Late-onset Alzheimer´s disease is associated with mitochondrial DNA 7028C/haplogroup H and D310 poly C tract heteroplasmy.

Eliecer Coto[1,5], Juan Gómez[1], Belén Alonso[1], Ana I. Corao[1], Marta Díaz[1], Manuel Menéndez[2], Carmen Martínez[3], María T. Calatayud[4], Germán Morís[4], Victoria Álvarez[1]

[1] Genética Molecular-Laboratorio de Medicina, Hospital Universitario Central Asturias, Oviedo, Spain.
[2] Neurología, Hospital A. Buylla, Mieres, Spain.
[3] Neurología, Hospital Cabueñes, Gijón, Spain.
[4] Neurología, Hospital Universitario Central Asturias, Oviedo, Spain.
[5] Departamento de Medicina, Universidad de Oviedo, Oviedo, Spain.

Corresponce to:
Dr. Eliecer Coto
Genética Molecular.
Hospital Universitario Central de Asturias
33006. Oviedo. Spain
Tel./ Fax. (34).985.27.36.57.
Email: ***eliecer.coto@sespa.princast.es***

Running head: mitochondrial DNA variants in Alzheimer´s disease

Key words: Alzheimer´s disease; mitochondrial polymorphisms; haplogroups; heteroplasmy; genetic risk.

Abbreviations: LOAD, late-onset Alzheimer´s disease; mtDNA, mitochondrial DNA; SNPs, single nucleotide polymorphisms; *APOE*, apolipoprotein E gene; ROS, reactive oxygen species; mtOD, mitochondrial oxidative damage.

Conflict of interest: The authors declare no conflict of interest.

Letter Neurogenetics (word count: 600).

There is a growing evidence that mitochondria play an important role in the pathogenesis of late-onset Alzheimer´s disease (LOAD) [**1]**. Certain haplogroups defined by combinations of common mtDNA SNPs have been associated with longevity and age-related diseases. Recent studies reported a significant association between haplogroup H and LOAD [**2-4**; **supplementary table 1]**. Differences between the mtDNA SNPs / haplogroups in the efficiency of oxygen consumption and ATP and ROS production could partly explain these associations [5]. Although the relative risk attributed to this haplogroup was low (<1.5) in all the studies, this could have important implications considering the high incidence of LOAD at the population level. It is therefore very important to determine the influence of mtDNA polymorphisms in LOAD by studying large cohorts from different populations. We report the results from Asturias (Northern Spain) based on a total of 500 LOAD-patients and 500 controls (**supplementary file**).

The mtDNA 7028 C was significantly more frequent among the patients (p=0.001; OR=1.52, 95% CI=1.18-1.95) (**table 1**). Allele frequencies for the other six SNPs that defined the mitochondrial haplogroups did not differ between patients and controls (data not shown). Haplogroup H (defined by 7028C) was significantly more frequent among the patients (**supplementary file**). Multiple logistic regression with *APOE- ε4* and mtDNA 7028 C status, age, and gender as covariates showed that *APOE- ε4* and 7028 C were associated with LOAD. We found a synergistic effect between the two risk variants: the OR for *APOE- ε4* + 7028 C (OR=5.30, 95%CI= 3.72-7.54) was higher than the OR for each separately. The non-*APOE- ε4* + 7028 T had a protective effect (OR=0.43; 95%CI= 0.33-0.75).

Patients and controls were classified as D310 7C or D310 >7C (poly-C tract at position 310 of the mtDNA) (**supplementary file**). We found a significantly higher frequency of the >7C alleles among the patients (**Table 1**). We also found a significantly higher frequency of 7028 C + D310 >7C among the patients (p<0.001; OR=1.45) while the 7028 T + D310 7C haplogroup was significantly more frequent among the elderly controls (p<0.001; OR=0.51) (**supplementary file**). We sequenced individuals classified as 7C and >7C. While all the D310 7C individuals showed a clear sequence with the 7C tract, some of the >7C patients showed a mixed sequence suggestive of the existence of heteroplasmy. We confirmed that these were naturally occurring mixed sequences (instead of being artifacts from the PCR) by cloning into plasmids the PCR products from several

individuals and sequencing individual clones (**supplementary file**). A higher frequency of D310 heteroplasmy in the blood cells of LOAD patients compared to healthy controls was recently reported [6]. The D310 poly C tract is on a region implicated in the replication and transcription of the mtDNA. The presence of eight or more Cs would create mtDNA replication instability, and this could explain the higher degree of heteroplasmy among individuals with longer C tracts. A functional effect of the D310 >7C alleles could also explain some of the reported differences between H and the other haplogroups.

