



Published in final edited form as:

*Neurobiol Aging*. 2016 October ; 46: 236.e1–236.e6. doi:10.1016/j.neurobiolaging.2016.06.018.

## Exome sequencing in a consanguineous family clinically diagnosed with early onset Alzheimer's disease identifies an homozygous *CTSF* mutation

Jose Bras<sup>a</sup>, Ruth Djaldetti<sup>b</sup>, Ana Margarida Alves<sup>c</sup>, Simon Mead<sup>d</sup>, Lee Darwent<sup>c</sup>, Alberto Lleo<sup>e</sup>, Jose Luis Molinuevo<sup>f</sup>, Rafael Blesa<sup>e</sup>, Andrew Singleton<sup>g</sup>, John Hardy<sup>c</sup>, Jordi Clarimon<sup>e</sup>, and Rita Guerreiro<sup>a,\*</sup>

<sup>a</sup>Department of Molecular Neuroscience, Institute of Neurology, University College London, Queen Square, London WC1N 3BG, UK and Department of Medical Sciences and Institute of Biomedicine – iBiMED, University of Aveiro, 3810-193 Aveiro, Portugal

<sup>b</sup>Department of Neurology, Rabin Medical Center, Beilinson Campus, and Felsenstein Research Center, Petah Tiqva, affiliated to the Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel

<sup>c</sup>Department of Molecular Neuroscience, Institute of Neurology, University College London, Queen Square, London WC1N 3BG, UK

<sup>d</sup>MRC Prion Unit, Department of Neurodegenerative Diseases, UCL Institute of Neurology, London, UK

<sup>e</sup>Memory Unit, Neurology Department and Sant Pau Biomedical Research Institute, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, Spain, and CIBERNED, Center for Networked Biomedical Research into Neurodegenerative Diseases, Madrid, Spain

<sup>f</sup>Alzheimer's Disease and Other Cognitive Disorders Unit, Department of Neurology, Institute of Neurosciences, Hospital Clínic i Universitari de Barcelona, IDIBAPS, Barcelona, Spain

<sup>g</sup>Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA

### Abstract

We have previously reported the whole genome genotyping analysis of two consanguineous siblings clinically diagnosed with early onset Alzheimer's disease. In this analysis we identified several large regions of homozygosity shared between both affected siblings, which we suggested could be candidate loci for a recessive genetic lesion underlying the early onset Alzheimer's disease in these cases. We have now performed exome sequencing in one of these siblings and identified the potential cause of disease: the *CTSF* c.G1243A:p.Gly415Arg mutation in homozygosity. Bi-allelic mutations in this gene have been shown to cause Type B Kufs disease, an adult-onset neuronal ceroid lipofuscinosis with some cases resembling the impairment seen in Alzheimer's disease.

\*Corresponding author: Rita Guerreiro, PhD, r.guerreiro@ucl.ac.uk, Department of Molecular Neuroscience, UCL Institute of Neurology, 1 Wakefield Street (1st Floor), London WC1N 1PJ England.

## Keywords

Early onset Alzheimer's disease; Kufs disease; Recessive; Exome sequencing; Homozygosity; *CTSF*

---

## 1. Introduction

The absence of consanguineous families in the deeply characterized datasets of North America and Europe has made mapping of recessive loci challenging in Alzheimer's disease (AD) (Ghani, et al., 2015). Nonetheless, a recent study suggested that nearly 90% of early-onset AD (EOAD) cases are likely the result of autosomal recessive inheritance (Wingo, et al., 2012). In 2009 we described the whole genome genotyping analysis of two EOAD siblings from a Jewish Israeli family originating from Morocco whose parents were first cousins and neurologically healthy (Clarimon, et al., 2009). At that time the formal assessment of a recessive aetiology in EOAD cases had not been systematically performed and the identification of large tracts of homozygosity shared between these two affected siblings added information to the possibility of recessive heritability being part of this disease. Currently, two mutations in *APP* are known to cause AD through an autosomal recessive pattern of inheritance. The p.Ala673Val has been found to cause disease only in the homozygous state, whereas heterozygous carriers were found to be unaffected. This mutation was also shown to affect APP processing with co-incubation of mutated and wild-type peptides conferring instability on A $\beta$  aggregates and inhibiting amyloidogenesis and neurotoxicity. The highly amyloidogenic effect of the p.Ala673Val mutation in the homozygous state and its anti-amyloidogenic effect in the heterozygous state accounted for the autosomal recessive pattern of inheritance of the disease (Di Fede, et al., 2009). Similarly, the p.Glu693 was found to segregate in a Japanese family with an autosomal recessive pattern. In this case the secretion of total A $\beta$  was markedly reduced by the mutation, but the variant A $\beta$  was more resistant to proteolytic degradation (Tomiyama, et al., 2008).

