the 60 families, including, in addition to FTD, a family with neuropathologically verified AD. Twelve families (10 with AD and 2 with FTD) with representative samples from affected and unaffected subjects and without *C9orf72* expansions were selected for whole exome sequencing. Exome sequencing did not reveal any variants that could be regarded unequivocally causative of the disease in the families, but revealed potentially damaging variants in *UNC13C* and *MARCH4*.

Introduction

Alzheimer's disease (AD) and frontotemporal dementia (FTD) are the two most common neurodegenerative dementias. AD is characterized by progressive loss of memory, typically presenting with deficits in anterograde episodic memory. Other cognitive functions, such as language, executive functions and visuospatial functions, deteriorate as the disease progresses [1]. Most AD patients first develop symptoms after 65 years of age (late-onset AD, LOAD), while less than 10% of patients present with early-onset AD (EOAD). Autosomal dominant inheritance and rare cases of autosomal recessive inheritance are seen in the EOAD group, due to variants in the amyloid precursor protein (*APP*), presenilin 1 (*PSEN1*) and presenilin 2 (*PSEN2*) genes. Familial cases are also seen in LOAD with an estimated heritability of 58 to 79% [2]. More than 20 disease-associated loci have been detected in genome-wide association studies (GWAS)[3-7] and meta-analyses [8] but, apart from the *APOE* ε4 and *TREM2* p.(R47H) risk alleles, most of these only have a modest effect. Although rare variants in *APP*, *PSEN1* and *PSEN2* have been detected in LOAD patients by targeted resequencing [9], variants affecting function are rare in LOAD and its pathobiology reflects the interplay of predisposing genetic variants and environmental factors. Sequencing studies have also shown that rare variants can be found in dementia-associated loci identified through GWAS [10].

In contrast to AD, FTD is more commonly observed in patients younger than 65 years [11]. FTD may present with changes in personality and behaviour (behavioural-variant FTD) or language difficulties (non-fluent variant primary progressive aphasia and semantic-variant progressive aphasia) [11]. Up to 40% of patients have a positive family history with autosomal dominant inheritance in 10% [12]. The most commonly mutated genes are *C9orf72* [13,14], *GRN* [15,16] and *MAPT* [17], while rare variants in *TARDBP* [18], *FUS* [19,20], *VCP* [21], *CHMP2B* [22], *UBQLN2* [23,24], *TBK1*[25], *SQSTM1* [26] and *CCNF* [27] have been detected in patients with either FTD and motor neuron disease (FTD-MND) or pure FTD. *C9orf72* expansions are prevalent

3

77   in Finnish patients with FTD or ALS, accounting for 48% and 46% of familial FTD and ALS,

78   respectively, and 19% and 21% of sporadic FTD and ALS[13].

79   Based on the known functions of disease-associated genes, several pathways involved in the

80   pathogenesis of AD and FTD have been identified. In AD, these include the amyloid β

81   pathway, the immune system (*CLU*, *CR1*, *ABCA7*, *CD33*, *EPHA1*, the *MS4A* gene cluster),

82   synaptic activity (*PICALM*, *CD33*, *CD2AP*, *EPHA1*, *BIN1*) and lipid metabolism (*CLU*,

83   *ABCA7*) (reviewed in [28]). In FTD, the disease-implicated pathways include RNA processing

84   and transcription regulation (*C9orf72, TARDBP, FUS*), microtubule function (*MAPT*),

85   immune response (*GRN*), lysosome-mediated and ubiquitin-mediated protein degradation and

86   autophagy (*GRN, VCP, CHMP2B*) (reviewed in [11]).

87   Here we describe a cohort of 60 Finnish families with possible inherited dementia. Our aim

88   was to clarify the genetic background of dementia in this cohort by analysing both known

89   dementia-associated genes and searching for rare or novel segregating variants using exome

90   sequencing. We show that *C9orf72* hexanucleotide repeat expansions are common in this

91   cohort but variants affecting function of the other most common AD and FTD genes are not

92   accountable for the disease in these families. We also present rare variants that segregate with

93   AD and FTD in small families. Although no definite conclusion can be achieved regarding

94   the causal involvement of these rare variants, these should be taken into account in future

95   studies trying to identify the genetic cause of familial dementias.

96   **Subjects and methods**

97   **Study cohort**

98   The study cohort is comprised of affected and unaffected members from 60 Finnish families

99   with possible inherited dementia. The families were recruited from neurology clinics in the

100 Helsinki and Uusimaa hospital district (Southern Finland) and via an advertisement in a

101 national newspaper in the late 1990s. The recruitment method proved particularly successful,

102 resulting in 60 families suited for the study. A total of 364 blood-derived DNA samples (107

103 from affected patients and 257 from unaffected family members) were available from the

104 families. A prerequisite for participation was a positive family history with two or more

105 living first-degree family members affected by dementia.

106 The clinical diagnosis was AD in most families (n = 38), FTD in 10 families, dementia with

107 Lewy bodies (DLB) in one family and unspecified dementia in 11 families (Supplementary

108 table 1). The diagnoses were based on clinical findings and brain CT imaging studies.

109 Extensive neuropsychological studies had been performed for some of the patients. Liquor

110 biomarkers were not available at the time of patient recruitment and sample collection.