In conclusion, we reported a significant association between the mtDNA 7028 C allele and LOAD. LOAD patients also showed a higher frequency of D310 >7C alleles, that were linked to a high degree of heteroplasmy at this polyC tract. The mtDNA 7028C + D310>7C haplogroup would thus represent a new genetic marker for LOAD.

Acknowledgements

This work was supported by grants from the Spanish Fondo de Investigaciones Sanitarias-Fondos FEDER European Union (FIS-05/008 and 08/0915). M.D. is the recipient of a Contrato de Apoyo a la Investigación-Fondo de Investigaciones Sanitarias. Authors wish to thank Asociación Parkinson Asturias and Obra Social Cajastur for their support.

References.

1. Swerdlow RH, Burns JM, and Khan SM (2010) The Alzheimer's disease mitochondrial cascade hypothesis. J Alzheimers Dis 2: S265-79.
2. Chagnon P, Gee M, Filion M, Robitaille Y, Belouchi M, and Gauvreau D (1999) Phylogenetic analysis of the mitochondrial genome indicates significant differences between patients with Alzheimer disease and controls in a French-Canadian founder population. Am J Medical Genet 85: 20–30.
3. Santoro A, Balbi V, Balducci E, Pirazzini C, Rosini F, Tavano F, Achilli A, Siviero P, Minicuci N, Bellavista E, Mishto M, Salvioli S, Marchegiani F, Cardelli M, Olivieri F, Nacmias B, Chiamenti AM, Benussi L, Ghidoni R, Rose G, Gabelli C, Binetti G, Sorbi S, Crepaldi G, Passarino G, Torroni A, and Franceschi C (2010) Evidence for sub-haplogroup h5 of mitochondrial DNA as a risk factor for late onset Alzheimer's disease. PLoS One 5: e12037.
4. Maruszak A, Canter JA, Styczyńska M, Zekanowski C, and Barcikowska M (2009) Mitochondrial haplogroup H and Alzheimer's disease--is there a connection?. Neurobiol Aging 30: 1749-1755.
5. Gómez-Durán A, Pacheu-Grau D, López-Gallardo E, Díez-Sánchez C, Montoya J, López-Pérez MJ, and Ruiz-Pesini E (2010) Unmasking the causes of multifactorial disorders: OXPHOS differences between mitochondrial haplogroups. Hum Mol Genet 19: 3343-3353.
6. Wang PN, Lee HC, Wang CH, Ping YH, Liu TY, Chi CW, Lin KN, and Liu HC (2009) Heteroplasmy of mitochondrial D310 mononucleotide repeat region in the blood of patients with Alzheimer's disease. J Alzheimer´s Dis 18: 345-353.

Table 1. Mean age, gender distribution, and frequency of APOE-ε4 carriers in LOAD-patients and controls.

	Index cohort		**Replication cohort**		**Total**	
	Patients (n=300)	Controls (n=250)	Patients (n=200)	Controls (n=250)	Patients (n=500)	Controls (n=500)
7028 C **7028 T**	**166 (0.54)** **134 (0.46)**	**112 (0.45)** **138 (0.55)**	**106 (0.53)** **94 (0.47)**	**108 (0.43)** **142 (0.57)**	**272 (0.53)** **228 (0.47)**	**220 (0.45)** **280 (0.55)**
APOE-ε34+ ε44 7028C 7028T	121 (0.40) 68 (0.56) 53 (0.44)	45 (0.18) 22 (0.49) 23 (0.51)	82 (0.41) 45 (0.55) 37 (0.45)	43 (0.17) 22 (0.50) 21 (0.50)	203 (0.41) 113 (0.56) 90 (0.44)	88 (0.18) 44 (0.50) 44 (0.50)
D310 >7C **D310 7C**	**195 (0.65)** **105 (0.35)**	**144 (0.57)** **106 (0.43)**	**131 (0.66)** **69 (0.34)**	**146 (0.58)** **104 (0.42)**	**326 (0.65)** **174 (0.35)**	**290 (0.58)** **210 (0.42)**
APOE-ε34+ ε44 D310 >7C D310 7C	121 (0.40) 79 (0.65) 42 (0.35)	45 (0.18) 24 (0.53) 21 (0.47)	82 (0.41) 54 (0.66) 28 (0.34)	43 (0.17) 23 (0.54) 20 (0.46)	203 (0.41) 133 (0.66) 70 (0.34)	88 (0.18) 47 (0.53) 41 (0.47)

7028 C : total: p=0.001; OR=1.52 (1.18-1.95)
D310>C: total: p=0.02; OR=1.36 (1.06-1.75)
7028 C + APOE-ε34/ε44: total: p<0.0001; OR=5.30 (3.72-7.54)
D310>7C + APOE-ε34/ε44: total: p<0.0001; OR=3.40 (2.37-4.87)

Supplementary Material.