The contribution of long runs of homozygosity (ROHs) to AD has also been tested in different populations with diverse results. The global burden of large genome-wide ROHs in outbred datasets from North America and Europe did not show a significant association with AD (Nalls, et al., 2009, Sims, et al., 2011). In the Wadi Ara population (an isolated Arab community from northern Israel) the controls studied were found to have more ROHs than the AD cases (Sherva, et al., 2011) and ROHs were found to be significantly associated with AD in a Caribbean Hispanic data set (Ghani, et al., 2013) and in an outbred African American population (Ghani, et al., 2015).

In this study we used exome sequencing in order to identify any potential pathogenic mutations within long ROHs shared between two siblings from a consanguineous family, clinically diagnosed with EOAD.

## 2. Methods

### 2.1. Patients

We analyzed a previously described Jewish Israeli family originating from Morocco (Clarimon, et al., 2009). The family was composed of seven siblings born to first-degree consanguineous parents who died after the age of 90 years without any signs of cognitive impairment. Two siblings were diagnosed with EOAD and agreed to participate in the genetic study. The others refused to give samples for genetic analyses. Written informed consent was obtained from both patients included in the study.

### 2.2. Genetic analyses

In our previous study we sequenced *PSEN1* (exons 3–12), *PSEN2* (exons 3–12), *APP* (exons 16 and 17), *TAU*, *PGRN*, and *PRNP* and found no pathogenic mutations in these genes. We moved on to perform whole genome genotyping using an Illumina Human-Hap240S genotyping array in the two siblings to identify long ROHs (>1MB) shared between them (Clarimon, et al., 2009). Now we have performed exome sequencing in one of the siblings. For this analysis we prepared genomic DNA according to Illumina's TruSeq Sample Preparation v.3 and performed the exome capture with Illumina's TruSeq Exome Enrichment according to the manufacturer's instructions. Sequencing was performed on an Illumina HiSeq2500 with 100-bp paired-end reads. Following quality control procedures, the sample yielded 10.2 Gb of high-quality, aligned data. This amount of data resulted in a mean target coverage of 35.8x, 83% of targets being covered at greater than or equal to 10x, and less than 0.2% of targets not being covered at least once. We performed sequence alignment and variant calling against the reference human genome (hg19) by using the Burrows-Wheeler Aligner (Li and Durbin, 2009) and the Genome Analysis Toolkit (DePristo, et al., 2011, McKenna, et al., 2010). Prior to variant calling, PCR duplicates were removed with the Picard software. On the basis of the hypothesis that the mutation was rare, we excluded all common variants (MAF > 5%) identified in dbSNP v.142 and in our in-house database of sequencing data for other diseases (n > 2,000). Given the apparent autosomal-recessive mode of inheritance in the family and the previous whole genome genotyping data, we focused the analysis on homozygous variants located in the shared ROHs. The vast majority of known coding exons within these ROHs (93%) were adequately covered (5x or more). We used Sanger sequencing to confirm the variant of interest identified by exome sequencing and to establish the presence of the variant in the second sibling. Because of the low amount of DNA available for analyses we performed whole genome amplification of both samples using the REPLI-g Amplification kit (Qiagen) before confirming the additional 10 variants identified by exome sequencing in the shared ROHs. For all variants we performed PCR amplification and purified the resulting products with ExoSAP-IT (USB), then performed direct Sanger sequencing of both strands using BigDye Terminator v.3.1 chemistry v.3.1 (Applied Biosystems) and an ABI 3730XL Genetic Analyzer (Applied Biosystems). Sequencing traces were analyzed with Sequencher software v.4.2 (Gene Codes).

### 3. Results

#### 3.1. Family description

As previously described, the family consisted of seven siblings from a first-cousin marriage (Clarimon, et al., 2009). At the time of the clinical study the mother was 90 years old with mild gait difficulties but otherwise healthy, with no cognitive loss. The father died at age 90 due to cerebral stroke. The proband is a 64-year-old man who was born in Morocco and migrated to Israel in 1956. At the age of 58 he presented with a 9-month history of insidious, progressive impairment of memory first noticed by the family. Early symptoms were repeated questioning, forgetting names and loss of interest in his family. Two months after the initial visit, the Mini-Mental State Examination score was 9/30. Cognitive examination revealed diffuse cognitive impairment, most notably of short and long-term memory, verbal and visual recall, and visuospatial abilities. Calculation, attention tasks, and information processing speed were markedly impaired. In addition, he had moderate dressing apraxia and right-left disorientation. According to the NINCDS-ADRDA criteria the clinical examination and neuropsychological profile were consistent with EOAD (McKhann, et al., 1984). One of his siblings initially presented clinical symptoms at the age of 58, when he was laid-off from work due to cognitive decline. One year later, he was institutionalized in a psychiatric institution with a clinical diagnosis of EOAD. He died at the age of 70 and did not present any extrapyramidal signs during the course of disease. The diagnoses were established based on the clinical presentation of the siblings and it was not possible to establish biomarker or neuropathological confirmation of these diagnoses.