111 Neuropathological data was available from patients belonging to seven of the 60 families.

112 The ages at onset are listed in Supplementary table 2.

113 **Ethical aspects**

114 Informed consent was obtained from all participants. The study was approved by the Ethics

115 committee of Neurology department at HUCH (4.6.1997 and 11.1.2012) and the HUCH

116 Ethics Committee of Medicine (Dnro104/13/03/01/14). Approval for using patient tissue

117 specimens was given by Valvira (Dnro 2855/06.01.03.01/2012). Approval for using medical

118 records and autopsy reports of the patients living outside the HUS district was obtained from

119 National Institute for Health and Welfare (Dnro THL/701/5.05.00/2013).

120 **Methods**

121 EDTA blood samples were drawn after obtaining informed consent from the participants.

122 Both affected patients and unaffected family members were recruited for the study. DNA

123 was extracted using standard protocols. Overview of the study scheme is shown in Figure

124 1. *APOE* genotypes were determined by PCR and *Cfo*I digestion as described in

125 Myllykangas *et al.* [29]. Screening of *C9orf72* expansions was done by repeat primed PCR

126 as described by Renton *et al.* [13].

127 Expansion was defined by two criteria that had to be fulfilled: 1. Characteristic saw-tooth

128 pattern in repeat-primed PCR extending over 30 G_4C_2 repeats on capillary electrophoresis

129 of the PCR products, and 2. Lack of large allele (>30 repeats) amplicon in standard PCR

130 across the repeat region. Standard PCR across the repeat region was performed using

131 LongAmp Taq Reaction Buffer (New England Biolabs) with the following PCR primers

132 5'-GGA GGG AAA CAA CCG CAG CCT GTA G-3'and 5'-ATG CCG CCT CCT CAC

133 TCA CCC ACT CG-3', 1.8 M of Betaine. The PCR products were run on 2 % agarose

134 gels.

135 We evaluated the pedigrees for the availability of samples from both affected and unaffected

136 individuals as well as the availability of neuropathological data. Twelve families without

137 *C9orf72* expansions (2 with FTD and 10 with AD) were selected for further genetic studies.

138 The FTD families were screened for variants in *GRN* and the AD families for variants in

139 exons 16 and 17 of *APP* and the coding regions of *PSEN1* and *PSEN2*. Exons and flanking

140 splice site regions were amplified by PCR and the purified PCR products were sequenced in

141 both directions using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied

142 Biosystems, CA, USA). All primer sequences and PCR conditions are available upon request.

143 Large structural and copy number variants were excluded by using HumanOmniExpress

144 Bead Chip (Illumina, San Diego, CA, USA). Loci known to have copy number variants that

145 associate with dementia (such as APP and SNCA) were checked visually. In addition, the

146 data was analysed with CNVPartition in Genome Studio (Illumina, San Diego, CA, USA) to

147 detect large (> 50 kb) CNVs. The identified CNVs were checked against the Database of

148 Genomic Variants (DGV) [30].

149 Whole exome sequencing (WES) of selected individuals was done at University College

150 London (UCL, London, UK). Exome enrichment was performed using TruSeq Exome

151 Capture kit (Illumina, San Diego, CA, USA). Sequencing was performed on a HiSeq 2000

152 (Illumina, San Diego, CA, USA). Reads were aligned to GRCh37/hg19 using BWA, variants

153 called according to GATK best practice guidelines and annotated with ANNOVAR [31]. *In*

154 *silico* pathogenicity predictions of nonsynonymous variants were done with SIFT [32],

155 Polyphen2 [33], MutationTaster [34], MutationAssessor [35], and CADD [36]. Variants were filtered

156 against population databases (1000Genomes, ESP and ExAC) and prioritised based on

157 variant type (missense, nonsense, splice site, frameshift, non-frameshift) and predicted

158 pathogenicity. We concentrated on variants found in genes implicated in GWAS or genes that

159 are highly expressed in the brain. We also assessed the known functions of the genes of

160 interest.

161 Selected variants (shared by the affected members in each family but not present in the

162 unaffected family member, if appropriate sample was available) from WES were confirmed

163 with Sanger sequencing and their segregation tested in a family setting. These variants and

164 the associated phenotypes have been submitted to ClinVar

165 (https://www.ncbi.nlm.nih.gov/clinvar/) with accession numbers SCV000576395,

166 SCV000576396, SCV000576397, SCV000576398, and SCV000576399. We also checked

167 the frequencies of these selected variants in SISu, a database of sequence variants in Finns

168 (Sequencing Initiative Suomi project (SISu), Institute for Molecular Medicine Finland

169 (FIMM), University of Helsinki, Finland (URL: http://sisuproject.fi), SISu v4.1, accessed in

170 09/2016).

171

Results

173 *APOE* genotyping

174 *APOE* genotyping was performed for 364 samples from 60 families. There were 34

175 individuals homozygous for the risk *APOE* genotype ε4 (20/107 affected, 18.7%; 14/257

176 unaffected, 5.4%). A total of 166 individuals were heterozygotes *APOE* ε3/ε4 (56/107

177 affected, 52.3%; 110/257 unaffected, 42.8%). The most common allele, *APOE* ε3, is not

178 associated with an increased risk for AD, and it was detected in 28 of 107 (26.1%) affected

179 patients and in 123 of 257 (47.8%) unaffected individuals. Five unaffected individuals

180 (5/257, 1.99%) were *APOE* ε2/ε4 heterozygotes and one unaffected (1/257, 0.3%) was ε2/ ε3.

181 No genotype was obtained for seven samples (3 affected and 4 unaffected). The genotypes in

182 each family are shown in Supplementary table 2.

183 *C9orf72*

184 All 60 families were included in the *C9orf72* hexanucleotide expansion screening.

185 Expansions were detected in 12 of the 60 families (20%). The distribution of expansions in

186 affected and unaffected individuals in each family is shown in Supplementary table 2.

187 Clinical diagnosis was FTD in seven families, AD or variant AD in four families and

188 degenerative dementia (ALS, DLB or AD resembling syndromes) in one family (Table 1).

189 The proportions of *C9orf72* expansions in each diagnosis group in shown in Figure 2. As

190 FTD or FTD/ALS are the typical clinical phenotypes associated with *C9orf72* expansions, we

191 only describe the five families with more atypical presentations in detail. No additional

192 information was available from family Fam-62.