Late-onset Alzheimer´s disease is associated with mitochondrial DNA 7028C/haplogroup H and D310 poly C tract heteroplasmy.

Eliecer Coto, Juan Gómez, Belén Alonso, Ana I. Corao, Marta Díaz, Manuel Menéndez, Carmen Martínez, María T. Calatayud, Germán Morís, Victoria Álvarez

Patients and controls.

The study included a total of 500 patients (65% women) who fulfilled the criteria for clinical probable Alzheimer´s disease established by the National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA). All the patients had an onset age ≥65 years (range 65-91), and were recruited by Neurologists from three Hospitals in the region of Asturias (Northern Spain, total population one million). The control group consisted of 500 individuals aged 65-90 years (55% women). They were from the general population or spouses of the patients and were apparently healthy, although we did not perform neuropsycological examination to exclude the existence of symptoms of cognitive impairment. The mtDNA polymorphisms were also determined in a total of 200 healthy blood donors aged 18-55 years (the younger controls; 52% women). Patients and controls were Spanish Caucasians and gave their informed consent to participate in the study. This research was approved by the Ethical Committee of Hospital Central Asturias.

The association between the mtDNA/haplogroup frequencies and LOAD was assessed following a two step approach. First, we compared a total of 300 patients recruited through the Hospital Universitario Central Asturias (HUCA) in the period 2000-2005 and 250 healthy controls (the index cohort). The replication cohort consisted of 200 patients recruited through Hospitals Cabueñes-Gijón and A. Buylla-Mieres and 250 controls. The main characteristics of the two cohorts are summarised in **table S1**.

Haplogroups classification.

Seven mtDNA single nucleotide polymorphisms (SNPs G4580A, C7028T, G8251A, G9055A, A12308G, G13368A, and G13708A) that defined the most common mitochondrial haplogroups in Caucasians were determined in the index cohort (see **table S2** and the mitomap database for

haplogroups definition; www.mitomap.org). The C7028C was also determined in the replication cohort.

DNA was obtained from the cells in 5 ml of blood and SNPs C7028T, G9055A, A12308G, G13368A, and G13708A were determined with a real time polymerase chain reaction (PCR) Taqman assay in an ABI7500 machine (Applied biosystems) (**table S3**; **figure S1)**. SNPs G4580A and G8251A were analysed through digestion with a restriction enzyme of a PCR fragment (PCR-RFLP) followed by electrophoresis on agarose gels, as reported (Castro et al. Int. J. Cardiol. 112, 202-206, 2006; Huerta et al. J. Neurol. Sci. 236, 49-54, 2005; Palacín et al. Mitochondrion 11, 176-181, 2011).

Mit D310 poly C tract polymorphism.

The D310 poly C is a tract of 7 or more C followed by a T and 5 Cs. We designated a forward primer that created an *AluI* site in PCR fragments with 303-309 C, that were seen on agarose gels as digested fragments (**table S3**; **figure S1)**. Non digested fragments were classified as >7 C alleles. The accuracy of this genotyping method was determined by amplifying and sequencing the mtDNA region between nucleotides 1-660 from 40 patients previously genotyped as 7C (n=20) and >7C (n=20) (**figure S2**).

The degree of heteroplasmy at the D310 polyC tract was assessed in 200 patients and 200 controls through amplification of DNA with primers GACATCATAACAAAAAATTTCCACCA (D310F, forward) and TTTGGGGTTTGGCAGAGATGT (D310R-FAM, reverse; annealing at 58ºC). The reverse primer was 5´-end labeled with 6-FAM, and the PCR products were size-fractioned in an ABI3130 equipment (Applied Bisystems) and the fluorescent profile of each fragment characterized with the *GenMapper* software (Applied Biosystem). Fragments containing 7C were visualized as peaks of 85 bp (1 bp longer for each extra C). The degree of heteroplasmy was determined by the existence of multiple peaks (**figure S3**).