#### 3.2. Exome and Sanger sequencing analyses

We identified 8312 variants in the exome sequenced sample with a high or medium effect in the gene (nonsynonymous missense, nonsense, affecting splicing sites, nonframeshift and frameshift insertions/deletions). From these, 3921 were homozygous. Based on the hypothesis that the mutation underlying the disease in this family is rare, we filtered the results to include variants with minor allele frequencies (MAFs) <0.001 in the 1000 Genomes Project and the ExAC datasets, resulting in 39 variants (12 indels, 26 nonsynonymous and 1 splicing variants), 14 of which were novel. From these, 11 variants were located in genes within the shared ROHs (Table 1). The bioinformatics analysis of these 11 variants involved the determination of genomic conservation for the variant locus; prediction of the effect in the protein using different predictive algorithms; information regarding the expression of the respective gene in brain tissue; and a constraint metric evaluating the probability of the gene being intolerant to variation. Details for this analysis are presented in Supplementary Table 1. Only one variant was found to follow all the inclusion and exclusion criteria: Cathepsin F (*CTSF*) p.[(Gly415Arg)];[(Gly415Arg)] (NM\_003793; c.[1243G>A];[1243G>A]). Sanger sequencing confirmed the presence of the variant in both siblings. This variant is present in 10 of 74678 alleles in the ExAC dataset (MAF=0.0001) and is never seen in homozygosity. In fact, no *CTSF* loss of function variants are present in homozygosity in this dataset (Exome Aggregation Consortium [ExAC], Cambridge, MA (URL: <http://exac.broadinstitute.org>), accessed August 2015). The mutation is predicted to be damaging/deleterious/disease causing by Polyphen-2, SIFT and

MutationTaster, respectively and is highly conserved both between species and between the human members of the papain family of cysteine proteases (Figure 1).

Because of the low amounts of DNA available for analyses, whole genome amplification of both samples was done before PCR and Sanger sequencing to confirm the additional 10 variants. Seven of these variants were confirmed before complete exhaustion of the DNA (Supplementary Table 1).

The subsequent analysis of our in-house exome sequencing data for Alzheimer's disease cases (542 EOADs and 185 late onset ADs) revealed a second variant of potential interest in an EOAD case. This homozygous mutation c.214-6C>T was present in the splice site region of *CTSF* exon 2. The same variant is present in homozygosity in one sample and in heterozygosity in 46 samples of the ExAC database. The mutation was found in a proband originally from Slovakia presenting with a history of dementia and possible myoclonus. The patient was 46 years old at the time of clinical onset, and 51 at time of sampling. There were white matter changes in subcortical regions on MRI. No family history of dementia and no other samples from family members were available for analysis, however the presence of this variant in homozygosity in the ExAC database argues against its pathogenicity in EOAD.

The *CTSF* locus on chromosome 11 has not been identified as associated with Alzheimer's disease by any of the reports analyzing long runs of homozygosity (Supplementary Table 2).

#### 4. Discussion

The availability of only two samples from this family for genetic analyses precludes the definitive conclusion about the genetic cause of recessive dementia in the two siblings studied. However, the results obtained from exome sequencing the proband indicate the possible involvement of *CTSF* in the phenotype. The brothers are likely to be AD phenocopies with Kufs' pathology, but the unavailability of samples for pathology analyses do not allow us to confirm this hypothesis.

Bi-allelic mutations in *CTSF* have been previously identified as the cause of type B Kufs disease (Smith, et al., 2013). This is the most common adult form of neuronal ceroid lipofuscinosis (NCLs), which are neurodegenerative diseases mainly characterized by the abnormal accumulation of lipopigment in lysosomes. The diagnosis of Kufs disease is often very challenging and the definitive diagnosis may require the neuropathological assessment of brain tissue (Smith, et al., 2013, Williams and Mole, 2012).

Kufs disease is genetically heterogeneous with mutations in *CLN6* known to cause recessive Type A disease and mutations in *DNAJC5* underlying some cases of dominant Kufs (Arsov, et al., 2011, Benitez, et al., 2011, Noskova, et al., 2011). Type A is characterized by progressive myoclonus epilepsy and cognitive decline; and type B is characterized by movement and behavioral abnormalities associated with dementia (Berkovic, et al., 1988). More recently, mutations in *CTSF* have been identified as the cause of recessive type B Kufs disease, characterized by dementia and motor features (Di Fabio, et al., 2014, Smith, et al., 2013).