193 Table 1. Families with *C9orf72* expansions. Two families (Fam-18 and Fam-73) marked with

194 an asterisk may have been initially misdiagnosed as AD. In Fam-25 (marked with two

195 asterisks) the neuropathological diagnoses are based on haematoxylin & eosin and ancillary

196 stainings. No immunohistochemical stainings for FTLD were available at the time of the

197 neuropathological examination.

Family ID	Clinical diagnosis	Neuropathological diagnosis
Fam-18	AD*+FTD	not available
Fam-22	FTD	not available
Fam-25	FTD	FTLD + AD-type lesions**
Fam-27	dementia (AD)	not available
Fam-31	degenerative dementia (ALS, DLB or AD resembling syndromes)	AD
Fam-33	FTD	not available
Fam-39	FTD	not available
Fam-50	AD	not available
Fam-62	AD	not available
Fam-71	FTD	not available
Fam-73	variant AD*	not available
Fam-77	FTD/ALS	not available

198

199 The index patient of family Fam-18 developed symptoms at 65 years. Clinical presentation

200 was compatible with AD, but CT and MRI studies showed frontal atrophy. Clinical data from

201 other family members was not available.

202 Patient records of two affected patients from family Fam-27 were available for review. Both

203 patients had clinical AD with late onset.

204 Neuropathological data of two affected patients from family Fam-31 were available. The

205 index patient had late-onset AD and the diagnosis was verified neuropathologically post

206 mortem in 1998 using the methods available at that time. Re-analysis of the archived

207 formalin-fixed, paraffin-embedded sample revealed Braak stage V tau-pathology and

208 CERAD stage B beta-amyloid load. TDP-43 staining was negative. p62 positive inclusions

209 were observed in the granular cerebellar cells. No DNA sample was available for study.

210 Formalin-fixed brain tissue from temporal lobe of another patient of the family was available

211 for immunohistochemistry. TDP-43 staining was negative, but moderate to severe tau-

212 pathology suggestive of AD was observed. This patient was shown to harbor a *C9orf72*

213 expansion and carried the *APOE* ε4/4 genotype.

214 Family Fam-50 included several affected individuals with onset of disease after 70 years of

215 age. DNA sample was available from one of them. One patient had visual hallucinations as

216 the first symptom and subsequently developed loss of concentration and memory deficit.

217 Neuropsychological examination was consistent with large-scale impairment and visual

218 defect.

219 The index patient of family Fam-73 was diagnosed with variant AD. However, brain SPECT

220 was suspective of FTD. The index patient's sister had been diagnosed with ALS. Thus, the

221 actual diagnostic spectrum of this family is consistent with FTD and ALS.

222 *Further studies on 12 families without C9orf72 expansions*

223 *Exclusion of GRN, APP, PSEN1, PSEN2*

224 Twelve families without *C9orf72* expansions and representative samples from both affected

225 and unaffected individuals were selected for further studies (Figure 1.). The clinical

226 diagnoses in these families were FTD (two families: Fam-13 and Fam-59) and AD (10

227 families: Fam-15, Fam-29, Fam-32, Fam-35, Fam-38, Fam-49, Fam-52, Fam-55, Fam-56,

228 and Fam-57). Sanger sequencing did not reveal any causal variants in *GRN* (FTD), *APP*

229 exons 16 and 17, or coding regions of *PSEN1* and *PSEN2* (AD families).

230 *Exclusion of large structural and copy number variants by SNP microarray*

231 None of the 12 index patients had a duplication of *APP*. Neither did we identify any deletions

232 or duplications involving other known dementia-associated genes, such as *SNCA*.

233 *Whole exome sequencing*

234 Whole exome sequencing (WES) data was generated for at least two affected patients from

235 each of the 12 families. The oldest unaffected family members from whom a DNA sample

236 was available (7 families) were exome sequenced as controls. We concentrated on rare

237 variants identified in WES shared by the affected patients in each family but not seen in the

238 analysed healthy family members, when available (list of rare variants in each family can be

239 found in a Supplementary table 3). Confirmation and segregation analyses were done with

240 Sanger sequencing. A large number of shared rare variants were identified in each family, but

241 we concentrated on variants in GWAS hit genes or in genes with known functions possibly

242 relevant for neurodegeneration. The validated variants are listed in Table 2 and presented in

243 detail below.

244 Table 2. Rare variants identified in exome sequencing and validated in a family setting.

245 Minor allele frequencies (MAFs) of each variant in 1000Genomes, ESP, SiSU and ExAC are

246 shown. Different prediction programs (SIFT, PolyPhen, MutationTaster, MutationAssessor,

247 CADD) were used to estimate the deleteriousness of the variants.

Gene	*CLU*	*PCDH11X*	*UNC13C*	*MARCH4*	*MARCH4*
Family	Fam-56	Fam-15	Fam-49	Fam-59	Fam-13
clinical diagnosis	AD	AD	AD	FTD	FTD
genomic location (hg19)	27462662	91133518	54306424	217234945	217148338
chromosome location	8p21.1	Xq21.31	15q21.3	2q35	2q35
reference sequence	NM_001831.3	NM_001168360.4	NM_001080534.2	NM_020814.2	NM_020814.2
cDNA change	c.608C>T	c.2279A>T	c.1324_1326del	c.39G>C	c.631A>G
amino acid change	p.(Thr203Ile)	p.(Asp760Val)	p.(Lys443del)	p.(Trp13Cys)	p.(Lys211Glu)
rs identifier	rs41276297	rs781770086	rs746069739	rs145386484	rs756981946

SIFT	tolerated	Damaging	-	tolerated	damaging
PolyPhen	Benign	possibly damaging	-	benign	damaging
MutationTaster	Neutral	damaging	-	damaging	damaging
MutationAssessor	medium effect	neutral	-	neutral	neutral
CADD phred score	8.521	14.33	-	2.416	19.62
1000G	0.005	absent	Absent	0.000599042	absent
ESP	0.0027	absent	Absent	7.7e-05	absent
SISu	0.000698324	0.00017454	Singleton	0.00124611	singleton
ExAC (Finnish)	0.00121	singleton	Absent	0.001285	absent
ExAC (total)	0.001673	singleton	Absent	0.0002082	singleton
Significance	likely benign	possibly benign	possibly deleterious	likely benign	possibly deleterious

248

250 *CLU*

251 A heterozygous *CLU* c.608C>T, p.(Thr203Ile) variant (rs41276297) was identified in two

252 affected patients of family Fam-56. The variant was not detected in the four unaffected

253 family members from whom a sample was available (Figure 3a). This variant is a previously

254 known, rare variant with a frequency of 0.00121 in Finnish samples in ExAC. Polyphen and

255 SIFT predicted no deleterious effect. This variant has also been detected in British AD

256 samples (reported as p.T255I) with a frequency of 0.003 as well as in unaffected controls

257 (frequency 0.006)[37]. One of the affected individuals also carried one *APOE* ε4 allele, while

258 the other was homozygous for ε3.

259 *PCDH11X*

260 Affected members of family Fam-15 carried a heterozygous c.2279A>T, p.(Asp760Val)

261 variant in *PCDH11X*. This variant (rs781770086) is present as a singleton in ExAC and in

262 SISu. Sanger sequencing confirmed the variant in the two affected patients (II:4 and III:4).

263 However, segregation analysis showed that the variant was also present in two currently

264   unaffected individuals (III:2 and III:5) and in one individual (III: 6) with unclear status. The

265   remaining two unaffected family members (II:7 and III:7) did not carry the variant (Figure

266   3b). Only one of the affected individuals carried *APOE* ε4.

267   *UNC13C*

268   In family Fam-49, a heterozygous 3-bp deletion in *UNC13C*, c.1324_1326del, p.(443del) was

269   detected in two affected patients. The variant was not seen in two unaffected family members

270   (Figure 3c). This in-frame deletion variant (rs746069739) is present as a singleton in ExAC

271   and in SISu (low-quality). In addition to the *UNC13C* variant, both affected individuals also

272   carried one *APOE* ε4 allele.

273   <mark>*WES findings in FTD families*</mark>

274   *MARCH4*

275   Two affected patients from <mark>the FTD</mark> family Fam-13 carried a heterozygous c.631A>G,

276   p.(Lys211Glu) variant, (rs756981946) in *MARCH4*. This variant was absent from the

277   unaffected family members (Figure 3d). The *APOE* genotypes of the two affected individuals

278   of Fam-13 were ε3/4 and ε3/3.Two affected members of the second FTD family, Fam-59,

279   carried a heterozygous c.39C>G variant, p.(Trp13Cys) (rs145386484) in *MARCH4*.

280   Segregation analysis showed that this variant was also present in 7 currently unaffected

281   family members (ages 45 – 73 years) and absent in other 9 unaffected family members

282   (Figure 3e). All studied individuals in Fam-59 were homozygous for *APOE* ε3.

283   *MARCH4* p.(Lys211Glu) variant is present in ExAC as a singleton in a non-Finnish European

284   sample and in SISu as a singleton (identifier: rs756981946). In silico predictions gave the

285   following results: Polyphen2 predicted the variant to be probably damaging (score 0.995),

286   SIFT tolerated (score 0.29), MutationTaster damaging (score 1.000), MutationAssessor

287 medium effect (score 2.22), CADD Phred-like scaled C-score was 19.62. These data

288 demonstrate that the variant is extremely rare and suggest that it might alter the normal

289 function of *MARCH4*.

290 The p.(Trp13Cys) variant is more common as it is reported in ExAC with a frequency of

291 0.001285 in Finnish samples. It has also been detected in other populations: European

292 (3/56914), South Asian (8/13234), African (3/7972) and Latino (1/9762). SIFT predicted this

293 variant to be tolerated (score 0.29), Polyphen2 benign (score 0.00), MutationTaster damaging

294 (score 0.981), MutationAssessor neutral (score -0.55), and CADD Phred-like scaled C-score

295 was 2.416. These predictions along with the fact that it was present in individuals over 70

296 years of age suggest that p.(Trp13Cys) might be a rare neutral variant.

297 **Discussion**

298 In contrast to early-onset AD, late-onset Alzheimer's disease is rarely caused by segregating

299 variants in families. The strongest identified risk factor is the *APOE ε4* allele. In few cases,

300 variants in *APP*, *PSEN1* and *PSEN2* have been reported in LOAD families[9]. GWAS studies

301 have identified approximately 20 loci associated with predisposition to AD but finding

302 variants that actually have a biological effect has proven difficult. In FTD, variants in

303 *C9orf72*, *MAPT* and *GRN* account for up to 60% of familial cases while variants in other

304 genes are rare [11].

305 In addition to the ALS/FTD entity, *C9orf72* expansions have been linked to several other

306 clinical manifestations including AD, Parkinson's disease and Huntington's disease

307 phenocopies (reviewed in [38]). We detected *C9orf72* expansions in 7/60 (11.6%) families with

308 either FTD or FTD/ALS but also in 3/60 (5%) families with clinical AD. In one family, Fam-

309 31, neuropathological examinations disclosed moderate to severe AD tau-pathology, and no

310 TDP-43-positive inclusions were seen. However, p62-positive inclusions were present in the

cerebellum, consistent with the *C9orf72* expansion. Several earlier reports have described

C9orf72 expansions in either clinically diagnosed [39-42] or neuropathologically confirmed AD

[43]. It is possible that the AD pathology is at least partly attributable to *APOE* as the one

affected individual with *C9orf72* expansion and AD-type neuropathology was homozygous

for the *APOE* ε4 allele (Supplementary table 2). Previous work has shown that *C9orf72*

expansions are seen in ~30% of Finnish FTD patients [13] and in 48.1% of familial FTD [44]. Our

results confirm this finding and suggest that *C9orf72* expansions may manifest as clinical AD

and some patients may also show concomitant AD pathology at the neuropathological

examination. Previous studies on *C9orf72* expansions in AD patients have suggested that the

clinical or neuropathological classification as AD may have been incorrect, and this appeared

to be the case in some of our families with clinical diagnosis of AD.

No variants in *APP* exons 16 and 17 or the coding regions of *PSEN1* and *PSEN2* were

observed in the 10 AD families selected from our cohort. Whole-genome genotyping also

showed no clearly causative CNVs. Both results are in agreement with previous studies. Only

a few *PSEN1* variants have been reported in Finnish AD families: two families carry the

'Cotton-wool' variant, Δ9Finn (c.869_955del) [45], p.(Met146Val) has been reported in a

Swedish family of Finnish descent [46,47] and p.(Pro264Leu) in one family [48]. Screening of

APP, *PSEN1* and *PSEN2* in a cohort of 140 EOAD patients revealed no variants that might

affect function [49]. In addition, duplication of *APP* was not detected in a cohort of 64 Finnish

EOAD patients [50].

GRN sequencing and exome sequencing did not reveal any pathogenic variants in the two

FTD families without *C9orf72* expansions. In agreement with our results, previous work

suggests that GRN variants are rare among Finnish FTD patients [51].

Exome sequencing revealed rare, potentially relevant variants in five families. Two variants were in genes previously linked to AD (*CLU* and *PCDH11X*) while three variants were in genes (*UNC13C* and *MARCH4*) that have not been directly linked to dementia but could be important in maintaining normal neuronal function.

In 2010, a large GWAS study indicated that *PCDH11X* was linked to LOAD in a combined American Caucasian cohort[52]. However, subsequent studies in different populations failed to confirm the findings of Carrasquillo *et al.* [53-56] Recently, Jiao *et al.* reported a SNP in *PCDH11X* to confer a risk to LOAD [57]. Thus, the possible role of *PCDH11X* in AD susceptibility is still somewhat unclear. Our results show that the rare p.(Asp760Val) variant is present in all affected individuals of family Fam-15 but also in two asymptomatic individuals and in one subject with unclear status.

The role of *CLU* as an AD risk gene has been established in independent datasets [3,4]. We noted co-segregation of a rare *CLU* variant and dementia in an AD family (Fam-56). Even though rare non-synonymous and small insertion/deletion variants have been reported to increase AD risk[58,59], the p.(Thr203Ile) variant is predictably not deleterious, but at present we cannot exclude its possible role in AD risk.

Two AD patients from family Fam-49 shared a 3bp in-frame deletion in *UNC13C*. The *UNC13C* gene is highly expressed in brain. Experimental evidence from cat and mouse models have suggested that its mammalian homologue, *Munc13-3*, has a role in controlling critical-period neuronal plasticity in visual cortex.[60,61] Gene expression studies in human AD and control brain samples showed increased *UNC13C* expression in hippocampal CA3 compared to CA1 in Alzheimer patients. This implicates that *UNC13C* might have a neuroprotective role in the brain [62]. The rare variant found in family 49 removes one amino

357   acid residue but does not disturb the reading frame. Both affected patients were also

358   heterozygous for the *APOE* ε4 allele, a likely risk factor in this family.

359   A rare segregating missense variant in *MARCH4* was identified in the FTD family Fam-13.

360   *MARCH4* is a member of membrane-associated RING-CH family of ubiquitin E3 ligases.

361   These ligases function in the last step of ubiquitination by recruiting the ubiquitin carrying E2

362   enzyme and transferring ubiquitin from E2 to the target protein[63]. *MARCH4* is predominantly

363   expressed in the adult human brain [64]. The ubiquitin-related protein degradation pathway has

364   been implicated in many neurodegenerative diseases, including FTD. Recent work by

365   Williams and coworkers described variants in a component of the ubiquitin E3 ligase

366   complex, *CCNF*, in a large ALS/FTD family and a few singleton patients [27]. Although the

367   *MARCH4* variant segregated with FTD in our small family, we cannot exclude the possibility

368   that we merely identified a rare neutral variant in a gene with function that could fit in the

369   model of FTD pathogenesis.

370   *C9orf72* repeat expansions are common among Finnish FTD patients and our results indicate

371   that expansions may also be seen in patients with clinical and neuropathological diagnoses of

372   AD. Our results suggest that unknown genetic factors are likely to be responsible for a

373   proportion of familial dementia in the Finnish cohort, but definitely causal or risk variants in

374   novel genes are yet to be identified. Exome sequencing is an efficient way to search for rare

375   coding variants, but thus far only few segregating risk variants (e.g. *TREM2* p.(Arg47His) [65]

376   and *TTC3* p.(Ser1038Cys) [66]) have been described in LOAD families. Our results

377   corroborate the view that even in large LOAD families with multiple affected individuals the

378   disease is likely caused by combination of multiple genetic and environmental risk factors.

379   The *APOE* ε4 risk allele can be assumed to account for multiple affected individuals in

380   several of the AD families in our study.

We detected rare segregating coding variants in *UNC13C* in an AD family and in *MARCH4* in an FTD family. However, replication in larger familial and case-control datasets and functional assays would be needed to prove their causality. The limitation or our study is the relatively small number of patients. Thus, we could only aim to find highly penetrant pathogenic variants. In addition, exome sequencing does not enable the identification of non-coding variants that might affect splicing or gene expression.

While our exome sequencing approach failed to identify any clearly causal variants in the 12 families, we believe that the rare variants found in our cohort will be of interest for other dementia researchers. Thus, we presented all the variants and genes of potential interest in the hope this may be useful for future studies and can facilitate analyses in other families and datasets.

Acknowledgements

References

404   1.   McKhann GM, Knopman DS, Chertkow H *et al*: The diagnosis of dementia due to Alzheimer's
405   disease: recommendations from the National Institute on Aging-Alzheimer's Association
406   workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 2011; **7:**
407   263-269.

408
409   2.   Gatz M, Reynolds CA, Fratiglioni L *et al*: Role of genes and environments for explaining
410   Alzheimer disease. *Arch Gen Psychiatry* 2006; **63:** 168-174.

411
412   3.   Lambert JC, Heath S, Even G *et al*: Genome-wide association study identifies variants at CLU
413   and CR1 associated with Alzheimer's disease. *Nat Genet* 2009; **41:** 1094-1099.

414
415   4.   Harold D, Abraham R, Hollingworth P *et al*: Genome-wide association study identifies
416   variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet* 2009; **41:** 1088-
417   1093.

418
419   5.   Seshadri S, Fitzpatrick AL, Ikram MA *et al*: Genome-wide analysis of genetic loci associated
420   with Alzheimer disease. *JAMA* 2010; **303:** 1832-1840.

421
422   6.   Hollingworth P, Harold D, Sims R *et al*: Common variants at ABCA7, MS4A6A/MS4A4E,
423   EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet* 2011; **43:** 429-
424   435.

425
426   7.   Naj AC, Jun G, Beecham GW *et al*: Common variants at MS4A4/MS4A6E, CD2AP, CD33 and
427   EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet* 2011; **43:** 436-441.

428
429   8.   Lambert JC, Ibrahim-Verbaas CA, Harold D *et al*: Meta-analysis of 74,046 individuals
430   identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 2013; **45:** 1452-1458.

431
432   9.   Lee JH, Kahn A, Cheng R *et al*: Disease-related mutations among Caribbean Hispanics with
433   familial dementia. *Mol Genet Genomic Med* 2014; **2:** 430-437.

434
435   10.   Vardarajan BN, Ghani M, Kahn A *et al*: Rare coding mutations identified by sequencing of
436   Alzheimer disease genome-wide association studies loci. *Ann Neurol* 2015; **78:** 487-498.

437
438   11.   Bang J, Spina S, Miller BL: Frontotemporal dementia. *Lancet* 2015; **386:** 1672-1682.

439
440   12.   Rohrer JD, Guerreiro R, Vandrovcova J *et al*: The heritability and genetics of frontotemporal
441   lobar degeneration. *Neurology* 2009; **73:** 1451-1456.

442
443   13.   Renton AE, Majounie E, Waite A *et al*: A hexanucleotide repeat expansion in C9ORF72 is the
444   cause of chromosome 9p21-linked ALS-FTD. *Neuron* 2011; **72:** 257-268.

445

446   14.   DeJesus-Hernandez M, Mackenzie IR, Boeve BF *et al*: Expanded GGGGCC hexanucleotide
447   repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron*
448   2011; **72:** 245-256.

449
450   15.   Baker M, Mackenzie IR, Pickering-Brown SM *et al*: Mutations in progranulin cause tau-
451   negative frontotemporal dementia linked to chromosome 17. *Nature* 2006; **442:** 916-919.

452
453   16.   Cruts M, Gijselinck I, van der Zee J *et al*: Null mutations in progranulin cause ubiquitin-
454   positive frontotemporal dementia linked to chromosome 17q21. *Nature* 2006; **442:** 920-924.

455
456   17.   Hutton M, Lendon CL, Rizzu P *et al*: Association of missense and 5'-splice-site mutations in
457   tau with the inherited dementia FTDP-17. *Nature* 1998; **393:** 702-705.

458
459   18.   Benajiba L, Le Ber I, Camuzat A *et al*: TARDBP mutations in motoneuron disease with
460   frontotemporal lobar degeneration. *Ann Neurol* 2009; **65:** 470-473.

461
462   19.   Van Langenhove T, van der Zee J, Sleegers K *et al*: Genetic contribution of FUS to
463   frontotemporal lobar degeneration. *Neurology* 2010; **74:** 366-371.

464
465   20.   Huey ED, Ferrari R, Moreno JH *et al*: FUS and TDP43 genetic variability in FTD and CBS.
466   *Neurobiol Aging* 2012; **33:** 1016.e1019-1017.

467
468   21.   Watts GD, Wymer J, Kovach MJ *et al*: Inclusion body myopathy associated with Paget disease
469   of bone and frontotemporal dementia is caused by mutant valosin-containing protein. *Nat*
470   *Genet* 2004; **36:** 377-381.

471
472   22.   Skibinski G, Parkinson NJ, Brown JM *et al*: Mutations in the endosomal ESCRTIII-complex
473   subunit CHMP2B in frontotemporal dementia. *Nat Genet* 2005; **37:** 806-808.

474
475   23.   Deng HX, Chen W, Hong ST *et al*: Mutations in UBQLN2 cause dominant X-linked juvenile and
476   adult-onset ALS and ALS/dementia. *Nature* 2011; **477:** 211-215.

477
478   24.   Synofzik M, Maetzler W, Grehl T *et al*: Screening in ALS and FTD patients reveals 3 novel
479   UBQLN2 mutations outside the PXX domain and a pure FTD phenotype. *Neurobiol Aging*
480   2012; **33:** 2949.e2913-2947.

481
482   25.   Gijselinck I, Van Mossevelde S, van der Zee J *et al*: Loss of TBK1 is a frequent cause of
483   frontotemporal dementia in a Belgian cohort. *Neurology* 2015; **85:** 2116-2125.

484
485   26.   Le Ber I, Camuzat A, Guerreiro R *et al*: SQSTM1 mutations in French patients with
486   frontotemporal dementia or frontotemporal dementia with amyotrophic lateral sclerosis.
487   *JAMA Neurol* 2013; **70:** 1403-1410.

488
489   27.   Williams KL, Topp S, Yang S *et al*: CCNF mutations in amyotrophic lateral sclerosis and
490   frontotemporal dementia. *Nat Commun* 2016; **7:** 11253.

491
492   28.   Cuyvers E, Sleegers K: Genetic variations underlying Alzheimer's disease: evidence from
493   genome-wide association studies and beyond. *Lancet Neurol* 2016; **15:** 857-868.

494
495   29.   Myllykangas L, Polvikoski T, Sulkava R *et al*: Genetic association of alpha2-macroglobulin
496   with Alzheimer's disease in a Finnish elderly population. *Ann Neurol* 1999; **46:** 382-390.

497
498   30.   MacDonald JR, Ziman R, Yuen RK, Feuk L, Scherer SW: The Database of Genomic Variants: a
499   curated collection of structural variation in the human genome. *Nucleic Acids Res* 2014; **42:**
500   D986-992.

501
502   31.   Wang K, Li M, Hakonarson H: ANNOVAR: functional annotation of genetic variants from
503   high-throughput sequencing data. *Nucleic Acids Res* 2010; **38:** e164.

504
505   32.   Kumar P, Henikoff S, Ng PC: Predicting the effects of coding non-synonymous variants on
506   protein function using the SIFT algorithm. *Nat Protoc* 2009; **4:** 1073-1081.

507
508   33.   Adzhubei IA, Schmidt S, Peshkin L *et al*: A method and server for predicting damaging
509   missense mutations. *Nat Methods* 2010; **7:** 248-249.

510
511   34.   Schwarz JM, Cooper DN, Schuelke M, Seelow D: MutationTaster2: mutation prediction for
512   the deep-sequencing age. *Nat Methods* 2014; **11:** 361-362.

513
514   35.   Reva B, Antipin Y, Sander C: Predicting the functional impact of protein mutations:
515   application to cancer genomics. *Nucleic Acids Res* 2011; **39:** e118.

516
517   36.   Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J: A general framework for
518   estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014; **46:** 310-
519   315.

520
521   37.   Guerreiro RJ, Beck J, Gibbs JR *et al*: Genetic variability in CLU and its association with
522   Alzheimer's disease. *PLoS One* 2010; **5:** e9510.

523
524   38.   Chi S, Jiang T, Tan L, Yu JT: Distinct neurological disorders with C9orf72 mutations: genetics,
525   pathogenesis, and therapy. *Neurosci Biobehav Rev* 2016; **66:** 127-142.

526
527   39.   Cacace R, Van Cauwenberghe C, Bettens K *et al*: C9orf72 G4C2 repeat expansions in
528   Alzheimer's disease and mild cognitive impairment. *Neurobiol Aging* 2013; **34:** 1712.e1711-
529   1717.

530

531   40.   Saint-Aubert L, Sagot C, Wallon D *et al*: A case of logopenic primary progressive aphasia with
532   C9ORF72 expansion and cortical florbetapir binding. *J Alzheimers Dis* 2014; **42:** 413-420.

533

534   41.   Harms M, Benitez BA, Cairns N *et al*: C9orf72 hexanucleotide repeat expansions in clinical
535   Alzheimer disease. *JAMA Neurol* 2013; **70:** 736-741.

536

537   42.   Majounie E, Abramzon Y, Renton AE *et al*: Repeat expansion in C9ORF72 in Alzheimer's
538   disease. *N Engl J Med* 2012; **366:** 283-284.

539

540   43.   Kohli MA, John-Williams K, Rajbhandary R *et al*: Repeat expansions in the C9ORF72 gene
541   contribute to Alzheimer's disease in Caucasians. *Neurobiol Aging* 2013; **34:** 1519.e1515-
542   1512.

543

544   44.   Majounie E, Renton AE, Mok K *et al*: Frequency of the C9orf72 hexanucleotide repeat
545   expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a
546   cross-sectional study. *Lancet Neurol* 2012; **11:** 323-330.

547

548   45.   Prihar G, Verkkoniem A, Perez-Tur J *et al*: Alzheimer disease PS-1 exon 9 deletion defined.
549   *Nat Med* 1999; **5:** 1090.

550

551   46.   Alzheimer's Disease Collaborative Group: The structure of the presenilin 1 (S182) gene and
552   identification of six novel mutations in early onset AD families. *Nat Genet* 1995; **11:** 219-222.

553

554   47.   Hiltunen M, Helisalmi S, Mannermaa A *et al*: Identification of a novel 4.6-kb genomic
555   deletion in presenilin-1 gene which results in exclusion of exon 9 in a Finnish early onset
556   Alzheimer's disease family: an Alu core sequence-stimulated recombination? *Eur J Hum
557   Genet* 2000; **8:** 259-266.

558

559   48.   Martikainen P, Pikkarainen M, Pöntynen K *et al*: Brain pathology in three subjects from the
560   same pedigree with presenilin-1 (PSEN1) P264L mutation. *Neuropathol Appl Neurobiol* 2010;
561   **36:** 41-54.

562

563   49.   Krüger J, Moilanen V, Majamaa K, Remes AM: Molecular genetic analysis of the APP, PSEN1,
564   and PSEN2 genes in Finnish patients with early-onset Alzheimer disease and frontotemporal
565   lobar degeneration. *Alzheimer Dis Assoc Disord* 2012; **26:** 272-276.