DNA cloning.

To confirm that the observed D310 heteroplasmy was due to the presence of heterogeneous of polyC tracts (instead of being PCR artifact), we amplified a fragment from the mtDNA (nucleotides 15 to 660) with primers GATCACAGGTCTATCACCCTATTAAC (cloning-F, forward) and GACCAAACCTATTTGTTTATGGGG (cloning-R, reverse; annealing at 60ºC). The PCR products were cloned into the pGEM-T vector (Promega) and transfected into *E. coli*. Bacteria were growth onto agar plates containing ampicillin and single colonies were picked and growth on

liquid media. Plasmids were isolated and amplified with primers D310F and D310R-FAM and the size of the PCR product was determined. The two strands of plasmid inserts were also sequenced with primers cloning-F and cloning-R (figure S4).

***APOE* genotyping.**

All the patients and controls were genotyped for the *APOE*-ε2/ ε3/ ε4 polymorphisms following a PCR-RFLP method (available upon request).

Statistical analysis.

Allele and haplogroup frequencies were compared between patients and controls through the χ^2 test. Odds ratios (ORs) and their 95% confidence intervals (CIs) were also calculated. We compared each haplogroup frequency with the frequency of all other haplogroups pooled into one group. The χ^2 test was also used to compare the differences in the distribution of 7028 + D310 alleles between patients and controls (**table S4**). A $p<0.01$ was considered as the level of statistical significance. The sample size required to reach a power of 80 was defined online (http://statpages.org/proppowr.html). The Student´s t-test was used to compare the difference in means between the groups. The logistic regression was used to determine the association of polymorphisms/haplogroups with LOAD after correcting by covariates (age, sex, *APOE* genotype). All the statistical analyses were performed with the R-software (http://www.R-project.org).

Table S1. Main characteristics of patients and controls in the first and second (replication) cohorts, and the total patients and controls.

	Index Patients (n=300)	Index Controls (n=250)	Replication patients (n=200)	Replication controls (n=250)	Total patients (n=500)	Total controls (n=500)	p value*
Mean age (years)	77	76	76	76	77	76	ns
Range (years)	65-91	65-90	65-90	65-89	65-91	65-90	ns
Male (%)	35%	45%	35%	45%	35%	45%	ns
APOE-ε34	106 (35%)	44 (18%)	71 (36%)	42 (17%)	177 (35%)	86 (17%)	<0.0001
APOE-ε44	15 (5%)	1 (<1%)	11 (5%)	1 (<1%)	26 (6%)	2 (<1%)	<0.0001
APOE-ε34+ε44	121 (40%)	45 (18%)	82 (41%)	43 (17%)	203 (41%)	88 (18%)	<0.0001

* p values, total patients vs. controls.

Table S2. Definition of the seven mitochondrial haplogroups based on the 7 SNPs, and haplogroup frequencies in LOAD-patients and healthy controls in the index cohort (parenthesis, %).

HAPLO GROUP	C7028T	G4580A	G8251A	G9055A	A12308G	G13368A	G13708A	Patients (n=300)	Elderly Controls (n=250)	Younger Controls (n=200)
H*	**C**	**G**	**G**	**G**	**A**	**G**	**G**	**163 (53)**	**109 (45)**	**92 (46)**
J	**T**	**G**	**G**	**G**	**A**	**G**	**A**	**29 (9)**	**29 (12)**	**22 (11)**
K	**T**	**G**	**G**	**A**	**G**	**G**	**G**	**10 (4)**	**18 (7)**	**16 (8)**
T	**T**	**G**	**G**	**G**	**A**	**A**	**G**	**30 (10)**	**23 (9)**	**20 (10)**
U	**T**	**G**	**G**	**G**	**G**	**G**	**G**	**32 (10)**	**29 (12)**	**22 (11)**
V	**T**	**A**	**G**	**G**	**A**	**G**	**G**	**11 (4)**	**12 (5)**	**10 (5)**
X	**T**	**G**	**G**	**G**	**A**	**G**	**G**	**2 (1)**	**5 (2)**	**2 (1)**
other								**15 (5)**	**14 (6)**	**8 (4)**

* p=0.012; OR=1.54 (1.10-2.16) H vs. non-H, patients vs. elderly controls

Table S3. Primer sequence and PCR-RFLP conditions for the three mtDNA polymorphisms SNPs that were analyzed through this technique, and primers and allele-specific probes for the SNPs analyzed through taqman assays. Taqman probes were labeled with VIC or FAM, and NFQ as the quencher.