Four homozygous or compound heterozygous *CTSF* mutations have been associated with Kufs disease: p.[(Gly458Ala)];[(Ser480Leu)], p.[(Gln321Arg)];[(Gln321Arg)], p.[(Tyr231Cys)];[(Ser319Leufs\*27)], and c.[213+1G>C];[213+1G>C]. Most of these are located in the peptidase C1 domain towards the C-terminal end of *CTSF* and are predicted to affect the protein function (Figure 2). The fourth mutation (c.[213+1G>C];[213+1G>C]) has been shown to create an mRNA lacking exon 1 and was predicted to form a truncated N-terminus of cathepsin F 20% shorter than the wild-type protein with a possible loss-of-function acting in a pseudo-dominant fashion (Di Fabio, et al., 2014).

Of the 9 reported cases from 4 families, 8 were women. The onset of the disease occurred on average at 32.7+/-15.3 years with seizures. Progressive cognitive impairment was present in all cases at an average age of 40+/-14.2 years (Table 2) (Di Fabio, et al., 2014, Smith, et al., 2013).

All identified mutations were associated with a clinical picture of Kufs disease characterized by an older age at onset, seizures and delayed cognitive decline. However, even within the same kindred, significant differences were observed in the clinical presentation of the disease, which seem to happen within a spectrum between typical Kufs disease and late onset cognitive impairment (Di Fabio, et al., 2014, Smith, et al., 2013).

Even though no seizures were reported in the two siblings studied here, and they had a typical clinical presentation of Alzheimer's disease with the first signs of dementia at 58 years, given the broad presentation of *CTSF* mutations and the challenging diagnosis of Kufs B disease, *CTSF* p.[(Gly415Arg)];[(Gly415Arg)] is the most probable genetic cause of disease. The difficulty in the diagnosis of Kufs disease and the overlap with Alzheimer's disease has been recently highlighted by the finding of a novel *PSEN1* mutation (p.Leu381Phe) in a family originally diagnosed with Kufs disease. Electron microscopy studies of the proband's skin biopsy showed lipofuscin containing phagocytic cells and distinct curvilinear lysosomal inclusion bodies, which were highly suggestive of NCL (Dolzhanskaya, et al., 2014). These results substantiate the suggestion that *CTSF* should be screened in cases of early-onset recessive dementia.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

This work was supported in part by an anonymous donor; the Medical Research Council (UK), and by the National Institute of Health Research's Biomedical Research Centre and Biomedical Research Unit (Dementia) at UCL Hospitals NHS Foundation Trust; fellowships from the Alzheimer's Society to J.B. and R.G and by the 2014 European Grand Prix for Young Researcher-SCOR awarded by the *Fondation pour la Recherche sur Alzheimer* to R.G.. The authors would like to thank the Exome Aggregation Consortium and the groups that provided exome variant data for comparison. A full list of contributing groups can be found at <http://exac.broadinstitute.org/about>.

