Mutation Analysis of Sporadic Early-Onset Alzheimer's Disease using the NeuroX Array

Imelda S. Barber[a], Anne Braae[a], Naomi Clement[a], Tulsi Patel[a], Tamar Guetta-Baranes[a], Keeley Brookes[a], Christopher Medway[a], Sally Chappell[a], Rita Guerreiro[b], Jose Bras[b], Dena Hernandez[c], Andrew Singleton[c], John Hardy[b], David M. Mann[d], ARUK Consortium[e] and Kevin Morgan[a]

a. School of Life Sciences, Queens Medical Centre, University of Nottingham, Nottingham, NG7 2UH, UK
b. Department of Molecular Neuroscience, Institute of Neurology, University College London, Queen Square, London, WC1N 1PJ, UK
c. Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA
d. Institute of Brain, Behaviour and Mental Health, Faculty of Medical and Human Sciences, Oxford Road, University of Manchester, Manchester, M13 9PT, UK
e. Alzheimer's Research United Kingdom (ARUK) Consortium

Details of ARUK consortium:
Peter Passmore[f], David Craig[f], Janet Johnston[f], Bernadette McGuinness[f], Stephen Todd[f], Reinhard Heun[g], Heike Kölsch[h], Patrick G. Kehoe[i], Emma R.L.C. Vardy[j], Nigel M. Hooper[d], Stuart Pickering-Brown[d], Julie Snowden[k,d], Anna Richardson[k,d], Matthew Jones[k,d], David Neary[k,d], Jennifer Harris[k,d], James Lowe[a], A. David Smith[l], Gordon Wilcock[l], Donald Warden[l] & Clive Holmes[m]

f. Centre for Public Health, School of Medicine, Queen's University, Belfast, BT9 7BL, UK
g. Royal Derby Hospital, Derby, DE22 3WQ, UK
h. Department of Psychiatry, University of Bonn, Sigmund-Freud-Strasse 25, Bonn 53105, Germany
i. School of Clinical Sciences, John James Laboratories, University of Bristol, Bristol, BS16 1LE, UK
j. Ageing and Complex Medicine, Salford Royal NHS Foundation Trust, Stott Lane, Salford, M6 8HD, UK
k. Cerebral Function Unit, Greater Manchester Neurosciences Centre, Salford Royal NHS Foundation Trust, Stott Lane, Salford, M6 8HD, UK
l. OPTIMA, Nuffield Department of Clinical Neurosciences, Level 6, West Wing, John Radcliffe Hospital, University of Oxford, Oxford, OX3 9DU, UK
m. Clinical and Experimental Science, University of Southampton, Southampton, SO17 1BJ, UK

Corresponding Author

Ms Imelda Barber
Department of Psychiatry
Warneford Hospital
Oxford
OX3 7JX
United Kingdom
Tel: +44 1865 613175
E-mail: imelda.barber@psych.ox.ac.uk

Abstract

We have screened sporadic early-onset Alzheimer's disease (sEOAD, n=408) samples using the NeuroX array for known causative and predicted pathogenic variants in 16 genes linked to familial forms of neurodegeneration. We found two sEOAD individuals harbouring a known causative variant in *PARK2* known to cause early-onset Parkinson's disease (EOPD); p.T240M (n=1) and p.Q34fs delAG (n=1). Additionally, we identified three sEOAD individuals harbouring a predicted pathogenic variant in *MAPT* (p.A469T) which has previously been associated with AD. It is currently unknown if these variants affect susceptibility to sEOAD, further studies would be needed to establish this. This work highlights the need to screen sEOAD individuals for variants that are more classically attributed to other forms of neurodegeneration.

Keywords

Alzheimer's disease, Parkinson's disease, sporadic, early-onset, NeuroX, screening

1. Introduction

Alzheimer's disease (AD) is the commonest form of dementia in the world. AD and other dementias were the fourth leading cause of death in high-income countries in 2012 (WHO, 2012). Sporadic early-onset Alzheimer's disease (sEOAD) has a disease onset ≤ 65 years of age and these individuals do not harbour a known causative mutation, the remaining sporadic cases are classified as late-onset Alzheimer's disease (LOAD). Both forms of AD have a complex aetiology with heritability estimated to be 92-100% for sEOAD (Wingo, et al., 2012) and 70% for LOAD (Gatz, et al., 2006). Given the difference in heritability and age of onset between sEOAD and LOAD, it is likely that sEOAD patients have a more penetrant genetic aetiology and thus provide a good cohort to explore the genetics of AD.

Many types of dementia have a neuropathology and/or clinical crossover, for example Parkinson's disease with/without dementia (PDD/PD) have alpha-synuclein deposits in the brain and a similar clinical presentation to dementia with Lewy bodies (DLB) (Jellinger, 2014). Mixed dementia has features linked to more than one type of dementia, such as AD with cerebrovascular lesions and/or Lewy bodies (Jellinger, 2014). It is not surprising then that some genetic loci identified thus far are associated with multiple types of dementias; the commonest example is that of *APOE* ε4 which is associated with AD (Corder, et al., 1993), posterior cortical atrophy (PCA) (Carrasquillo, et al., 2014) and DLB (Bras, et al., 2014). It is believed that sporadic AD could be in part due to the aggregate of multiple causative variants, therefore it is easy to imagine that different types of dementia have overlapping genetics; whereby a portion of variants that contribute to one dementia are also seen to contribute to a different type of dementia. Alternatively different types of dementia could be a result of pleiotropy, for example it has recently been reported that the alpha-synuclein gene (*SNCA*) is associated with DLB but this is a different haplotype to that associated with PDD (Bras, et al., 2014).

The NeuroX is a customised Illumina HumanExome DNA microarray; the first version of the chip contains 267,607 markers, most of which genotype rare missense variants, notably most of the genes in the human genome have at least one variant genotyped. The NeuroX includes standard content (242,901) designed by Illumina together with custom content (24,706) designed to be 'neuro-specific'. The custom content was selected to genotype specific variants and genes linked to neurologic diseases, the inclusion criterion was determined using literature searches and genotyping data available before 2014. A more descriptive explanation of the NeuroX can be found in the consortia's published paper (Nalls, et al., 2015). The NeuroX provides a convenient approach to screen for causative mutations and test for genetic crossover and/or pleiotropy amongst neurologic diseases.

We report the screening of 408 sEOAD individuals with the aim to identify causative or predicted pathogenic variants in 16 selected genes using the NeuroX. These 16 genes are linked to familial forms of neurodegeneration including Alzheimer's disease (*APP*, *PSEN1* and *PSEN2*), frontotemporal dementia/amyotrophic lateral sclerosis (*C9orf72*, *CHMP2B*, *FUS*, *GRN*, *MAPT*, *TARDBP* and *VCP*), Parkinson's disease (*LRRK2*, *PARK2*, *PARK7*, *PINK1* and *SNCA*) and prion disease (*PRNP*); all have pathogenic variants highlighted in freely accessible online databases.

2. Materials and Methods

Methods were conducted according to the manufacturer's instructions unless otherwise stated. All sporadic early-onset Alzheimer's disease (sEOAD) samples were screened for known causative variants in *APP* exons 16 and 17 by Sanger sequencing and in *PSEN1*/*PSEN2* using the NeuroX data, individuals harbouring a causative variant in either of these genes were removed prior to this analysis (Barber, et al., 2016).

2.1. Samples

sEOAD individuals (n=408) had an age of disease onset ≤ 65 years of age (**Table 1**). For 28 individuals where AAO was not documented, it was derived assuming 8 years disease duration from age at death (Ryman, et al., 2014).