Polymorphism	PCR – Primers Forward/Reverse	Annealing (ºC)	Restriction enzyme	Allele size (base pairs)
G4580A	ACCTATCACACCCCATCCTAAA AGGATTATGGATGCGGTT	60	*NheI*	G:200+100 A:300
G8251A	ATGCAATTCCCGGACGTCT TTCACTGTAAAGAGGTGTTGGTTCT	54	*HaeIII*	G:149+114 A:263
D310 polyC	AACAAAAAATTTCCACCAAACCCCAGC GCGGGGGTTGTATTGATGAG	62	*AluI*	7C=190+28 >7C=218

SNP	Primers (forward/reverse)	probes
C7028T	GCAAACTCATCACTAGACATCGTACT AAGCCTCCTATGATGGCAAATACAG	C: VIC-5´-TAGTGGAAGTGGGCTACAA-3´-NFQ T: FAM-5´ TAGTGGAAGTGAGCTACAA-3´-NFQ
A9055G	GCCACCTACTCATGCACCTA GTGTAGAGGGAAGGTTAATGGTTGAT	G: VIC-5´- ATTGGAAGCGCCACCC-3´-NFQ A: FAM-5´- TTGGAAGCACCACCC-3´-NFQ
A12308G	CATGGCTTTCTCAACTTTTAAAGGATAACA GGTGGTTATAGTAGTGTGCATGGTTATTA	A: VIC-5´- CCCCAAAAATTTTG-3´-NFQ G: FAM-5´- CCCAAGAATTTTG-3´-NFQ
A13368G	ACGCCTTCTTCAAAGCCATACTATT TTTCGAATATCTTGTTCATTGTTAAGGTTGTG	G:VIC-5´- TGCTCCGGGTCCATC-3´-NFQ A: FAM-5´- TGCTCCGGATCCATC-3´-NFQ
A13708G	CCCCACCCTACTAAACCCCATTAA TGTTAGTAATGAGAAATCCTGCGAATAGG	G:VIC-5´- ACGCCTGGCAGCCG-3´-NFQ A: FAM-5´- ACGCCTGACAGCCG-3´-NFQ

Table S4. Frequencies of the D310 poly C according to the C7028T polymorphism. The frequency of heteroplasmy was determined from 50 individuals with each of the four haplogroups in patients and elderly controls.

C7028T	D310	Patients N=500	Heteropl. (%)	Elderly Controls N=500	Heteropl. (%)	Younger Controls (n=200)
C	**>7C**	165 (0.33)	90%	127 (0.25)	80%	52 (0.26)
C	**7C**	107 (0.21)	2%	93 (0.19)	0	44 (0.22)
T	**>7C**	161 (0.32)	80%	163 (0.33)	70%	64 (0.32)
T	**7C**	67 (0.14)	<1%	117 (0.23)	0	40 (0.20)
	D310 >7C **7028 C**	326 (0.65) 272 (0.54)		290 (0.58) 220 (0.44)		116 (0.58) 96 (0.48)

7028T + D310 7C : P<0.001 OR=0.51 (0.36-0.70)

7028C + D310 >7C: p<0.001; OR=1.45 (1.10-1.90)