## References

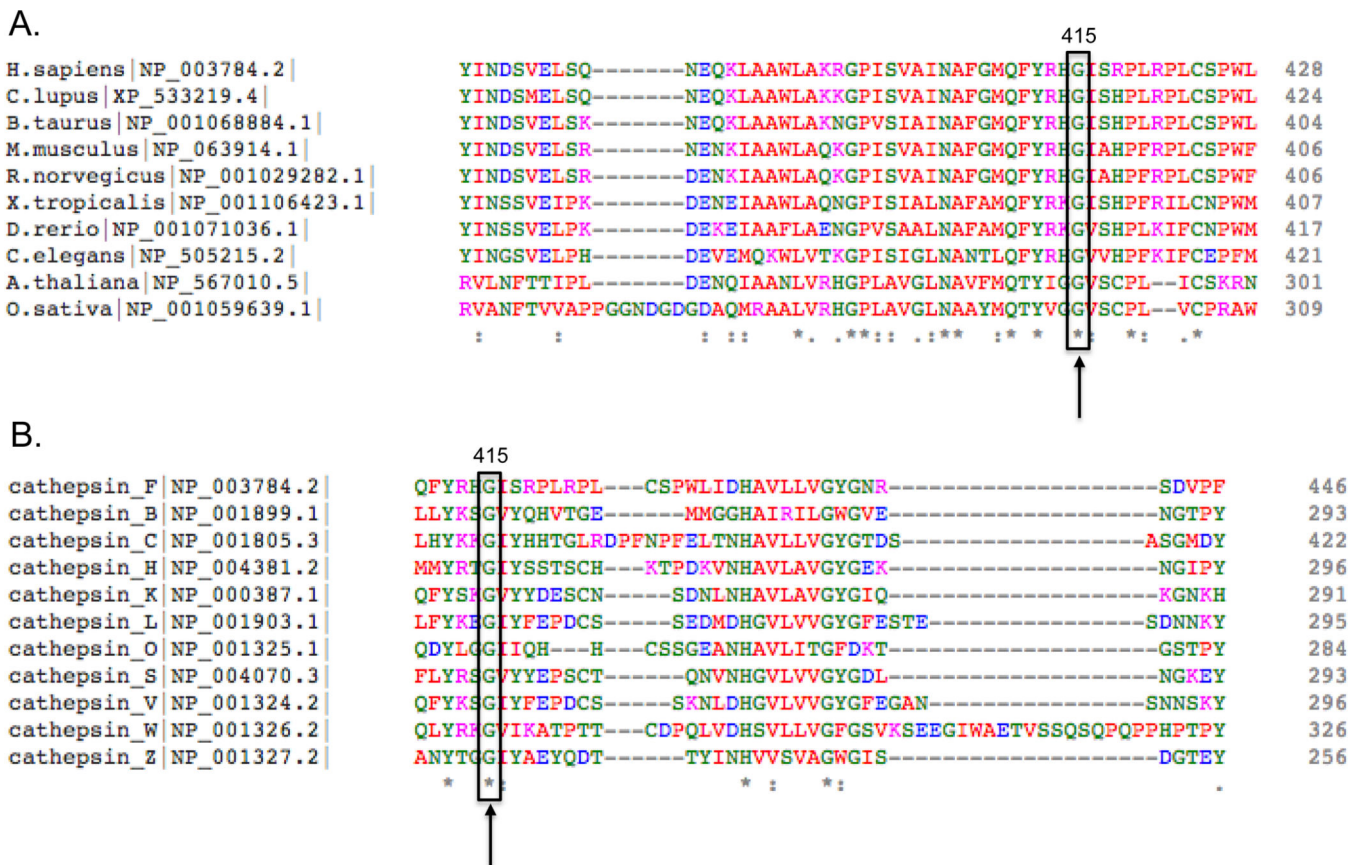
Arsov T, Smith KR, Damiano J, Franceschetti S, Canafoglia L, Bromhead CJ, Andermann E, Vears DF, Cossette P, Rajagopalan S, McDougall A, Sofia V, Farrell M, Aguglia U, Zini A, Meletti S,