	N	Mean AAO (±SD)		Females (%)		*APOE* ε4+ (%)		*APOE* ε4 MAF	*APOE* ε4ε4 (%)		Definite	Probable
Bristol	21	53.3	(5.3)	9	(42.9)	10	(47.6)	0.31	3	(14.3)	21	0
Manchester	328	57.1	(5.5)	156	(47.6)	196	(59.8)	0.58	46	(14.0)	53	275
Nottingham	26	58.2	(6.3)	12	(46.2)	11	(42.3)	0.23	1	(3.8)	5	21
Oxford	33	55.6	(4.2)	19	(57.6)	19	(57.6)	0.33	3	(9.1)	24	9
All	408	56.8	(5.5)	196	(48.0)	236	(57.8)	0.53	53	(13.0)	103	305

Table 1

Demographics of sporadic early-onset Alzheimer's disease (sEOAD) cohort

The cohort contains individuals from multiple centres; each centre is represented one per row. The number of individuals (N) from each centre is given along with the mean age of onset with standard deviation (Mean AAO (±SD)), this is followed by the number and percentage of female individuals per centre (Females (%)), the number and percentage of individuals harbouring at least one *APOE* ε4 allele (APOE ε4+ (%)), *APOE* ε4 minor allele frequency (*APOE* ε4 MAF) and number and percentage of individuals with *APOE* ε4ε4 genotype (*APOE* ε4ε4). The final columns state the number of individuals classed as either post mortem verified (Definite) or probable (Probable) Alzheimer's disease according to NINCDS-ADRDA and CERAD guidelines. Key: N, number of individuals; SD, standard deviation.

sEOAD individuals were diagnosed as either definite or probable Alzheimer's disease (AD) according to National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's disease and Related Disorders Association (NINCDS-ADRDA), and the Consortium to Establish a Registry for Alzheimer's disease (CERAD) guidelines. All samples used in this study were received with informed consent and experimental procedures were approved by the local ethics committee, Nottingham Research Ethics Committee 2 (REC reference 04/Q2404/130). All experimental procedures were conducted in accordance with approved guidelines.

DNA was extracted from blood or brain tissue using a standard phenol chloroform extraction method. DNA quality and quantity was assessed by gel electrophoresis and NanoDrop™ 3300 spectrometry respectively.

2.2. Quality control of NeuroX data

Quality control (QC) of the NeuroX intensity data was conducted in Genome Studio version 2011.1 using the genotyping module version 1.9.4 with 'exclude female Y-SNPs from SNP Statistics' checked to ensure variants on the Y chromosome aren't incorrectly labelled as having a low call rate. Standard content was clustered using the CHARGE cluster file version 1.0 (Grove, et al., 2013). All quality control of the markers in Genome Studio were conducted using only the best quality samples (≥ 99% call rate). Standard content with call rate < 100% and all custom content were clustered using Genome Studio's clustering algorithm, this was followed by manual assessment and clustering for all non-autosomal markers, and manual assessment and clustering of autosomal markers matching any of the following criteria: ≤ 99% call rate, excess heterozygote calls, excess heterozygote calls relative to expected HWE, deficient heterozygote calls relative to expected HWE, low intensity, unexpected cluster positions, wide clusters or low cluster separation (Grove, et al., 2013). Once quality control had been conducted in Genome Studio the genotyping calls were exported from Genome Studio to PLINK format in the forward orientation with all samples and all variants included. The MapInfo (location) was updated for 29 markers that originally had no location

(**Supplementary Table 1**), and the chromosome was updated for 121 markers from chromosome X to the pseudo autosomal region (**Supplementary Table 2**). Additional quality control was performed in PLINK v1.07. Samples were first removed if they had a call rate < 98% followed by markers that had a call rate < 95%. Further samples were removed if they failed the following criteria: samples with identity by decent > 18.75% or heterozygosity rate outside ± 3 SD of the mean, both determined using an LD pruned version of the dataset (indep-pairwise 50 5 0.2) with only common autosomal variants (MAF > 0.1). Further common markers were removed if they had significant deviation (p.value < 1.2E-6) from Hardy-Weinberg equilibrium in control samples (data not shown). The final NeuroX dataset had 265,049 markers with an average sample call rate of 99.9%.

2.3. Samples harbouring a known causative variant

A total of 1196 variations are documented in the PD online mutation database (PDMDB) [http://www.molgen.vib-ua.be/PDMutDB/ accessed August 2013], AD&FTD online mutation database (AD&FTDMDB) [http://www.molgen.vib-ua.be/ADMutations/ accessed August 2013] (Cruts, et al., 2012) and Human Prion Mutation Database [http://www.mad-cow.org/prion_point_mutations.html accessed November 2014] combined. These databases document variants across 16 genes known to cause familial forms of neurodegeneration (*APP, C9orf72, CHMP2B, FUS, GRN, LRRK2, MAPT, PARK2, PARK7, PINK1, PRNP, PSEN1, PSEN2, SNCA, TARDBP* and *VCP*), and includes three genes that are linked to Alzheimer's disease (*APP, PSEN1* and *PSEN2*). The databases include variants that don't cause disease (not pathogenic), have unknown pathogenicity (pathogenic nature unclear) and are known to cause disease (known causative) (**Supplementary Table 3**). 1075 of these database variants are SNPs, small insertions or small deletions and therefore had the potential to be genotyped on the NeuroX.

The genomic position (relative to the reference build GRCh37, 27 Feb 2009), reference allele and alternative allele were successfully calculated for 265,828 markers (99%) on the NeuroX using an in-house script. A second in-house script was used to establish if any of these markers genotyped the

1075 database variants. It was established that 412 (38%) of these variants were genotyped on the NeuroX and included 38% of AD related variants, 32% of FTD/ALS related variants, 47% of PD related variants and 12% of prion related variants. Of the 412 variants, 407 passed quality control procedures and inspection of the cluster plots found they all clustered well.

The 407 database variants were filtered to retain only those labelled as causative and were also polymorphic in our sEOAD cohort, four variants fit this criterion. The sEOAD samples harbouring these four variants were identified and their genotype verified with Sanger sequencing. As expected these four variants were not located in *APP, PSEN1* or *PSEN2* as our sporadic cohort had been previously filtered to remove individuals harbouring these variants.

2.4. Samples harbouring a predicted pathogenic variant

It was established that 662 variants were genotyped on the NeuroX which passed QC and were located in one of the 16 genes linked to neurodegenerative diseases (**Table 2**), this list excluded the 407 known causative variants analysed previously. The gene ID, European minor allele frequency (MAF) from the 1000 Genomes Project (Genomes Project, et al., 2012), Polyphen score and SIFT score were retrieved for all 662 variants using ENSEMBL's variant effect predictor (VEP) (McLaren, et al., 2010) which was installed locally and run over the command line using reference genome build GRCh37, Polyphen HDIV database and predicting one consequence per variant.

Variants were filtered to retain only those that were polymorphic in our sEOAD cohort, had a 1000 Genomes European MAF < 0.01, and were predicted to be probably pathogenic; which is defined here as a 'deleterious effect' by SIFT (≤0.05) or 'probably damaging' by PolyPhen (≥0.909). Three variants fit this criterion. The sEOAD samples harbouring these three variants were identified and their genotype verified with Sanger sequencing.

2.5. Sanger sequencing

Genomic DNA was amplified in a final volume of 15 µl using the following constituents and final concentrations: 2 ng/µl gDNA, 1 pM forward primer, and 1 pM reverse primer, 1x Buffer (BioLabs), 0.2 mM dNTPs (Thermo Scientific), 0.1 U/µl LongAmp Taq DNA polymerase (New England Biolabs) and molecular grade water to the required volume. The reaction was subjected to the following thermal conditions: initial denaturation step of 94°C for 2 min, followed by 35 cycles of 94°C for 30 sec, 57°C for 15 sec and 72°C for 45 sec, finished with a final extension step at 72°C for 7 min. A reaction containing no gDNA was included as a negative control. The PCR products were cleaned using ExoSAP-IT (Affymetrix). Primers used for sequencing were the same as those used for amplification (Eurogenomics). The primers used for p.T240M were forward 5'GCTCGTGTGGCAGAACAATA 3' and reverse 5'ACACCCCACCTCTGACAAG 3'; the product was 202 bp (base pairs) long. The primers used for p.G255V were forward 5' TGGCAATCAAGACCAGAGTG 3' and reverse 5' GATTCATGCAATCCTCCACA 3'; the product was 243 bp long. The primers used for p.Q34fs were forward 5' TCAGGCATGAATGTCAGATTG 3' and reverse 5' CCTTCCAATTTCCTTGGTCA 3'; the product was 272 bp long. The primers used for p.G107V were forward 5'AGAGCTGAGGCACCTTGGTA 3' and reverse 5' ATGGCAGGCAATTCCAGTT 3'; the product was 235 bp long. The primers used for p.S427F were forward 5' TCCACACGTTCCTCTGCTAA 3' and reverse 5' AGCAGCCTGGTTCTTTTCAA 3'; the product was 230 bp long. The primers used for p.A469T were forward 5'

GGCTGGTGTTGACTCTTGGT 3’ and reverse 5’ TCTTACCAGAGCTGGGTGGT 3’; the product was 206 bp long.

Amplicons were sequenced using the Sanger di-deoxy method in the forward and reverse directions. Cleaned PCR products were sequenced in a final volume of 10 µl using 4 µl PCR product and the following constituents: 0.5 pM forward/reverse primer, 1x BigDye Sequencing Buffer (Life Technologies), 0.25x BigDye Terminator v3.1 (Life Technologies) and molecular grade water to the required volume. The reaction was subjected to the following thermal conditions: 25 cycles of 96°C for 30 sec, 50°C for 15 sec and finally 60°C for 4 min. The reactions were cleaned using Performa DTR Gel filtration Cartridges (Edge Biosystems). The eluent was dried and sequencing was performed on an ABI 3130 automated sequencer.

3. Results

sEOAD samples (n=408) were genotyped on the NeuroX DNA microarray. These samples were screened for variants known to cause disease (known causative variants) of 13 genes linked to familial forms of neurodegeneration (non-AD familial genes) and variants predicted to be pathogenic (predicted pathogenic variants) in 16 genes linked to familial forms of neurodegeneration (familial genes).

3.1. Validated genotypes

The NeuroX identified three sEOAD samples harbouring a known causative variant in two of the 13 non-AD familial genes and ten sEOAD samples harbouring a predicted pathogenic variant in one of the 16 familial genes. All samples were clinically diagnosed as having probable AD. Sanger sequencing confirmed that nine samples harboured a variant. **Figure 1** shows the sequence chromatograms for the thirteen samples and **Table 2** lists their genotypes.

A	Marker	Genomic Position	Variant	Base Change	MAF	Protein Change	Gene	Disease	Sample	Seq	Gender	AAO	*APOE* ε
	exm593516	6:162394349	rs137853054	aCg/aTg	2.8e-04	p.T240M	*PARK2*	PD (R)	M177 (Het)	C	Female	52	44
	NeuroX_PARK2_Gln34fs_del_AG	6:162864411-2	rs55777503	cga.cAG.ggg/cga.cgg.ggt	2.9e-04	p.Q34fs	*PARK2*	PD (R)	M099 (Het)	C	Female	56	44
	NeuroX_PARK2_Gln34fs_del_A	6:162864412	rs748142049	cga.cAg.ggg/cga.cgg.ggg	8.2e-06	p.Q34fs	*PARK2*	PD (R)	M099 (Het)	U	Female	56	44
	NeuroX_16:31196410	16:31196410	-	gGc/gTc	U/K	p.G225V	*FUS*	ALS (R)	M215 (Het)	U	Female	53	34

B	Marker	Genomic Position	rsID	Base Change		Protein Change	Gene	Disease	Individual	Seq	Gender	AAO	*APOE* ε
	exm1330962	17:44055753	rs144397565	gGc/gTc	2.4e-05	p.G107V	*MAPT*	FTD (D)	M820 (Het)	C	Male	65	34
	exm1331018	17:44067341	rs143956882	tCc/tTc	0.001	p.S427F	*MAPT*	FTD (D)	M357 (Het)	C	Male	53	33
									M143 (Het)	C	Female	57	34
									M697 (Het)	U	Female	48	24
									M245 (Het)	C	Female	63	44
	exm1331027	17:44068850	rs143624519	Gcc/Acc	0.001	p.A469T	*MAPT*	FTD (D)	M168 (Het)	C	Female	58	33
									M172 (Het)	U	Female	58	34
									M382 (Het)	U	Female	52	33
									M691 (Het)	C	Male	55	34
									M699 (Het)	C	Male	58	33

Table 2

Samples harbouring a variant in a familial gene

Samples identified by the NeuroX as harbouring a known causative variant in one of the 13 none AD familial genes (A) or a predicted pathogenic variant in one of the 16 familial genes. The name of the NeuroX marker (Marker) is followed by information about the variant genotyped by the marker, including the genomic position of the variant in reference genome build GRCh37 and format chromosome:base position (Genomic Position), and the reference ID of the variant according to dbSNP (Variant). The variant is given at nucleotide level (Base Change) on the sense strand, the minor allele frequency in the general population according to ExAC (MAF) and amino acid level (Protein Change). Also listed is the gene the variant resides in (Gene) and the disease most associated with the gene, the Mendelian pattern of inheritance is given in brackets (Disease). The sample(s) harbouring the variant (Sample) is followed by the results of Sanger sequencing (Seq) and patient information including gender (Gender), age at onset (AAO) and *APOE* ε status (*APOE* ε). Key: rsID, reference single nucleotide polymorphism identification; U/K, unknown; FTD, frontotemporal dementia; D, dominant; R, recessive; C, Sanger sequencing confirmed the genotype of this sample; U, Sanger sequencing did not confirm the genotype of this sample.

Sanger sequencing confirmed that eight samples harboured the minor allele indicated by the NeuroX and one sample (M820) harboured an alternative minor allele. M820 appeared heterozygous for G>A base change (rs144397565, p.G170D), however sequencing confirmed it was heterozygous for G>T base change (rs144397565, p.G170V). Evidently this position is trimorphic, as this marker uses the Infinium II probe design it was able to detect both minor alleles but was unable to differentiate between them.

3.2. Invalidated genotypes

Of the variants not confirmed by Sanger sequencing, one sample (M099) was identified as heterozygous for p.Q34fs (delA) in addition to the confirmed variant p.Q34fs (delAG) at the same position. Notably the probe sequence for both markers was identical. As both deletions are followed by the same nucleotide (G), the single nucleotide extension design meant this probe was capable of detecting both variants, in this instance both probes detected the delAG variant.

Four samples (M215, M697, M172, M382) failed to verify with initial Sanger sequencing; the cluster plots were examined to establish call quality, which were found to be good (**Figure 2**). Three of the four samples (M215, M697 and M382) were re-sequenced from the original DNA stock; all three gave the same result as originally obtained (unconfirmed), thereby obviating a sample 'mix-up'. We were unable to re-sequence M172 as no additional DNA was available. It is interesting to note that variant p.G225V is located within a GGC repeat region and this could be the reason for the discrepancy between the NeuroX genotype and sequencing result. Alternatively, the NeuroX genotypes could be correct for these samples and the PCR reaction may have resulted in allele dropout (Blais, et al., 2015), however it would be advisable to use a different primer pair as this might permit detection of the other allele.

In addition to the above, twenty-five samples were identified as harbouring the variant p.Q130fs (rs63750768, gaT.AGT/ga) in the *GRN* gene. Similar to the situation with M820 (as described in **Section 3.1**), Ghani and colleagues confirmed that this marker also genotypes a common alternative

minor allele at the same position (rs25646, gaT/gaC) (Ghani, et al., 2015). Consequently, this variant was discarded from further investigation.

4. Discussion

4.1. Known causative variants

Two sEOAD samples were confirmed heterozygous for a known causative variants; p.T240M (n=1) and p.Q34fs (delAG, n=1). Both of these variants are located in the gene *PARK2*. Known causative variants in *PARK2* include point mutations and exon rearrangements/deletions/duplications, often as homozygote or compound heterozygotes in early-onset PD (EOPD).

Variant p.T240M has been seen in EOPD as a compound heterozygote with various exon deletions or duplications (Amboni, et al., 2009,Deng, et al., 2006,Periquet, et al., 2003,Sironi, et al., 2008), it has also been seen as a homozygote (Madegowda, et al., 2005) and a heterozygote (Camargos, et al., 2009). A compound heterozygote with an exon deletion has also been seen in one healthy individual (Deng, et al., 2006). Variant p.Q34fs (delAG) has been seen in EOPD as a compound heterozygote with exon deletions or SNPs (Abbas, et al., 1999,Guo, et al., 2008,Guo, et al., 2010,Hedrich, et al., 2002,Illarioshkin, et al., 2003,Koziorowski, et al., 2010,Lesage, et al., 2008,Lohmann, et al., 2009,Scherfler, et al., 2004), it has also been seen as a homozygote (Koziorowski, et al., 2010,Scherfler, et al., 2004) and a heterozygote (Brooks, et al., 2009,Bruggemann, et al., 2009). A compound heterozygote with an exon deletion has also been seen in LOPD (Lesage, et al., 2008).

Previous findings would suggest that p.T240M and p.Q34fs (delAG) elicit risk for PD, in particular EOPD. Finding these variants in our sEOAD cohort would suggest that they could also elicit risk to sEOAD; however we only found each of them in one sample (0.25% of our sEOAD cohort) as a heterozygote and we don't know if these individuals were compound heterozygotes. These variants could elicit risk to sEOAD, however a large case-control association study would be needed to establish this.

Both of these patients (M117 and M099) had an *APOE* ε4ε4 status and there was nothing unusual about their presentation or progress, which suggests that a misdiagnosis is unlikely. However, M099 had a mother who was said to have had motor neuron disease (MND), so there was likely to have been physical signs in her, possibly consistent with a known causative variant in *PARK2*.

4.2. Predicted pathogenic variants

Seven sEOAD samples were confirmed heterozygous for a predicted pathogenic variant; p.G107V (n=1), p.S427F (n=3) or p.A469T (n=2). All these variants are located in the gene *MAPT* and are named in reference to the longest tau transcript (tau-g). The majority of known causative variants in *MAPT* are SNPs with autosomal dominant inheritance and result in FTD.

Variant p.G107V was predicted to be probably damaging (1.00) by Polyphen and deleterious (0) by SIFT, it is located in exon 4 of *MAPT* where no known causative variants have been documented. Variant p.S427F was predicted to be deleterious (0.02) by SIFT and probably damaging (0.99) by Polyphen, it is located in exon 4a which is spliced out in the transcript htau40 and thus not present in the human brain (Liu and Gong, 2008,Pittman, et al., 2006). It is unlikely this variant has a functional effect in the brain if the transcripts containing this variant are not present.

Variant p.A469T is also called p.A152T in htau40. It was predicted to be deleterious (0.05) by SIFT and benign (0.30) by Polyphen. This variant significantly increases the risk for both FTD (p.value=0.0005, OR=3.0 (CI: 1.6–5.6)) and AD (p.value=0.004, OR=2.3 (CI: 1.3–4.2)) when compared to controls (Coppola, et al., 2012). *In vitro* site-directed mutagenesis of human tau cDNA showed the variant resulted in less efficient binding to microtubules and a pronounced increase in the formation of tau oligomers (Coppola, et al., 2012). Furthermore, Isogenic human iPSCs generated from fibroblasts saw the variant result in axonal degeneration and cell death (Fong, et al., 2013). Whether heterozygous p.A469T in humans would cause the same effect is unknown. Notably, this variant is located in exon 7, the downstream residue (p.T153) is part of a Threonine-Proline motif that is phosphorylated during the cell cycle (Illenberger, et al., 1998) and the upstream residue (p.I151) is

seen to interact with microtubules using nuclear magnetic resonance (Mukrasch, et al., 2009). Variant p.A469T could affect the functioning of the upstream or downstream residue which might explain the experimental observations for this variant. This variant could elicit risk to sEOAD, however a large case-control association study would be needed to establish this.

4.3. Considerations

Although DNA microarrays are cost efficient, they have several drawbacks as evident from this study. The results of Sanger sequencing found that a high proportion of variants failed to verify, this emphasises the need to verify all genotypes called by DNA microarray technologies. The version of the NeuroX used in this study was only able to successfully genotype 182 of the 523 (35%) known causative variants documented in three online databases (**Supplementary Table 3**), since that time the number of variants documented in the online databases has increased by 147, albeit not all of these will be important (causative), but highlights the additional issue of having to redesign chips as new personal variants come to light, which questions the cost effectiveness of DNA microarray technology for extremely rare variants. Next generation sequencing (NGS) technology overcomes the limitations and drawbacks of DNA microarray technology, and the once debilitating cost of NGS has almost dissipated as the price has plummeted in recent years. There is no doubt that studies like this should be conducted, however given the drawbacks of DNA microarray technology they would be better conducted using NGS technologies.

This study has made use of an *in-silico* approach to classify the pathogenicity of variants; however *in-silico* predictions alone are insufficient to properly appraise a variant. Richards *et al* have developed an approach that can help define the definition of 'pathogenic' in the clinical and research setting with regard to Mendelian disorders (Richards, et al., 2015), this approach makes use of additional types of data including population, functional and segregation data. As our understanding of complex diseases increases no doubt an approach incorporating several lines of data will be used to define pathogenic variants for non-Mendelian disorders such as sEOAD and LOAD.

5. Conclusion

We have screened sporadic early-onset Alzheimer disease (sEOAD) individuals for known causative and predicted pathogenic variants in 16 genes linked to neurodegenerative diseases. We have identified nine sEOAD samples harbouring a known causative variant or a predicted pathogenic variant in *PARK2* and *MAPT*. These variants could elicit risk to sEOAD in addition to PD and FTD, however further studies would be needed to establish this. This work highlights the need to screen sEOAD individuals for variants that are more classically attributed to other forms of neurodegeneration as there could be a degree of genetic overlap.

Acknowledgements

The University of Nottingham lab is funded by Alzheimer's Research UK and the Big Lottery Fund. IB's PhD studentship is jointly funded by Alzheimer's Research UK and the School of Life Sciences at The University of Nottingham. JB and RG's Fellowships are funded by The Alzheimer's Society.

Figure Legends

Figure 1
Sequence chromatograms
Sequence chromatograms of all 13 individuals thought to harbour a pathogenic (A) or predicted pathogenic variant (B). Each individual was sequenced in forward (top row) and reverse (bottom row) orientations. Note that the forward orientation does not always correspond to the sense strand. The images were taken from Sequence Scanner (Applied Biosystems) with the variant at the centre and surrounded by two or three bases either side. Four individuals (M215, M697, M172 and M382) show wild-type sequence according to the chromatogram.

Figure 2
SNP cluster plot for genotypes unconfirmed with Sanger sequencing
SNP cluster graphs generated by GenomeStudio for markers NeuroX_16:31196410 (p.G225V located in *FUS*) (A), exm1331018 (p.S427F located in *MAPT*) (B) and exm1331027 (p.A469T located in *MAPT*) (C). Each coloured circle represents one individual, those coloured red are homozygous mutant (TT), those coloured purple are heterozygous (GT), those coloured blue are homozygous wildtype (GG), and finally those coloured black are not called. The plots show all ten samples called as heterozygotes cluster well and does not explain why four samples failed to verify with Sanger sequencing (M215, M697, M172 and M382).

References

Abbas, N., Lucking, C.B., Ricard, S., Durr, A., Bonifati, V., De Michele, G., Bouley, S., Vaughan, J.R., Gasser, T., Marconi, R., Broussolle, E., Brefel-Courbon, C., Harhangi, B.S., Oostra, A.B., Fabrizio, E., Bohme, G.A., Pradier, L., Wood, N.W., Filla, A., Meco, G., Denefle, P., Agid, Y., Brice, A., French Parkinson's Dis Genetics Study, G., European Consortium Genetic Susceptibility, P. 1999. A wide variety of mutations in the parkin gene are responsible for autosomal recessive parkinsonism in Europe. Hum Mol Genet 8(4), 567-74. doi:10.1093/hmg/8.4.567.

Amboni, M., Pellecchia, M.T., Cozzolino, A., Pieillo, M., Vitale, C., Barone, P., Varrone, A., Garavaglia, B., Gambelli, S., Federico, A. 2009. Cerebellar and Pyramidal Dysfunctions, Palpebral Ptosis and Weakness As Presenting Symptoms of PARK-2. Movement Disorders 24(2), 303-5. doi:10.1002/mds.22342.

Barber, I.S., Garcia-Cardenas, J.M., Sakdapanichkul, C., Deacon, C., Erazo, G.Z., Guerreiro, R., Bras, J., Hernandez, D., Singleton, A., Guetta-Baranes, T., Braae, A., Clement, N., Patel, T., Brookes, K., Medway, C., Chappell, S., Mann, D.M., Morgan, K., Consortium, A. 2016. Screening exons 16 and 17 of the amyloid precursor protein gene in sporadic early-onset Alzheimer's disease. Neurobiol Aging 39, E1-E.

Blais, J., Lavoie, S.B., Giroux, S., Bussieres, J., Lindsay, C., Dionne, J., Laroche, M., Giguere, Y., Rousseau, F. 2015. Risk of Misdiagnosis Due to Allele Dropout and False-Positive PCR Artifacts in Molecular Diagnostics: Analysis of 30,769 Genotypes. J Mol Diagn 17(5), 505-14. doi:10.1016/j.jmoldx.2015.04.004.

Bras, J., Guerreiro, R., Darwent, L., Parkkinen, L., Ansorge, O., Escott-Price, V., Hernandez, D.G., Nalls, M.A., Clark, L.N., Honig, L.S., Marder, K., Van Der Flier, W.M., Lemstra, A., Scheltens, P., Rogaeva, E., St George-Hyslop, P., Londos, E., Zetterberg, H., Ortega-Cubero, S., Pastor, P., Ferman, T.J., Graff-Radford, N.R., Ross, O.A., Barber, I., Braae, A., Brown, K., Morgan, K., Maetzler, W., Berg, D., Troakes, C., Al-Sarraj, S., Lashley, T., Compta, Y., Revesz, T., Lees, A., Cairns, N., Halliday, G.M., Mann, D., Pickering-Brown, S., Dickson, D.W., Singleton, A., Hardy, J. 2014. Genetic analysis implicates APOE, SNCA and suggests lysosomal dysfunction in the etiology of dementia with Lewy bodies. Hum Mol Genet 23(23), 6139-46. doi:10.1093/hmg/ddu334.

Brooks, J., Ding, J., Simon-Sanchez, J., Paisan-Ruiz, C., Singleton, A.B., Scholz, S.W. 2009. Parkin and PINK1 mutations in early-onset Parkinson's disease: comprehensive screening in publicly available cases and control. J Med Genet 46(6), 375-81. doi:10.1136/jmg.2008.063917.

Bruggemann, N., Mitterer, M., Lanthaler, A.J., Djarmati, A., Hagenah, J., Wiegers, K., Winkler, S., Pawlack, H., Lohnau, T., Pramstaller, P.P., Klein, C., Lohmann, K. 2009. Frequency of heterozygous Parkin mutations in healthy subjects: need for careful prospective follow-up examination of mutation carriers. Parkinsonism Relat Disord 15(6), 425-9. doi:10.1016/j.parkreldis.2008.11.014.

Camargos, S.T., Dornas, L.O., Momeni, P., Lees, A., Hardy, J., Singleton, A., Cardoso, F. 2009. Familial Parkinsonism and early onset Parkinson's disease in a Brazilian movement disorders clinic: phenotypic characterization and frequency of SNCA, PRKN, PINK1, and LRRK2 mutations. Mov Disord 24(5), 662-6. doi:10.1002/mds.22365.

Carrasquillo, M.M., Khan, Q.u.A., Murray, M.E., Krishnan, S., Aakre, J., Pankratz, V.S., Thuy, N., Ma, L., Bisceglio, G., Petersen, R.C., Younkin, S.G., Dickson, D.W., Boeve, B.F., Graff-Radford, N.R., Ertekin-Taner, N. 2014. Late-onset Alzheimer disease genetic variants in posterior cortical atrophy and posterior AD. Neurology 82(16), 1455-62. doi:10.1212/wnl.0000000000000335.

Coppola, G., Chinnathambi, S., Lee, J.J., Dombroski, B.A., Baker, M.C., Soto-Ortolaza, A.I., Lee, S.E., Klein, E., Huang, A.Y., Sears, R., Lane, J.R., Karydas, A.M., Kenet, R.O., Biernat, J., Wang, L.S., Cotman, C.W., Decarli, C.S., Levey, A.I., Ringman, J.M., Mendez, M.F., Chui, H.C., Le Ber, I., Brice, A., Lupton, M.K., Preza, E., Lovestone, S., Powell, J., Graff-Radford, N., Petersen, R.C.,

Boeve, B.F., Lippa, C.F., Bigio, E.H., Mackenzie, I., Finger, E., Kertesz, A., Caselli, R.J., Gearing, M., Juncos, J.L., Ghetti, B., Spina, S., Bordelon, Y.M., Tourtellotte, W.W., Frosch, M.P., Vonsattel, J.P., Zarow, C., Beach, T.G., Albin, R.L., Lieberman, A.P., Lee, V.M., Trojanowski, J.Q., Van Deerlin, V.M., Bird, T.D., Galasko, D.R., Masliah, E., White, C.L., Troncoso, J.C., Hannequin, D., Boxer, A.L., Geschwind, M.D., Kumar, S., Mandelkow, E.M., Wszolek, Z.K., Uitti, R.J., Dickson, D.W., Haines, J.L., Mayeux, R., Pericak-Vance, M.A., Farrer, L.A., Alzheimer's Disease Genetics, C., Ross, O.A., Rademakers, R., Schellenberg, G.D., Miller, B.L., Mandelkow, E., Geschwind, D.H. 2012. Evidence for a role of the rare p.A152T variant in MAPT in increasing the risk for FTD-spectrum and Alzheimer's diseases. Hum Mol Genet 21(15), 3500-12. doi:10.1093/hmg/dds161.

Corder, E.H., Saunders, A.M., Strittmatter, W.J., Schmechel, D.E., Gaskell, P.C., Small, G.W., Roses, A.D., Haines, J.L., Pericakvance, M.A. 1993. Gene dose of apolipoprotein E type-4 allele and the risk of Alzheimer's disease in late-onset families Science 261(5123), 921-3. doi:10.1126/science.8346443.

Cruts, M., Theuns, J., Van Broeckhoven, C. 2012. Locus-specific mutation databases for neurodegenerative brain diseases. Hum Mutat 33(9), 1340-4. doi:10.1002/humu.22117.

Deng, H., Le, W.D., Hunter, C.B., Ondo, W.G., Guo, Y., Xie, W.J., Jankovic, J. 2006. Heterogeneous phenotype in a family with compound heterozygous parkin gene mutations. Arch Neurol 63(2), 273-7. doi:10.1001/archneur.63.2.273.

Fong, H., Wang, C.Z., Knoferle, J., Walker, D., Balestra, M.E., Tong, L.M., Leung, L., Ring, K.L., Seeley, W.W., Karydas, A., Kshirsagar, M.A., Boxer, A.L., Kosik, K.S., Miller, B.L., Huang, Y.D. 2013. Genetic Correction of Tauopathy Phenotypes in Neurons Derived from Human Induced Pluripotent Stem Cells. Stem Cell Rep 1(3), 226-34. doi:10.1016/j.stemcr.2013.08.001.

Gatz, M., Reynolds, C.A., Fratiglioni, L., Johansson, B., Mortimer, J.A., Berg, S., Fiske, A., Pedersen, N.L. 2006. Role of genes and environments for explaining Alzheimer disease. Arch Gen Psychiatry 63(2), 168-74. doi:10.1001/archpsyc.63.2.168.

Genomes Project, C., Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., McVean, G.A. 2012. An integrated map of genetic variation from 1,092 human genomes. Nature 491(7422), 56-65. doi:10.1038/nature11632.

Ghani, M., Lang, A.E., Zinman, L., Nacmias, B., Sorbi, S., Bessi, V., Tedde, A., Tartaglia, M.C., Surace, E.I., Sato, C., Moreno, D., Xi, Z., Hung, R., Nalls, M.A., Singleton, A., St George-Hyslop, P., Rogaeva, E. 2015. Mutation analysis of patients with neurodegenerative disorders using NeuroX array. Neurobiol Aging 36(1), 545 e9-14. doi:10.1016/j.neurobiolaging.2014.07.038.

Grove, M.L., Yu, B., Cochran, B.J., Haritunians, T., Bis, J.C., Taylor, K.D., Hansen, M., Borecki, I.B., Cupples, L.A., Fornage, M., Gudnason, V., Harris, T.B., Kathiresan, S., Kraaij, R., Launer, L.J., Levy, D., Liu, Y., Mosley, T., Peloso, G.M., Psaty, B.M., Rich, S.S., Rivadeneira, F., Siscovick, D.S., Smith, A.V., Uitterlinden, A., van Duijn, C.M., Wilson, J.G., O'Donnell, C.J., Rotter, J.I., Boerwinkle, E. 2013. Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. PLoS One 8(7), e68095. doi:10.1371/journal.pone.0068095.

Guo, J.F., Xiao, B., Liao, B., Zhang, X.W., Nie, L.L., Zhang, Y.H., Shen, L., Jiang, H., Xia, K., Pan, Q., Yan, X.X., Tang, B.S. 2008. Mutation analysis of Parkin, PINK1, DJ-1 and ATP13A2 genes in Chinese patients with autosomal recessive early-onset Parkinsonism. Mov Disord 23(14), 2074-9. doi:10.1002/mds.22156.

Guo, J.F., Zhang, X.W., Nie, L.L., Zhang, H.N., Liao, B., Li, J., Wang, L., Yan, X.X., Tang, B.S. 2010. Mutation analysis of Parkin, PINK1 and DJ-1 genes in Chinese patients with sporadic early onset parkinsonism. J Neurol 257(7), 1170-5. doi:10.1007/s00415-010-5485-8.

Hedrich, K., Marder, K., Harris, J., Kann, M., Lynch, T., Meija-Santana, H., Pramstaller, P.P., Schwinger, E., Bressman, S.B., Fahn, S., Klein, C. 2002. Evaluation of 50 probands with early-onset Parkinson's disease for Parkin mutations. Neurology 58(8), 1239-46.

Illarioshkin, S.N., Periquet, M., Rawal, N., Lucking, C.B., Zagorovskaya, T.B., Slominsky, P.A., Miloserdova, O.V., Markova, E.D., Limborska, S.A., Ivanova-Smolenskaya, I.A., Brice, A. 2003.

Mutation analysis of the parkin gene in Russian families with autosomal recessive juvenile parkinsonism. Movement Disorders 18(8), 914-9. doi:10.1002/mds.10467.

Illenberger, S., Zheng-Fischhofer, Q.Y., Preuss, U., Stamer, K., Baumann, K., Trinczek, B., Biernat, J., Godemann, R., Mandelkow, E.M., Mandelkow, E. 1998. The endogenous and cell cycle-dependent phosphorylation of tau protein in living cells: Implications for Alzheimer's disease. Mol Biol Cell 9(6), 1495-512.

Jellinger, K.A. 2014. Neuropathology of Dementia Disorders. Journal of Alzheimer's Disease & Parkinsonism 04(01). doi:10.4172/2161-0460.1000135.

Koziorowski, D., Hoffman-Zacharska, D., Slawek, J., Szirkowiec, W., Janik, P., Bal, J., Friedman, A. 2010. Low frequency of the PARK2 gene mutations in Polish patients with the early-onset form of Parkinson disease. Parkinsonism Relat Disord 16(2), 136-8. doi:10.1016/j.parkreldis.2009.06.010.

Lesage, S., Lohmann, E., Tison, F., Durif, F., Durr, A., Brice, A., French Parkinson's Disease Genetics Study, G. 2008. Rare heterozygous parkin variants in French early-onset Parkinson disease patients and controls. J Med Genet 45(1), 43-6. doi:10.1136/jmg.2007.051854.

Liu, F., Gong, C.X. 2008. Tau exon 10 alternative splicing and tauopathies. Mol Neurodegener 3, 10. doi:10.1186/1750-1326-3-8.

Lohmann, E., Thobois, S., Lesage, S., Broussolle, E., du Montcel, S.T., Ribeiro, M.J., Remy, P., Pelissolo, A., Dubois, B., Mallet, L., Pollak, P., Agid, Y., Brice, A., French Parkinson's Dis Genetics, S. 2009. A multidisciplinary study of patients with early-onset PD with and without parkin mutations. Neurology 72(2), 110-6. doi:10.1212/01.wnl.0000327098.86861.d4.

Madegowda, R.H., Kishore, A., Anand, A. 2005. Mutational screening of the parkin gene among South Indians with early onset Parkinson's disease. J Neurol Neurosurg Psychiatry 76(11), 1588-90. doi:10.1136/jnnp.2004.046888.

McLaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P., Cunningham, F. 2010. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics 26(16), 2069-70. doi:10.1093/bioinformatics/btq330.

Mukrasch, M.D., Bibow, S., Korukottu, J., Jeganathan, S., Biernat, J., Griesinger, C., Mandelkow, E., Zweckstetter, M. 2009. Structural Polymorphism of 441-Residue Tau at Single Residue Resolution. PLoS Biol 7(2), 399-414. doi:10.1371/journal.pbio.1000034.

Nalls, M.A., Bras, J., Hernandez, D.G., Keller, M.F., Majounie, E., Renton, A.E., Saad, M., Jansen, I., Guerreiro, R., Lubbe, S., Plagnol, V., Gibbs, J.R., Schulte, C., Pankratz, N., Sutherland, M., Bertram, L., Lill, C.M., DeStefano, A.L., Faroud, T., Eriksson, N., Tung, J.Y., Edsall, C., Nichols, N., Brooks, J., Arepalli, S., Pliner, H., Letson, C., Heutink, P., Martinez, M., Gasser, T., Traynor, B.J., Wood, N., Hardy, J., Singleton, A.B., International Parkinson's Disease Genomics, C., Parkinson's Disease meta-analysis, c. 2015. NeuroX, a fast and efficient genotyping platform for investigation of neurodegenerative diseases. Neurobiol Aging 36(3), 1605 e7-12. doi:10.1016/j.neurobiolaging.2014.07.028.

Periquet, M., Latouche, M., Lohmann, E., Rawal, N., De Michele, G., Ricard, S., Teive, H., Fraix, V., Vidailhet, M., Nicholl, D., Barone, P., Wood, N.W., Raskin, S., Deleuze, J.F., Agid, Y., Durr, A., Brice, A., French Parkinsons Dis Genetics, S., European Consortium Genetic, S. 2003. Parkin mutations are frequent in patients with isolated early-onset parkinsonism. Brain 126, 1271-8. doi:10.1093/brain/awg136.

Pittman, A.M., Fung, H.-C., de Silva, R. 2006. Untangling the tau gene association with neurodegenerative disorders. Hum Mol Genet 15, R188-R95. doi:10.1093/hmg/ddl190.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., Rehm, H.L., Committee, A.L.Q.A. 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genetics in medicine : official journal of the American College of Medical Genetics 17(5), 405-24. doi:10.1038/gim.2015.30.

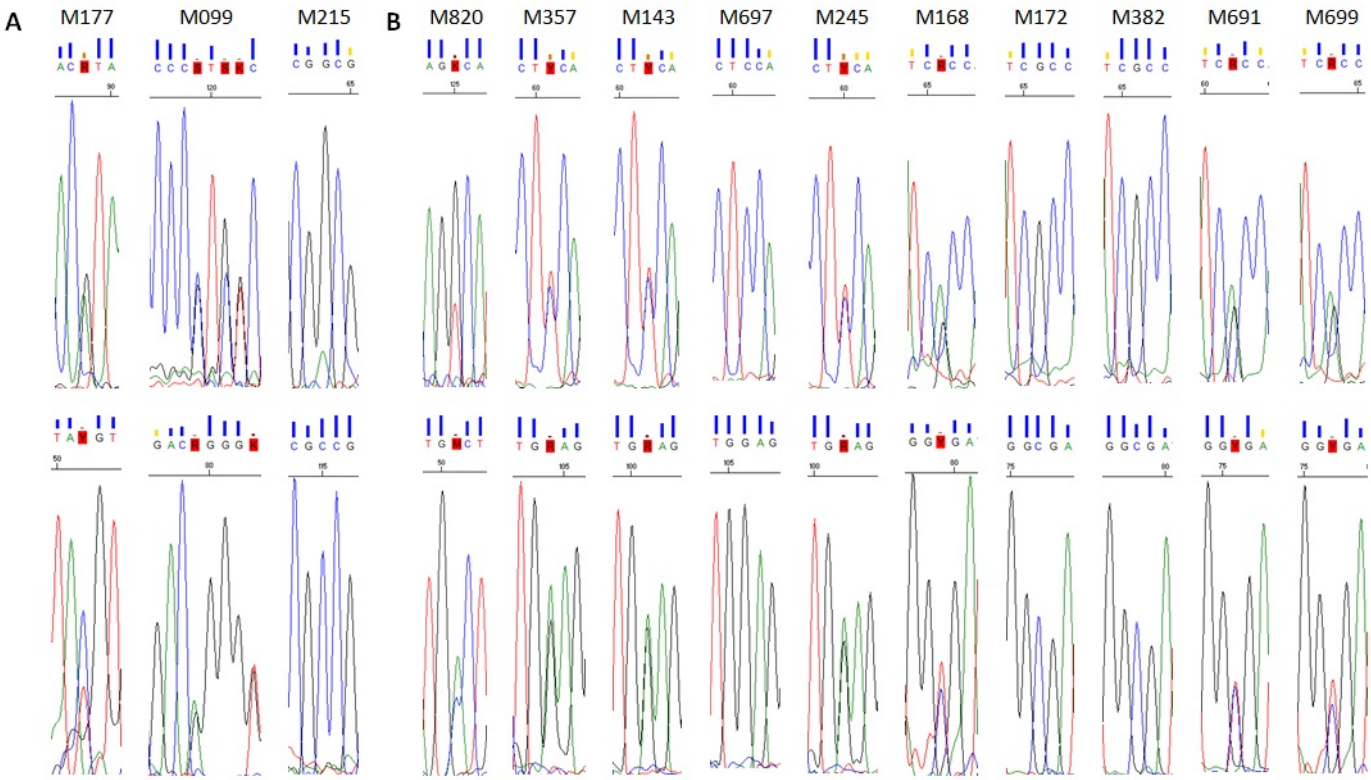
Ryman, D.C., Acosta-Baena, N., Aisen, P.S., Bird, T., Danek, A., Fox, N.C., Goate, A., Frommelt, P., Ghetti, B., Langbaum, J.B.S., Lopera, F., Martins, R., Masters, C.L., Mayeux, R.P., McDade, E., Moreno, S., Reiman, E.M., Ringman, J.M., Salloway, S., Schofield, P.R., Sperling, R., Tariot, P.N., Xiong, C.J., Morris, J.C., Bateman, R.J., Dominantly Inherited Alzheimer, N. 2014. Symptom onset in autosomal dominant Alzheimer disease A systematic review and meta-analysis. Neurology 83(3), 253-60.

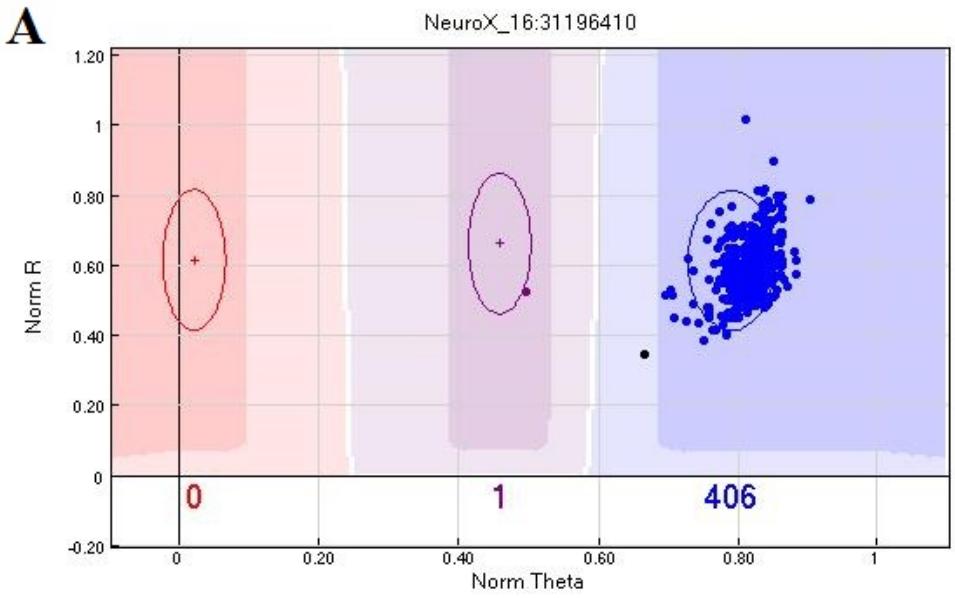
Scherfler, C., Khan, N.L., Pavese, N., Eunson, L., Graham, E., Lees, A.J., Quinn, N.P., Wood, N.W., Brooks, D.J., Piccini, P.P. 2004. Striatal and cortical pre- and postsynaptic dopaminergic dysfunction in sporadic parkin-linked parkinsonism. Brain 127, 1332-42. doi:10.1093/brain/awh150.

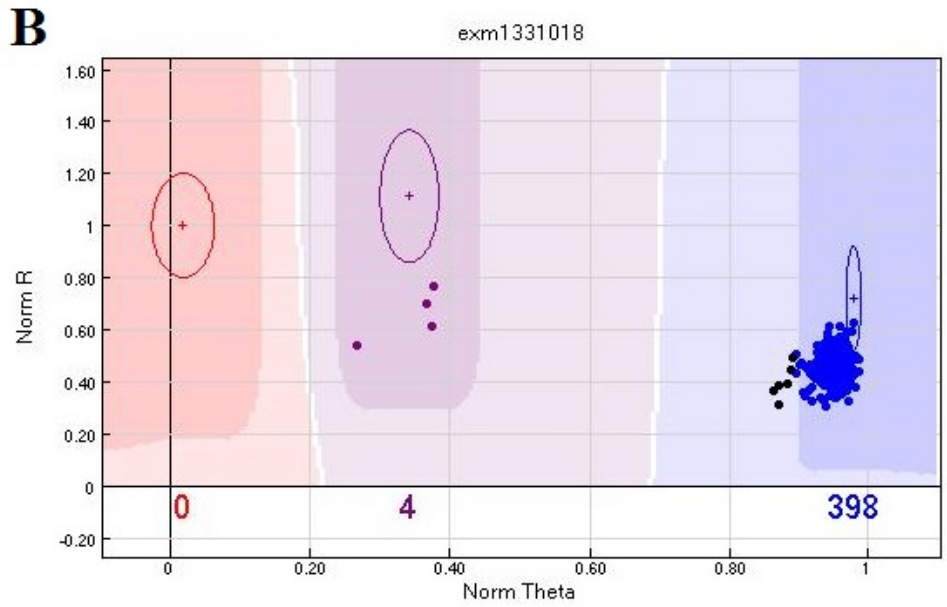
Sironi, F., Primignani, P., Zini, M., Tunesi, S., Ruffmann, C., Ricca, S., Brambilla, T., Antonini, A., Tesei, S., Canesi, M., Zecchinelli, A., Mariani, C., Meucci, N., Sacilotto, G., Cilia, R., Isaias, I.U., Garavaglia, B., Ghezzi, D., Travil, M., Decarli, A., Coviello, D.A., Pezzoli, G., Goldwurm, S. 2008. Parkin analysis in early onset Parkinson's disease. Parkinsonism & Related Disorders 14(4), 326-33. doi:10.1016/j.parkreldis.2007.10.003.

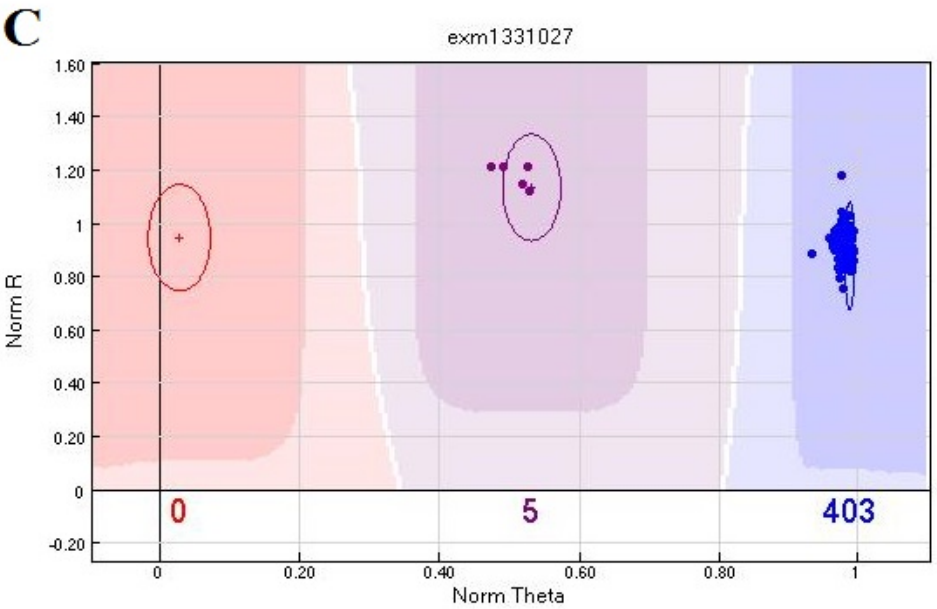
WHO. 2012. Top Ten leading Causes of Death by Country Income Group. Top Ten Causes of Death. World Health Organisation.

Wingo, T.S., Lah, J.J., Levey, A.I., Cutler, D.J. 2012. Autosomal recessive causes likely in early-onset Alzheimer disease. Arch Neurol 69(1), 59-64. doi:10.1001/archneurol.2011.221.


A
M177
M099
M215
B
M820
M357
M143
M697
M245
M168
M172
M382
M691
M699


A
NeuroX_16:31196410
Norm R
Norm Theta
0
1
406


B
exm1331018
Norm R
Norm Theta
0
4
398


C
exm1331027
Norm R
Norm Theta
0
5
403

Supplementary

Marker	MapInfo
NeuroX_GRN_Ala237fs	4031641
NeuroX_GRN_Asn118fs	42427596
NeuroX_GRN_Asn119del	42427601
NeuroX_GRN_Thr52Hisfs	42426809
NeuroX_LRRK2_Glu2490fs	40761451
NeuroX_LRRK2_IVS30-6C_T	40704227
NeuroX_LRRK2_IVS31+3A_G	40704454
NeuroX_LRRK2_IVS32+14G_A	40707989
NeuroX_LRRK2_IVS33+6T_A	40709108
NeuroX_LRRK2_IVS37-9A_G	40716953
NeuroX_LRRK2_IVS38+7C_T	40717115
NeuroX_LRRK2_IVS46-14T_A	40753048
NeuroX_LRRK2_IVS46-8delT	40753054
NeuroX_PARK2_Ala291fs	161990449
NeuroX_PARK2_Asn428fs	161781121
NeuroX_PARK2_Cys238fs	162394356
NeuroX_PARK2_Cys323fs	161969996
NeuroX_PARK2_Gln34fs_del_A	162864412
NeuroX_PARK2_Gln34fs_del_AG	162864411
NeuroX_PARK2_Gly179fs	162475205
NeuroX_PARK2_Pro133del	162683570
NeuroX_PARK2_Trp74fs	162683748
NeuroX_PARK2_Val324fs	161969998
NeuroX_PARK7_Pro158del	8045015
NeuroX_PINK1_23bp_del_ex7	20975486
NeuroX_PINK1_534_535insQ	20977040
NeuroX_PINK1_Asp525fs	20977011
NeuroX_PINK1_Cys549fs	20977085
NeuroX_PINK1_Lys520fs	20976995

Supplementary Table 1
Markers with updated MapInfo

List of 29 markers which had their MapInfo updated from zero to their correct base position (genome reference build GRCh37)

Marker		
exm1624485	exm2264787	exm1648550
exm1624661	exm-rs525869_ver4	exm1667365
exm1624677	exm1624877	exm1624484
exm1624783	exm1667344	exm1624788
exm1624797	exm2273163	exm1624791
exm1624862	exm2273277	exm1624872
exm1624867	exm2264764	exm1624917
exm1624886	exm1624934	exm1624943_ver3
exm1624907	exm1667357	exm1624953
exm1624944	exm1625534	exm1624965
exm1624960	exm1625323	exm1624996_ver3
exm1624975	exm1624879	exm1625022
exm1624995	exm1624887	exm1625063
exm1624999_ver3	exm1648651	exm1625610
exm1625001	exm1625041	exm1625631
exm1625200	exm1625556_ver2	exm1648554
exm1625277	exm1624641	exm1667351
exm1625538	exm1625029	exm1667363
exm1625569	exm1625819	exm2209863
exm1625571	exm1624434	exm2248602
exm1625573	exm1625064	exm2248962
exm1648538	exm1625554	
exm1667350	exm1625555_ver3	
exm1667504	exm1625807	
exm1667508	exm1667356	
exm1667537	exm1667360	
exm1667541	exm2206194	
exm2262791	exm1625045	
exm2263170	exm1625545	
exm2263174	exm1624766	
exm2263176	exm1625068	
exm2263276	exm1625216	
exm2263278_ver3	exm1625222	
exm2263279	exm1625253	
exm2263280	exm1648569	
exm2263281	exm2208735	
exm2268448	exm1624792	
exm2273075	exm1624794	
exm2273161	exm1624804	
exm2273221	exm1624869	
exm2273222	exm1624931	
exm2273223	exm1624954_ver3	
exm2273224	exm1625005	
exm2273278	exm1625025_ver2	
exm1624939_ver2	exm1625046	
exm1624946	exm1625162	
exm1625030_ver4	exm1625204	
exm-rs5941436_ver2	exm1625383_ver2	
exm-rs2573905	exm1625553	
exm1625510	exm1625780	

Supplementary Table 2

Markers with updated chromosome

List of 121 markers updated from chromosome X to the pseudo autosomal region (XY)

						Pathogenic Nature Unclear			Causative			Not Pathogenic		
Database	**Gene**	**Disease**	**Database**	**Designed on NeuroX**	**Passed QC**	**Database**	**Designed on NeuroX**	**Passed QC**	**Database**	**Designed on NeuroX**	**Passed QC**	**Database**	**Designed on NeuroX**	**Passed QC**
ADFTDPD	APP	AD	31	10	10	6	3	3	24	6	6	1	1	1
	C9orf72	FTD/ALS	12	0	0	11	0	0	0	0	0	1	0	0
	CHMP2B	FTD	12	7	7	4	0	0	4	3	3	4	4	4
	FUS	FTD/ALS	44	21	20	18	3	3	22	15	15	4	3	2
	GRN	FTD	146	33	33	45	10	10	66	13	13	35	10	10
	LRRK2	PD	128	71	71	54	15	15	6	3	3	68	53	53
	MAPT	FTD	73	11	10	27	10	9	44	1	1	2	0	0
	PARK2	PD	144	70	70	22	4	4	59	27	27	63	39	39
	PARK7	PD	21	6	5	5	1	0	1	0	0	15	5	5
	PINK1	PD	130	52	52	30	12	12	20	7	7	80	33	33
	PSEN1	AD	197	76	74	4	3	3	185	69	67	8	4	4
	PSEN2	AD	25	9	9	5	1	1	13	5	5	7	3	3
	SNCA	PD	5	1	1	1	1	1	3	0	0	1	0	0
	TARDBP	FTD/ALS	45	26	26	9	3	3	34	21	21	2	2	2
	VCP	FTD/ALS	20	14	14	0	0	0	18	13	13	2	1	1
HPP	PRNP	Prion	42	5	5	18	4	4	24	1	1	0	0	0
Total			1075	412	407	259	70	68	523	184	182	293	158	157

Supplementary Table 3

Coverage on the NeuroX of variants from familial online databases

Pathogenic nature and NeuroX coverage of variants from three online databases, including the Alzheimer Disease & Frontotemporal Dementia Mutation Database (AD&FTDMDB), Parkinson Disease Mutation Database (PDMutDB) and Human Prion Database (HPDB). Variants from the online databases were clustered into the gene they reside in (Gene) and the disease most associated with the gene (Disease). The number of variants in each gene is given (Database), followed by the number designed on the NeuroX (Designed on NeuroX) and finally the number that passed quality control in the final NeuroX dataset (Passed QC). These variants are then subcategorised depending on the pathogenicity documented in the database; those that cause disease (Causative), those that have unclear pathogenic nature (Pathogenic Nature Unclear), and those that do not cause disease (Not Pathogenic). Key: AD; Alzheimer's disease, FTD; frontotemporal dementia, ALS; amyotrophic lateral sclerosis, PD; Parkinson's disease.