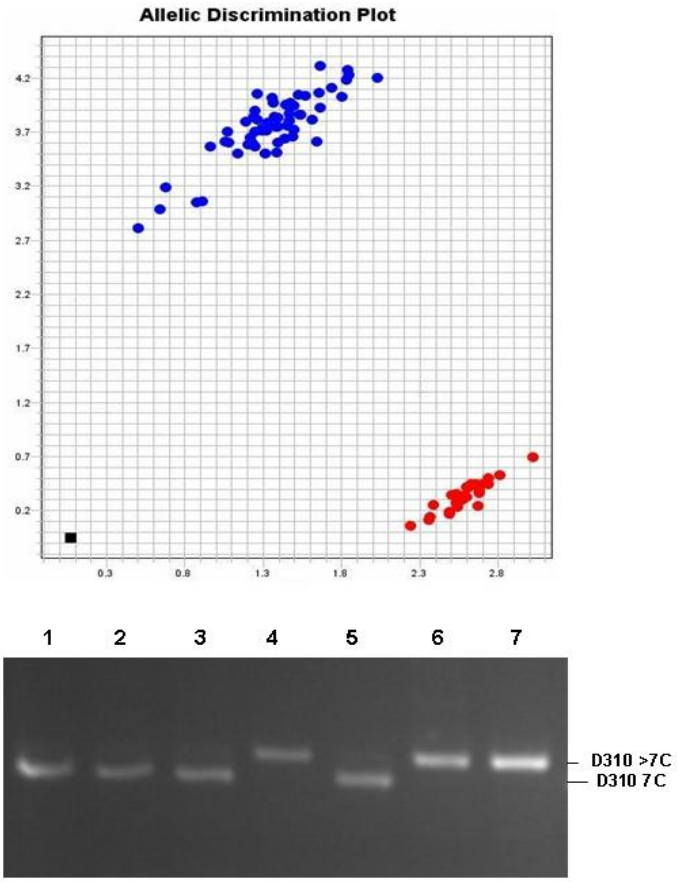

Figure S1. Taqman assay for C7028T (upper figure): blue dots, 7028C; red dots, 7028T; black dot, negative (no DNA) control.

PCR-RFLP to determine the D310 variant. After digestion with *AluI* and electrophoresis on agarose gels the D310 allele was seen as a digested fragment, while D310 >7C gave non-digested fragments.

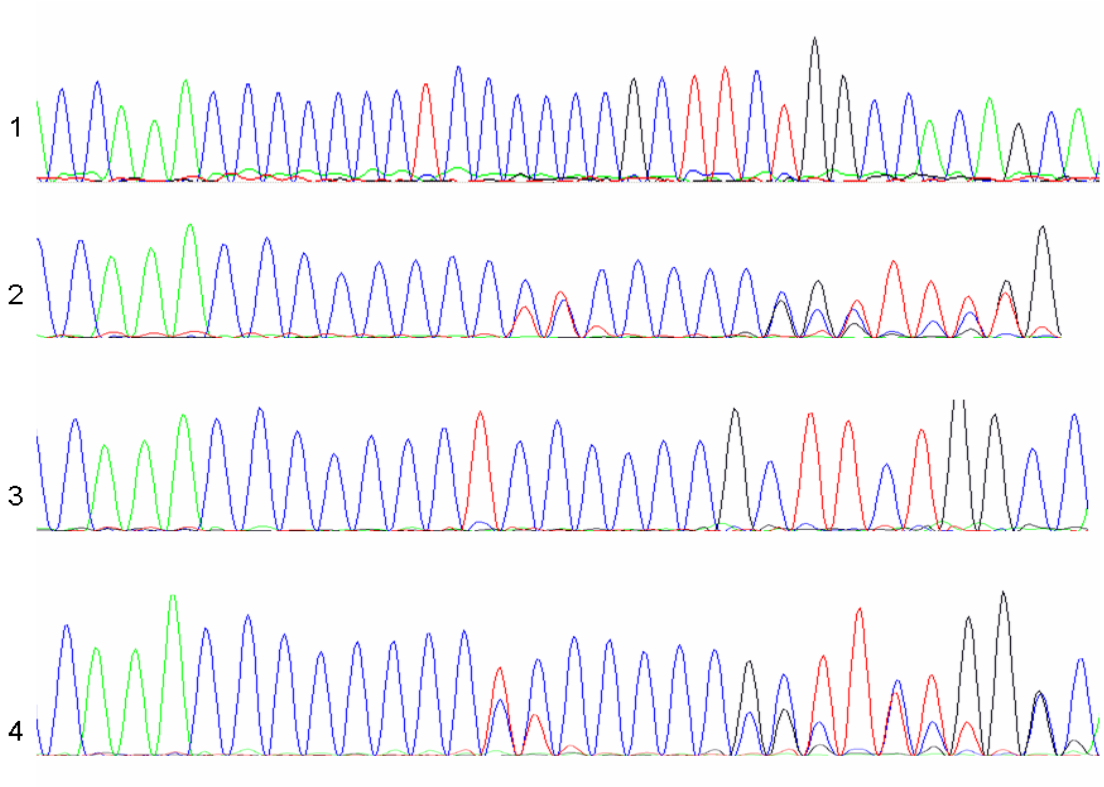

Figure S2. Sequences of the mtDNA D310 region showing homoplasmy for the 7 Cs allele (lanes 1, 3) and heteroplasmy for alleles >7Cs (lanes 2, 4).

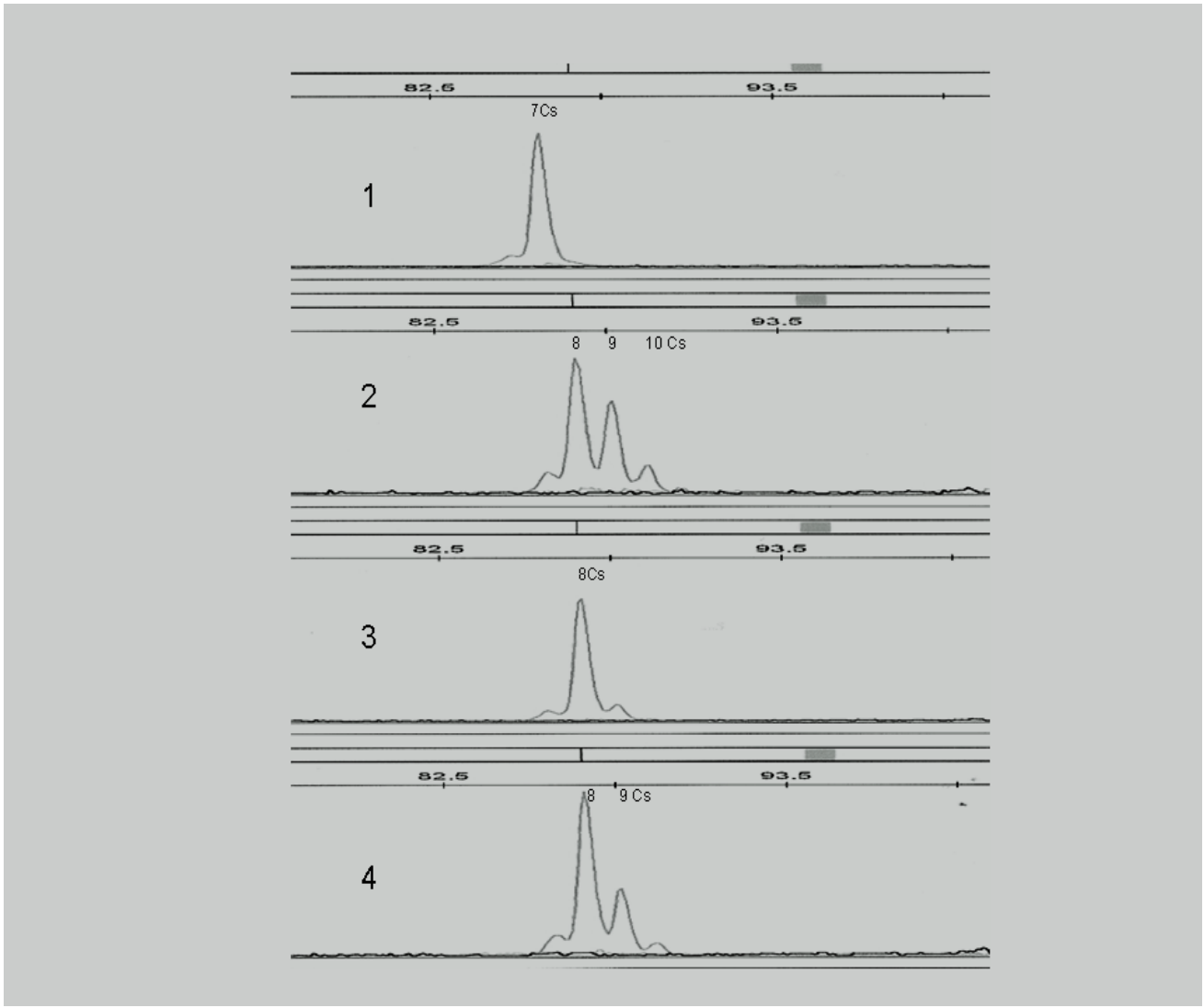

Figure S3. PCR fluorescent fragments showing homoplasmy for mtDNA with D310 7C (lane 1) and D310 8C (lane 3), and heroplasmy with alleles of 8-10 Cs (lanes 2 and 4; see also the supplementary figures 2 and 3).