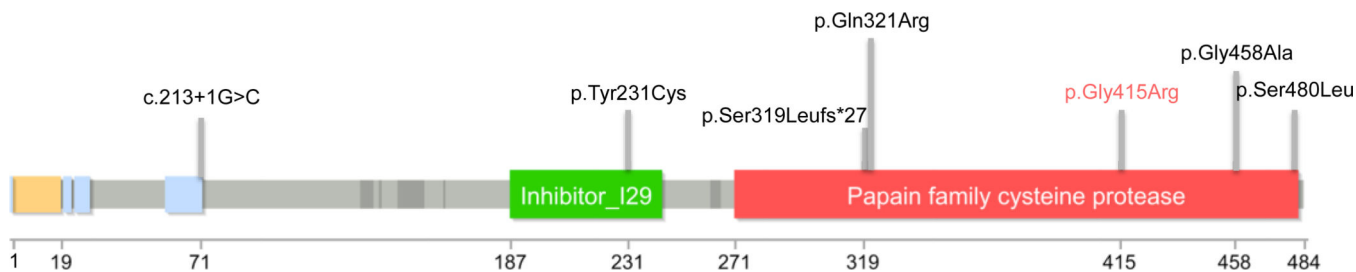
- Morbin M, Mullen S, Andermann F, Mole SE, Bahlo M, Berkovic SF. Kufs disease, the major adult form of neuronal ceroid lipofuscinosis, caused by mutations in CLN6. *Am J Hum Genet.* 2011; 88(5):566–573. [PubMed: 21549341]
- Benitez BA, Alvarado D, Cai Y, Mayo K, Chakraverty S, Norton J, Morris JC, Sands MS, Goate A, Cruchaga C. Exome-Sequencing Confirms DNAJC5 Mutations as Cause of Adult Neuronal Ceroid-Lipofuscinosis. *PLoS One.* 2011; 6(11):e26741. PONE-D-11-16499 [pii]. [PubMed: 22073189]
- Berkovic SF, Carpenter S, Andermann F, Andermann E, Wolfe LS. Kufs' disease: a critical reappraisal. *Brain.* 1988; 111(Pt 1):27–62. [PubMed: 3284607]
- Clarimon J, Djaldetti R, Lleo A, Guerreiro RJ, Molinuevo JL, Paisan-Ruiz C, Gomez-Isla T, Blesa R, Singleton A, Hardy J. Whole genome analysis in a consanguineous family with early onset Alzheimer's disease. *Neurobiology of aging.* 2009; 30(12):1986–1991. [PubMed: 18387709]
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernysky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011; 43(5):491–498. [PubMed: 21478889]
- Di Fabio R, Moro F, Pestillo L, Meschini MC, Pezzini F, Doccini S, Casali C, Pierelli F, Simonati A, Santorelli FM. Pseudo-dominant inheritance of a novel CTSF mutation associated with type B Kufs disease. *Neurology.* 2014; 83(19):1769–1770. [PubMed: 25274848]
- Di Fede G, Catania M, Morbin M, Rossi G, Suardi S, Mazzoleni G, Merlin M, Giovagnoli AR, Prioni S, Erbetta A, Falcone C, Gobbi M, Colombo L, Bastone A, Beeg M, Manzoni C, Francescucci B, Spagnoli A, Cantu L, Del Favero E, Levy E, Salmona M, Tagliavini F. A recessive mutation in the APP gene with dominant-negative effect on amyloidogenesis. *Science.* 2009; 323(5920):1473–1477. [PubMed: 19286555]
- Dolzhanskaya N, Gonzalez MA, Sperziani F, Stefl S, Messing J, Wen GY, Alexov E, Zuchner S, Velinov M. A novel p.Leu(381)Phe mutation in presenilin 1 is associated with very early onset and unusually fast progressing dementia as well as lysosomal inclusions typically seen in Kufs disease. *J Alzheimers Dis.* 2014; 39(1):23–27. [PubMed: 24121961]
- Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M. Pfam: the protein families database. *Nucleic acids research.* 2014; 42(Database issue):D222–D230. [PubMed: 24288371]
- Ghani M, Reitz C, Cheng R, Vardarajan BN, Jun G, Sato C, Naj A, Rajbhandary R, Wang LS, Valladares O, Lin CF, Larson EB, Graff-Radford NR, Evans D, De Jager PL, Crane PK, Buxbaum JD, Murrell JR, Raj T, Ertekin-Taner N, Logue M, Baldwin CT, Green RC, Barnes LL, Cantwell LB, Fallin MD, Go RC, Griffith PA, Obisesan TO, Manly JJ, Lunetta KL, Kamboh MI, Lopez OL, Bennett DA, Hendrie H, Hall KS, Goate AM, Byrd GS, Kukull WA, Foroud TM, Haines JL, Farrer LA, Pericak-Vance MA, Lee JH, Schellenberg GD, St George-Hyslop P, Mayeux R, Rogaeva E. Alzheimer's Disease Genetics, C. Association of Long Runs of Homozygosity With Alzheimer Disease Among African American Individuals. *JAMA Neurol.* 2015; 72(11):1313–1323. [PubMed: 26366463]
- Ghani M, Sato C, Lee JH, Reitz C, Moreno D, Mayeux R, St George-Hyslop P, Rogaeva E. Evidence of recessive Alzheimer disease loci in a Caribbean Hispanic data set: genome-wide survey of runs of homozygosity. *JAMA Neurol.* 2013; 70(10):1261–1267. [PubMed: 23978990]
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009; 25(14):1754–1760. [PubMed: 19451168]
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research.* 2010; 20(9):1297–1303. [PubMed: 20644199]
- McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology.* 1984; 34(7):939–944. [PubMed: 6610841]
- Nalls MA, Guerreiro RJ, Simon-Sanchez J, Bras JT, Traynor BJ, Gibbs JR, Launer L, Hardy J, Singleton AB. Extended tracts of homozygosity identify novel candidate genes associated with late-onset Alzheimer's disease. *Neurogenetics.* 2009; 10(3):183–190. [PubMed: 19271249]