566

567   50.   Blom ES, Viswanathan J, Kilander L *et al*: Low prevalence of APP duplications in Swedish and
568   Finnish patients with early-onset Alzheimer's disease. *Eur J Hum Genet* 2008; **16:** 171-175.

569

570   51.   Krüger J, Kaivorinne AL, Udd B, Majamaa K, Remes AM: Low prevalence of progranulin
571   mutations in Finnish patients with frontotemporal lobar degeneration. *Eur J Neurol* 2009; **16:**
572   27-30.

573

574 52. Carrasquillo MM, Zou F, Pankratz VS *et al*: Genetic variation in PCDH11X is associated with
575   susceptibility to late-onset Alzheimer's disease. *Nat Genet* 2009; **41:** 192-198.

576

577 53. Beecham GW, Naj AC, Gilbert JR, Haines JL, Buxbaum JD, Pericak-Vance MA: PCDH11X
578   variation is not associated with late-onset Alzheimer disease susceptibility. *Psychiatr Genet*
579   2010; **20:** 321-324.

580

581 54. Lescai F, Pirazzini C, D'Agostino G *et al*: Failure to replicate an association of rs5984894 SNP
582   in the PCDH11X gene in a collection of 1,222 Alzheimer's disease affected patients. *J*
583   *Alzheimers Dis* 2010; **21:** 385-388.

584

585 55. Wu ZC, Yu JT, Wang ND *et al*: Lack of association between PCDH11X genetic variation and
586   late-onset Alzheimer's disease in a Han Chinese population. *Brain Res* 2010; **1357:** 152-156.

587

588 56. Miar A, Alvarez V, Corao AI *et al*: Lack of association between protocadherin 11-X/Y
589   (PCDH11X and PCDH11Y) polymorphisms and late onset Alzheimer's disease. *Brain Res* 2011;
590   **1383:** 252-256.

591

592 57. Jiao B, Liu X, Zhou L *et al*: Polygenic Analysis of Late-Onset Alzheimer's Disease from
593   Mainland China. *PLoS One* 2015; **10:** e0144898.

594

595 58. Bettens K, Brouwers N, Engelborghs S *et al*: Both common variations and rare non-
596   synonymous substitutions and small insertion/deletions in CLU are associated with increased
597   Alzheimer risk. *Mol Neurodegener* 2012; **7:** 3.

598

599 59. Bettens K, Vermeulen S, Van Cauwenberghe C *et al*: Reduced secreted clusterin as a
600   mechanism for Alzheimer-associated CLU mutations. *Mol Neurodegener* 2015; **10:** 30.

601

602 60. Yang CB, Zheng YT, Li GY, Mower GD: Identification of Munc13-3 as a candidate gene for
603   critical-period neuroplasticity in visual cortex. *J Neurosci* 2002; **22:** 8614-8618.

604

605 61. Yang CB, Kiser PJ, Zheng YT, Varoqueaux F, Mower GD: Bidirectional regulation of Munc13-3
606   protein expression by age and dark rearing during the critical period in mouse visual cortex.
607   *Neuroscience* 2007; **150:** 603-608.

608

609 62. Miller JA, Woltjer RL, Goodenbour JM, Horvath S, Geschwind DH: Genes and pathways
610   underlying regional and cell type changes in Alzheimer's disease. *Genome Med* 2013; **5:** 48.

611

612 63. Chaugule VK, Walden H: Specificity and disease in the ubiquitin system. *Biochem Soc Trans*
613   2016; **44:** 212-227.

614

615   64.   Bartee E, Mansouri M, Hovey Nerenberg BT, Gouveia K, Früh K: Downregulation of major
616   histocompatibility complex class I by human ubiquitin ligases related to viral immune
617   evasion proteins. *J Virol* 2004; **78:** 1109-1120.

618
619   65.   Korvatska O, Leverenz JB, Jayadev S *et al*: R47H Variant of TREM2 Associated With Alzheimer
620   Disease in a Large Late-Onset Family: Clinical, Genetic, and Neuropathological Study. *JAMA*
621   *Neurol* 2015; **72:** 920-927.

622
623   66.   Kohli MA, Cukier HN, Hamilton-Nelson KL *et al*: Segregation of a rare TTC3 variant in an
624   extended family with late-onset Alzheimer disease. *Neurol Genet* 2016; **2:** e41.

625

626

627

Legends to the figures

Figure 1. Schematic presentation of the study describing the workflow of genetic examinations. WES = whole exome sequencing.

Figure 2. Proportions of *C9orf72* expansions in clinical frontotemporal dementia (FTD), Alzheimer's disease (AD) and unspecified dementia in a cohort of 60 families. DLB = dementia with Lewy bodies.

Figure 3. Pedigrees of the families with rare variants verified by Sanger sequencing. DNA samples were available from individuals marked with an asterisk. *APOE* genotypes are also marked in the pedigree.

a. Family Fam-56 with the *CLU* p.(Thr203Ile) variant. Heterozygous variant (-/+), homozygous wild-type allele (-/-).

b. Family Fam-15 with the *PCDH11X* p.(Asp760Val) variant. Heterozygous variant (-/+), homozygous wild-type allele (-/-), hemizygous variant (+), hemizygous wild-type allele (-).

c. Family Fam-49 with the *UNC13C* p.(443del) variant. Heterozygous variant (-/+), homozygous wild-type allele (-/-).

d. Family Fam-13 with the *MARCH4* p.(Lys211Glu) variant. Heterozygous variant (-/+), homozygous wild-type allele (-/-).

e. Family Fam-59 with the *MARCH4* p.(Trp13Cys) variant. Heterozygous variant (-/+), homozygous wild-type allele (-/-).