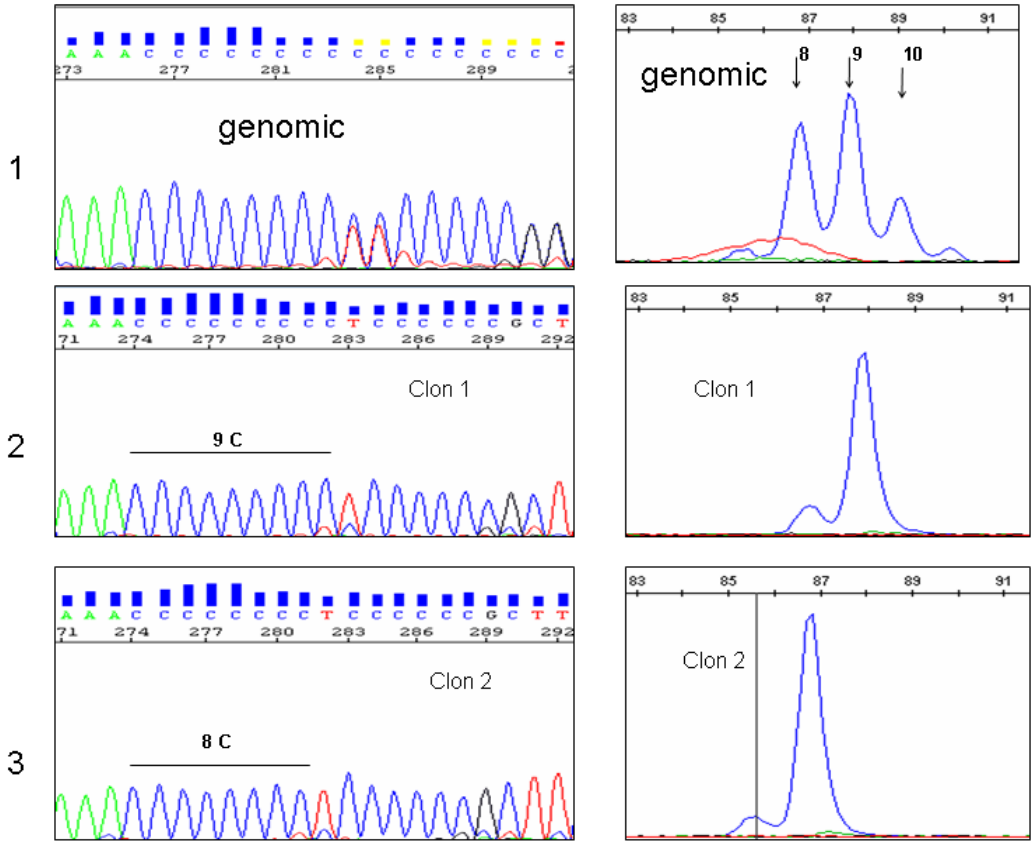

Figure S4. Sequence and size of a PCR fragment from total DNA showing the mtDNA heteroplasmy in a patients (lane 1). Sequence and size of the PCR fragment from plasmid clones from the same individuals, showing the homoplasmy for alleles of 8 and 9 Cs (lanes 2 and 3).

Supplementary table. Summary of the haplogroup frequencies in five studies.

(ND, value not indicated in the report).

	Maruszak et al. Poland	**Krüger et al. Finland**	**Mancuso et al. Italy**	**Chinnery et al. England**	**Coto et al. Spain**
Patients / Controls	222 / 252	128 / 99	209 / 191	185 / 179	300 / 250
Patients Mean age (range)	72 (ND)	58 (38 – 64)	74 (50-94)	ND	77 (65-91)
Patients, Female	68%	ND	62%	ND	65%
Controls Mean age (range)	70 (ND)	40 (19 – 64)	69 (55-92)	ND	76 (65-90)
Controls, Female (%)	74%	ND	51%	ND	55%
APOE 4 + Patients / Controls (%)	61 / 20	ND	23 / 8	ND	40 / 18
Haplogroups, % Patients / controls					
H	51 / 43	44 / 34	44 / 41	48 / 41	53 / 45
J	7 / 11	5 / 2	8 / 7	9 / 14	12 / 9
K	3 / 5	4 / 3	6 / 9	11 / 12	4 / 7
T	9 / 10	6 / 8	7 / 11	9 / 10	10 / 9
U	12 / 17	32 / 34	14 / 13	18 / 18	10 / 12
V	3 / 3	5 / 6	4 / 2	ND	4 / 5
W	4 / 1	4 / 8	2 / 2	2 / 1	2 / 3

Maruszak A et al. Neurobiol. Aging 30, 1749-1755, 2009.
Mancuso M et al. Neurol. Sci. 28, 142–147, 2007.
Krüger J et al. Mol. Neurodegener. 5, e8, 2010.
Chinnery PF et al. Neurology 55, 302–304, 2000.