- Noskova L, Stranecky V, Hartmannova H, Pristoupilova A, Baresova V, Ivanek R, Hulkova H, Jahnova H, van der Zee J, Staropoli JF, Sims KB, Tyynela J, Van Broeckhoven C, Nijssen PC, Mole SE, Elleder M, Kmoch S. Mutations in DNAJC5, encoding cysteine-string protein alpha, cause autosomal-dominant adult-onset neuronal ceroid lipofuscinosis. *Am J Hum Genet.* 2011; 89(2): 241–252. [PubMed: 21820099]
- Sherva R, Baldwin CT, Inzelberg R, Vardarajan B, Cupples LA, Lunetta K, Bowirrat A, Naj A, Pericak-Vance M, Friedland RP, Farrer LA. Identification of novel candidate genes for Alzheimer's disease by autozygosity mapping using genome wide SNP data. *J Alzheimers Dis.* 2011; 23(2): 349–359. [PubMed: 21098978]
- Sims R, Dwyer S, Harold D, Gerrish A, Hollingworth P, Chapman J, Jones N, Abraham R, Ivanov D, Pahwa JS, Moskvina V, Dowzell K, Thomas C, Stretton A, Lovestone S, Powell J, Proitsi P, Lupton MK, Brayne C, Rubinsztein DC, Gill M, Lawlor B, Lynch A, Morgan K, Brown KS, Passmore PA, Craig D, McGuinness B, Todd S, Johnston JA, Holmes C, Mann D, Smith AD, Love S, Kehoe PG, Hardy J, Mead S, Fox N, Rossor M, Collinge J, Livingston G, Bass NJ, Gurling H, McQuillin A, Jones L, Holmans PA, O'Donovan M, Owen MJ, Williams J. No evidence that extended tracts of homozygosity are associated with Alzheimer's disease. *American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics.* 2011; 156B(7):764–771.
- Smith KR, Dahl HH, Canafoglia L, Andermann E, Damiano J, Morbin M, Bruni AC, Giaccone G, Cossette P, Saftig P, Grotzinger J, Schwake M, Andermann F, Staropoli JF, Sims KB, Mole SE, Franceschetti S, Alexander NA, Cooper JD, Chapman HA, Carpenter S, Berkovic SF, Bahlo M. Cathepsin F mutations cause Type B Kufs disease, an adult-onset neuronal ceroid lipofuscinosis. *Hum Mol Genet.* 2013; 22(7):1417–1423. [PubMed: 23297359]
- Tomiyama T, Nagata T, Shimada H, Teraoka R, Fukushima A, Kanemitsu H, Takuma H, Kuwano R, Imagawa M, Ataka S, Wada Y, Yoshioka E, Nishizaki T, Watanabe Y, Mori H. A new amyloid beta variant favoring oligomerization in Alzheimer's-type dementia. *Ann Neurol.* 2008; 63(3):377–387. [PubMed: 18300294]
- Williams RE, Mole SE. New nomenclature and classification scheme for the neuronal ceroid lipofuscinoses. *Neurology.* 2012; 79(2):183–191. [PubMed: 22778232]
- Wingo TS, Lah JJ, Levey AI, Cutler DJ. Autosomal recessive causes likely in early-onset Alzheimer disease. *Arch Neurol.* 2012; 69(1):59–64. [PubMed: 21911656]





**Figure 1. Conservation alignment for the CTSF p.Gly415Arg variant**  
 The CTSF residue found to be mutated is highly conserved throughout evolution, as shown by alignment of the protein sequences of CTSF orthologues in various organisms (A), as well as in the known human members of the papain family of cysteine proteases (B). Protein alignments were performed using the Clustal Omega software. In both panels (A and B) the symbols under the aminoacid sequences represent: “\*” (asterisks) indicate positions which have a single, fully conserved residue; “:” (colons) indicate conservation between groups of strongly similar properties - scoring > 0.5 in the Gonnet PAM 250 matrix; and “.” (periods) indicate conservation between groups of weakly similar properties - scoring =< 0.5 in the Gonnet PAM 250 matrix. Gaps between the different protein sequences are represented by “-” (dashes) within the sequences. In panel A only residues 376 to 428 are represented, from a total of 484 amino acids in the human protein and corresponding region in the other organisms. In panel B only residues 410 to 446 are represented for Cathepsin F and corresponding region in the other members of the papain family of cysteine proteases. The boxes and underneath arrows indicate the residue found to be mutated in the siblings studied here (G415).



**Figure 2. Protein representation with the location of the *CTSF* mutations reported to be associated with Kufs disease**

*CTSF* is represented with the depicted regions as predicted by Pfam (UniProt Q9UBX1) (Finn, et al., 2014). The variant identified in this study is represented in red. Most mutations are located in the peptidase C1 domain towards the C-terminal end of the protein and are predicted to affect the protein function.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 1**

Variants identified by exome sequencing that map to previously determined regions of homozygosity shared between both affected siblings.

Chr	Start	End	Gene	Variants from exome sequencing	
				NM_reference	Variant
Chr 1	64,831,392	76,777,065			
	120,351,977	146,978,395			
	222,206,887	223,349,930			
Chr 2	99,769	7,554,772			
	194,236,059	195,771,061			
Chr 3	16,507,325	19,786,794	<i>TBC1D5</i>	NM_014744	c.C509T;p.T170M
	61,615,203	71,702,005			
	80,535,924	81,815,563			
Chr 4	101,669,900	115,238,487	<i>ATG3</i>	NM_001278712	c.920dupT;p.L307fs
	109,954,423	129,247,395	<i>QRFPR</i>	NM_198179	c.A770G;p.H257R
Chr 5	34,551,865	55,308,631	<i>OXC71</i>	NM_000436	c.T91G;p.S31A
	67,406,227	90,609,868			
Chr 6	11,274,139	20,983,442			
	33,181,258	34,327,790			
Chr 8	131,984,210	135,151,005			
	46,587	3,844,391			
Chr10	73,459,856	79,225,270	<i>DUSP13</i>	NM_001007272	c.G337C;p.D113H
	106,135,117	112,191,992			
Chr11	19,083,477	78,642,077	<i>OR5F1</i>	NM_003697	c.C398T;p.S133F
			<i>TNKS1BP1</i>	NM_033396	c.G600T;p.R200S
			<i>CTSF</i>	NM_003793	<b>c.G1243A;p.G415R</b>
			<i>NUMA1</i>	NM_001286561	c.A6128G;p.K2043R

Chr	ROH		Gene	Variants from exome sequencing	
	Start	End		NM_reference	Variant
Chr13	60,847,162	62,345,896			
	85,729,553	95,634,313			
Chr14	74,702,249	75,774,346			
Chr16	66,947,783	68,839,302			
Chr17	37,386,899	43,423,890	<i>MLX</i>	NM_170607	c.G263A:p.R88H
			<i>FZD2</i>	NM_001466	c.G757C:p.V253L

Chr: chromosome; Start-End: Chromosome positions (hg 19) of the beginning and end of all long regions of homozygosity (>1Mb) shared between both affected siblings.

**Table 2**

Main features of patients reported within families with *CTSF* mutations.

		Smith et al. Ku4	Smith et al. Ku10	Smith et al. Ku16	Di Fabio et al.					
		Proband	Older sister	Proband	Proband	Proband IV-3	Mother III-4	Maternal aunt III-5	Cousin V-1	Cousin IV-1
		Italy	France-Canada	Australia	Italy (Fondi)					
<b>Origin</b>		Italy	France-Canada	Australia	Italy (Fondi)					
<b>Consanguinity</b>		No	Yes	No	Yes					
<b>Gender</b>		Female	Female	Female	Female	Female	Female	Female	Female	Male
<b>AAE/AAD</b>		AAD 42	AAE ~54	AAE 40	AAE 42	AAE 56	AAE 74	AAE 45	AAE 45	AAD 67
<b>AAO</b>		20	24	35	23	25	66	21	21	48
<b>First manifestation</b>		Tremor	Focal seizures	Cognitive decline	Tonic-clonic seizures	Tonic-clonic seizures	Tonic-clonic seizures (after head trauma)	Tonic-clonic seizures	Tonic-clonic seizures	Tonic-clonic seizures
<b>Seizures</b>		Rare (from 28)	Yes	Rare (from 39)	23	25	66	21	21	48
<b>Cognitive impairment</b>		30	35	35	30	35	69	39 (after head trauma)	39 (after head trauma)	55
<b>Other features</b>		Ataxia and dysarthria; no myoclonus; no epileptiform activity	From age 35 she had progressive dementia with mood disturbance and motor features including tremor, ataxia, and extra-pyramidal rigidity with mild hyperflexia.	Mild cerebellar dysarthria and mild gait ataxia with tremor and past-pointing	Cerebellar dysarthria; transient perioral dyskinesias and segmental myoclonic jerks	Behavioral disturbances, emotional lability and intellectual deterioration	Confined to a wheelchair, aphasic and with severe dementia. Postural tremor in the upper limbs; action myoclonus and perioral dyskinesias	Rapidly progressive cognitive decline after head trauma; cerebellar dysarthria, occasional dysphasia with confabulation, ideomotor apraxia, inappropriate laughing, and urinary incontinence	Rapidly progressive cognitive decline after head trauma; cerebellar dysarthria, occasional dysphasia with confabulation, ideomotor apraxia, inappropriate laughing, and urinary incontinence	Severe dementia
<b>Mutation</b>		p.[(Gly458Ala)]; [(Ser480Leu)]	p.[(Gln321Arg)]; [(Gln321Arg)]	p.[(Tyr231Cys)]; [(Ser319Leufs*27)]	c.[213+1G>C]; [213+1G>C]	NA	c.[213+1G>C]; [213+1G>C]	c.[213+1G>C]; [213+1G>C]	c.[213+1G>C]; [213+1G>C]	NA

AAE/AAD: Age at examination/Age at death; AAO: age at onset; NA: not available

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript