

BRAIN COMMUNICATIONS

Genomic analysis of patients in a South Indian Community with autosomal dominant cortical tremor, myoclonus and epilepsy suggests a founder repeat expansion mutation in the *SAMD12* gene

Radha Mahadevan,^{1,†} Rahul C. Bhojar,^{2,†} Natarajan Viswanathan,³ Raskin Erusan Rajagopal,⁴ Bobby Essaki,¹ Varun Suroliya,^{2,5} Rachel Chelladurai,¹ Saravanan Sankaralingam,¹ Ganesan Shanmugam,⁶ Sriramakrishnan Vayanakkan,¹ Uzma Shamim,² Aradhana Mathur,² Abhinav Jain,^{2,7} Mohamed Imran,^{2,7} Mohammed Faruq,^{2,7} Vinod Scaria,^{2,3,7,†} Sridhar Sivasubbu^{2,7,2,†} and Shantaraman Kalyanaraman^{4,†}

†These authors contributed equally to this work.

Autosomal Dominant Cortical Tremor, Myoclonus and Epilepsy is a non-progressive disorder characterized by distal tremors. Autosomal Dominant Cortical Tremor, Myoclonus and Epilepsy has been reported globally with different genetic predispositions of autosomal dominant inheritance with a high degree of penetrance. In south India, Autosomal Dominant Cortical Tremor, Myoclonus and Epilepsy has been reported in a large cohort of 48 families, in which the genetic defect was not identified. This report pertains to the whole-genome analysis of four individuals followed by repeat-primed PCR for 102 patients from a familial cohort of 325 individuals. All the patients underwent extensive clinical evaluation including neuropsychological examinations. The whole-genome sequencing was done for two affected and two unaffected individuals, belonging to two different families. The whole-genome sequencing analysis revealed the repeat expansion of TTTTA and TTTC A in intron 4 of the *SAMD12* gene located on chromosome 8 in the patients affected with Autosomal Dominant Cortical Tremor, Myoclonus and Epilepsy, whereas the unaffected family members were negative for the similar expansion. Further, the repeat-primed PCR analysis of 102 patients showed the expansion of the TTTC A repeats in the intron 4 of *SAMD12* gene. All patients registered for this study belong to a single community called “Nadar” whose nativity is confined to the southern districts of India, with reported unique genetic characteristics. This is the largest and most comprehensive single report on clinically and genetically characterized Autosomal Dominant Cortical Tremor, Myoclonus and Epilepsy patients belonging to a unique ethnic group worldwide.

- 1 Department of Neurology, Tirunelveli Medical College, Tirunelveli 627011, Tamil Nadu, India
- 2 Genomics and Molecular Medicine, CSIR-Institute of Genomics and Integrative Biology (CSIR-IGIB), Delhi 110025, India
- 3 Institute of Neurology, Madras Medical College, Chennai 600003, Tamil Nadu, India
- 4 Multidisciplinary Research Unit, Tirunelveli Medical College, Tirunelveli 627011, Tamil Nadu, India
- 5 Department of Neurology, All India Institute of Medical Sciences, New Delhi 110029, India
- 6 Galaxy Hospital, Tirunelveli 627003, Tamil Nadu, India
- 7 Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, Uttar Pradesh, India

Received July 14, 2020. Revised October 15, 2020. Accepted October 25, 2020. Advance Access publication December 19, 2020

© The Author(s) (2020). Published by Oxford University Press on behalf of the Guarantors of Brain.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Correspondence to: Shantaraman Kalyanaraman, MD, Multidisciplinary Research Unit, Tirunelveli Medical College, Tirunelveli 627011, Tamil Nadu, India
E-mail: shantaraman_kal@tvmc.ac.in

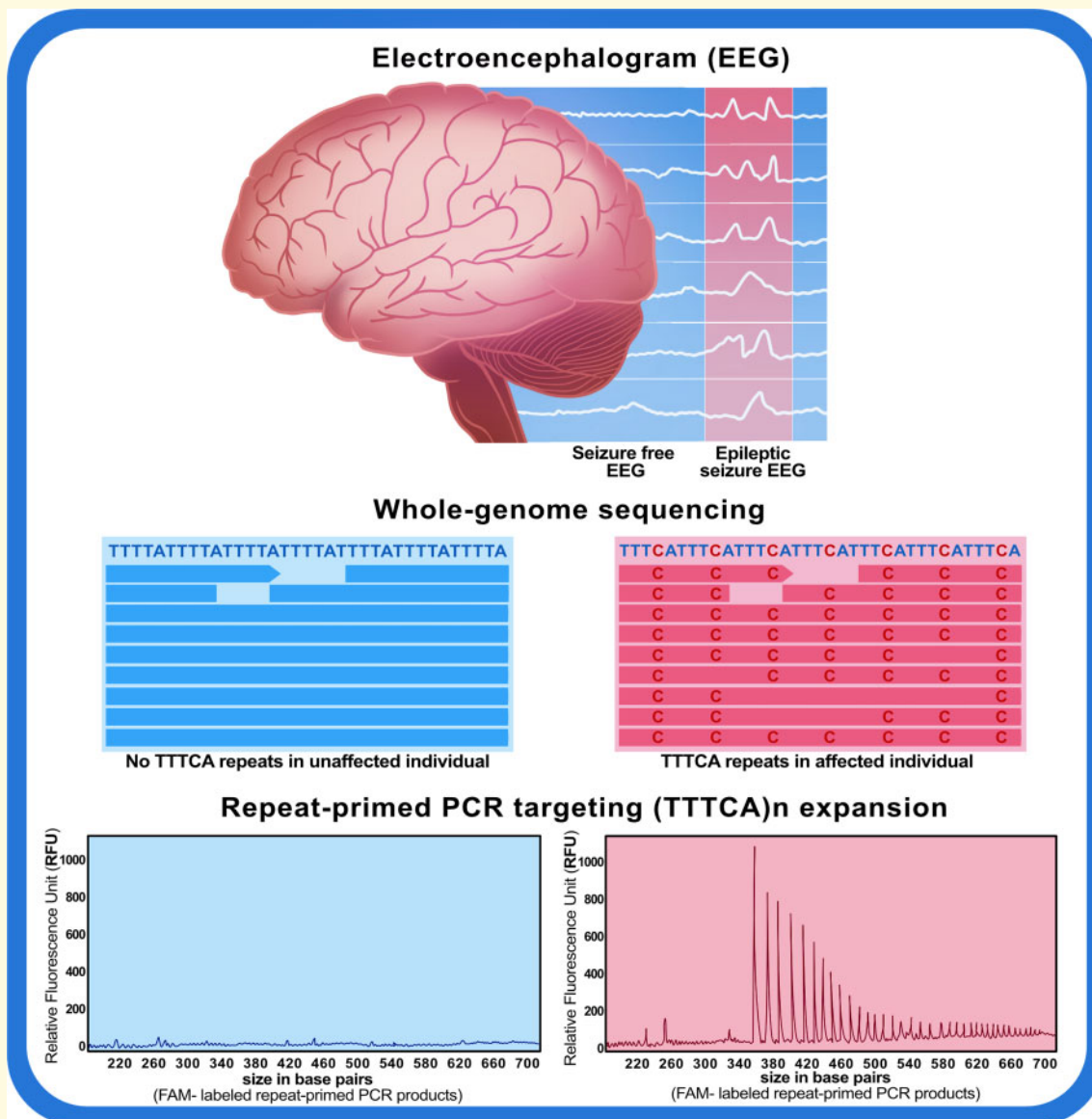
Correspondence may also be addressed to: Sridhar Sivasubbu, PhD, Genomics and Molecular Medicine, CSIR-Institute of Genomics and Integrative Biology (CSIR-IGIB), Delhi 110025, India
E-mail: s.sivasubbu@igib.res.in

Correspondence may also be addressed to: Vinod Scaria, MBBS, PhD, Genomics and Molecular Medicine, CSIR-Institute of Genomics and Integrative Biology (CSIR-IGIB), Delhi 110025, India
E-mail: vinods@igib.res.in

Keywords: autosomal dominant cortical tremor, myoclonus and epilepsy (ADCME); repeat-primed PCR; SAMD12; TTTCA repeat expansion

Abbreviations: ADCME = autosomal dominant cortical tremor, myoclonus and epilepsy; FAME = familial adult myoclonic epilepsy; SAMD12 = sterile alpha motif domain containing 12

Graphical Abstract



Introduction

Autosomal dominant cortical tremor, myoclonus and epilepsy [ADCME (MIM: 607876)] [Syn: benign adult familial myoclonus, and epilepsy; familial adult myoclonic epilepsy (FAME); familial cortical myoclonic tremor and epilepsy (FCMTE)] is a benign, non-progressive disorder reported in the age group of 12–50 years (Guerrini *et al.*, 2001; Striano and Zara, 2016). This condition has been characterized by distal action and postural tremors, arrhythmic and segmental photosensitive myoclonus predominantly in the upper limbs and seizures that are mostly generalized tonic-clonic in type. Few patients have complex partial seizures, all of which are exacerbated by anxiety, fatigue, emotional stress and sleep deprivation (Guerrini *et al.*, 2001; Striano and Zara, 2016). A higher prevalence of state-trait anxiety, depression and personality disturbances have been observed in patients with ADCME compared to the other epileptic syndromes; such serious psychiatric burden impairs their quality of life (Coppola *et al.*, 2016).

ADCME has been reported globally with different genetic predispositions of autosomal dominant inheritance with a high degree of penetrance (Coppola *et al.*, 2011). The condition has been previously reported in various ethnicities such as Japanese [8q UBR5] (Ishiura *et al.*, 2018), Italian [ADRA2B, 2q11.2] (Madia *et al.*, 2008; De Fusco *et al.*, 2014), Spanish [ACMSD, 2q21.3] (Martí-Massó *et al.*, 2013), Dutch [CTNND2, 5p15] (van Rootselaar *et al.*, 2017), French [5p15.3.1-p15.1] (Depienne *et al.*, 2010), Thai [3q26.32-q28] (Yeetong *et al.*, 2013) and Chinese [8q SLC30A8, 22q PLA2G6] (Cen *et al.*, 2015; Gao *et al.*, 2016) with pentanucleotide repeat expansions. In south India, this entity has been reported by our group in a large cohort of 48 families, with detailed clinical workup; however, the genetic defect was not identified (Mahadevan *et al.*, 2016).

Electroencephalography in patients with ADCME shows either generalized spike and wave discharges with photosensitivity or focal discharges, while the somatosensory evoked potential reported with giant cortical potentials (Ikeda *et al.*, 1990). Though these patients typically have no structural neuroimaging abnormalities, magnetic resonance spectroscopy is reported to show choline peaks in the dentate nucleus of the cerebellum (Striano *et al.*, 2009). The pathophysiology of ADCME has not been elucidated yet, however, amongst the other hypotheses, elevated cortical hyperexcitability may explain its biology. The pathogenesis of seizures may be attributed to the chronic excitation of the cerebello-thalamocortical circuits which would result in decreased cortical inhibition normally applied by the cerebellum (Striano *et al.*, 2005, 2013). These patients typically improve on treatment with Sodium Valproate and Clonazepam (Coppola *et al.*, 2011).

Studies undertaken till date for characterizing genetic variants causing epilepsies have documented repeat expansion variants in multiple genes (Ishiura *et al.*, 2018; Corbett *et al.*, 2019; Florian *et al.*, 2019; Lei *et al.*, 2019; Yeetong *et al.*, 2019). The recent series of reports on ADCME from Japanese and Chinese populations (Ishiura *et al.*, 2018; Lei *et al.*, 2019) motivated us to explore whether the genetic defect in cases of ADCME in South India could be attributed to repeat expansions. In the present study, a cohort of 66 families (325 individuals including 102 patients) affected with ADCME have been investigated. Of these, 33 families have been included from our previous report (Mahadevan *et al.*, 2016) for the genetic characterization of the disease. The present study reports the whole-genome analysis of two affected and two unaffected individuals from a familial cohort of 325 individuals. The whole-genome sequencing followed by ExpansionHunter (Dolzhenko *et al.*, 2017) and repeat-primed PCR (RP-PCR) analysis showed the insertion and repeat expansion of TTTTA and TTTCA pentanucleotide repeats in intron 4 of the sterile alpha motif domain containing 12 (*SAMD12*) gene in affected patients only. Similar repeat expansion in the *SAMD12* gene was reported in 82 patients from 48 Japanese families (Ishiura *et al.*, 2018). In the families with Benign adult familial myoclonus, and epilepsy, which were previously identified negative for the repeat expansions in *SAMD12* gene (Ishiura *et al.*, 2018), the expansions of TTTCA and TTTTA were identified in trinucleotide repeat-containing adaptor 6A [*TNRC6A* (MIM: 610739)] and Rap guanine nucleotide exchange factor 2 [*RAPGEF2* (MIM: 609530)] genes. In a recent study, an expansion of a TTTCA pentamer within the first intron of *StAR* related lipid transfer domain containing 7 [*STARD7* (MIM: 616712)] gene on chromosome 2 was also shown to be linked with the Familial Adult Myoclonic Epilepsy 2 [FAME2 (MIM: 607876)] (Corbett *et al.*, 2019). Similarly, in series of reports, an intronic TTTCA expansion was found in the membrane-associated ring-CH-type finger 6 [*MARCH6* (MIM: 613297)] and in YEATS domain containing 2 (*YEATS2* (MIM: 613373)) genes that are known to be associated with Familial Adult Myoclonic Epilepsy 3 [FAME3 (MIM: 613608)] and Familial Adult Myoclonic Epilepsy 4 [FAME4 (MIM: 615127)], respectively (Florian *et al.*, 2019; Yeetong *et al.*, 2019). In our study, we did not find such repeat expansion in *TNRC6A*, *RAPGEF2*, *MARCH6* and *YEATS2* genes in two individuals who were negative for the TTTCA pentanucleotide repeat expansion in *SAMD12* gene. Further, in order to check TTTCA pentanucleotide repeat expansion in intron 4 of *SAMD12* gene, the RP-PCR was performed for all the 102 patients enrolled in the study. RP-PCR analysis revealed the presence of TTTCA repeat expansion in intron 4 of *SAMD12* gene in all the patients investigated.

Materials and methods

Patients and clinical workup

Patients with suspected ADCME presented at the Neurology out-patient department in Tirunelveli Medical College, a tertiary care centre in the state of Tamil Nadu. Patients were taken up for genetic and genomic analysis based on the clinical and electrophysiological criteria, during a 9 year period (2009–2018) with prior institutional ethical approval. They were subsequently referred to the Multi-Disciplinary Research Unit of the medical college for genomic analysis in 2016 (Ref No 76/Neu/2011-13/05). Total 3 ml of the blood sample was drawn by venipuncture in acid citrate dextrose tubes (Becton Dickinson, NJ, USA) after a written informed consent for genome analysis was obtained. The clinical information on the onset of disease, its progression, associated aggravating factors, a detailed neurological evaluation including a mini-mental state examination and detailed lobar function examination, neuropsychological evaluations using relevant scales were documented. The social and demographic data along with detailed family history was obtained by interview and pedigree charts were drawn. Electroencephalography was recorded in all the patients and evaluated for paroxysmal abnormalities, responses to intermittent photic stimulation at frequencies between 10–25 Hz and background activity. Somatosensory evoked potential was recorded and evaluated for giant cortical potentials. Computerized tomography (CT) of the brain was performed in all patients while magnetic resonance imaging (MRI) with magnetic resonance spectroscopy (MRS) and susceptibility-weighted images (SWI) were performed in selected patients. The treatments for each patient, the responses as well as the changes in neurological and psychological status were recorded.

DNA isolation

For whole-genome sequencing and repeat-primed polymerase chain reaction (RP-PCR) screening, the genomic DNA of all reported individuals were isolated from the blood using salting out method (Miller *et al.*, 1988) and diluted to 20 ng/ μ l concentration.

Whole-genome sequencing

The 20 ng/ μ l diluted DNA samples (total 1 μ g) of four individuals from two families (B.III.19- affected, B.III.17- unaffected of family 1 represented in Fig. 1A. and D.III.2- affected, D.III.4- unaffected of family 2 represented in Fig. 1B) were fragmented by adaptive focused acoustics sonication (Covaris S220), and paired-end PCR-free libraries were prepared with TruSeq DNA PCR-free LP kit, as detailed in the standard procedure (Cat no. Illumina Inc. USA 20015962). The whole-genome libraries were further analysed and quantified with

the library quantification assay kit (KAPA Biosystems). All the libraries were pooled together at 350 pM concentration and processed for paired-end (2 \times 151 base pair reads) sequencing on Illumina NovaSeq6000 (S4 flow cell) instrument.

Whole-genome sequence analysis

The sequencing files were generated in the binary base call (bcl) format by Real-Time Analysis software and were demultiplexed into fastq files according to the sample index by the tool bcl2fastq. The raw reads underwent adapter trimming with an average Phred score of Q20 using the tool trimmomatic-0.36 that filters out the bad quality reads as well as unpaired reads (Bolger *et al.*, 2014). Filtered reads were aligned on the human reference genome GRCh38/hg38 using the Burrows-Wheeler Aligner (BWA) (Li and Durbin, 2009). After alignment, PCR duplicates were marked and removed using the java-based tool Picard (<http://broadinstitute.github.io/picard/>). GenomeAnalysisTK (GATK) v3.8 was used for variant calling (McKenna *et al.*, 2010) and postprocessing (DePristo *et al.*, 2011; Van der Auwera *et al.*, 2013). The SNVs and small indels were annotated by multiple databases such as RefGene (Pruitt *et al.*, 2007), dbsnp (dbsnp152) (Sherry, 2001), ljb26_all, ClinVar v.20180603 (Landrum *et al.*, 2018) and allele frequencies were fetched from 1000 Genomes project (A global reference for human genetic variation), ExAC (Lek *et al.*, 2016), Esp6500 (<https://evs.gs.washington.edu/EVS/>), and gnomAD (Karczewski *et al.*, 2020) using ANNOVAR (Wang *et al.*, 2010).

Short tandem repeat expansion detection using ExpansionHunter

For detecting large short tandem repeat (STR) expansions that could cause multiple diseases, a software package ExpansionHunter (v3.0.1) (<https://github.com/Illumina/ExpansionHunter>; Dolzhenko *et al.*, 2017) was used, which can identify the repeats from the PCR-free whole-genome sequencing (WGS) short-read data at the predefined genomic locus. It can detect repeats even if it is larger than the read length by using an algorithm that takes into consideration not only spanning reads but also flanking and in-repeat reads. The algorithm uses a binary alignment mapping (BAM) file as an input that contains the aligned reads from a WGS sample to the reference genome assembly FASTA file and a repeat catalogue file. The tool then performs a targeted search for the reads that span, flank, and are fully contained in-repeat and genotypes repeats at the locus of interest. In this case, BAM files of four samples aligned to GRCh38/hg38 reference genome FASTA were taken as an input for ExpansionHunter to estimate large repeat expansions. We also updated the repeat catalogue file that matches the

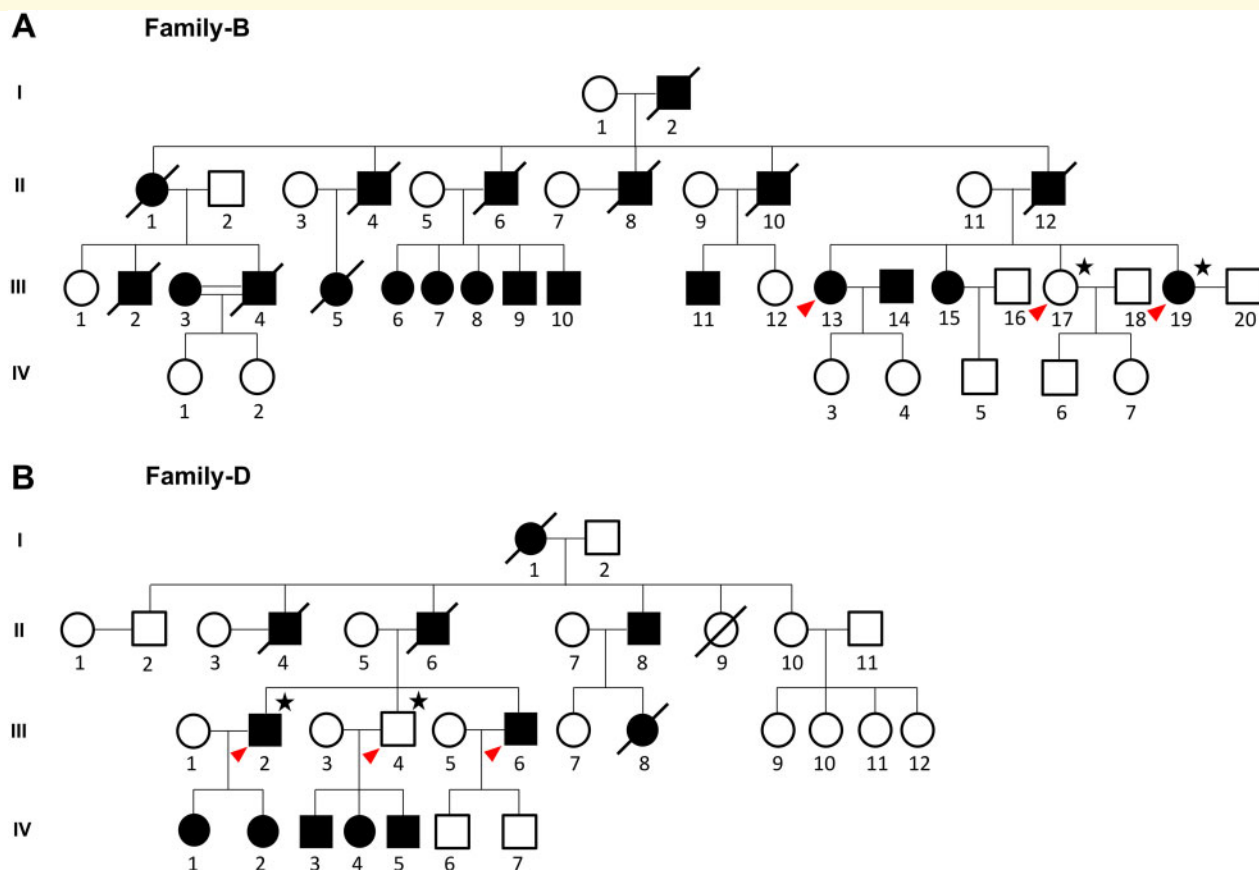


Figure 1 Pedigree tree of ADCME families subjected to WGS and RP-PCR analysis. WGS analysis was done for two individuals each from two families, **(A)** family B (B.III.19- affected and B.III.17- unaffected) and **(B)** family D (D.III.2- affected and D.III.4- unaffected), respectively. All the individuals who underwent sequencing and RP-PCR analysis are marked with a star and red arrowhead, respectively.

patient's condition to get genotype and phenotype correlations.

Repeat-primed polymerase chain reaction for the (TTTTA)_n and (TTTCA)_n repeat

The RP-PCR screening was performed for 102 patients to identify the TTTTA and TTTCA repeats in intron 4 of the *SAMD12* gene. The TTTTA repeats were amplified by RP-PCR using FAM-labelled primer (P1), repeat-containing primer (R1) and M13 anchor primer (Supplementary Table 1). The schematic representation of all the primer binding sites has been shown in Fig. 3A. PCR was performed with 20 ng genomic DNA, with LA Taq DNA polymerase (TaKaRa), Mg²⁺ free buffer (TaKaRa), 200 μM dNTP mixture (TaKaRa), 1 M betaine (Sigma-Aldrich), and MgCl₂ (TaKaRa). The touchdown PCR was carried out using the following conditions: initial denaturation for 5 min at 94°C, 10 cycles of 10 s at 98°C, 30 s at 64°C (decreasing 2°C per cycle), and 1.5 min at 72°C, followed by 27 cycles of 10 s at 98°C, 30 s at 50°C and 1.5 min at 72°C.

For TTTCA repeat expansion screening, FAM-labelled primer (P2), repeat-containing primer (R2) and M13 anchor primer (Supplementary Table 1) were used. The modified PCR parameters were used that comprised 100 ng of genomic DNA, 0.1 μl 5 U GoTaq Flexi DNA polymerase (Promega), 1 μl 5× buffer (Promega), 1 μl 2 mM dNTP mixture (Promega), 0.6 μl 25 mM MgCl₂ (Promega), 0.8 μl 10 pM labelled (P2) and M13 primers each and 0.1 μl 10 pM repeat-containing primer (R2) in a final 10 μl reaction volume. PCR was carried out using the following conditions: denaturation for 5 min at 96°C, followed by 38 cycles of 45 s at 96°C, 30 s at 60°C, and 4 min at 72°C.

PCR products were analysed on an ABI 3730 DNA Analyzer (Applied Biosystems) and visualized using GeneMapper software (version 4, Applied Biosystems).

Ancestry mapping using chromosomal painting

The TTTCA pentanucleotide expansion in the *SAMD12* gene has been earlier reported from China and Japan, which motivated us to explore the ancestry of our

patients based on the similarity in the haplotype pattern. Haplotype prediction tool fineSTRUCTURE [version 4.0.1] (Lawson *et al.*, 2012) was used, which exploits the linkage disequilibrium along with standard principal component analysis (PCA) and model-based STRUCTURE based approach for better inference of haplotype ancestry. Haplotype prediction on chromosome 8 variants for four samples (two affected and two unaffected) was performed using the chromosome 8 of the 1000 Genomes Project as a reference. This dataset is composed of 2504 individuals from five major populations, African (AFR), American (AMR), East Asian (EAS), European (EUR), and South Asian (SOU) and 26 minor populations. The whole-genome sequence corresponding to chromosome 8 from our samples was individually merged with chromosome 8 derived from the 1000 genome dataset. Further, the data were pruned to get the maximum genotype rate and phased using SHAPEIT tool (Delaneau *et al.*, 2012). Four-stage pipeline, that is a combination of chromopainter, chromocombine, and fineSTRUCTURE were used to paint each individual. Further, the combine files were assigned to population based on Markov chain Monte Carlo (MCMC) algorithm, and finally tree building was based on the maximum posterior inferred number of population. Finally, 5 MB upstream and downstream on both sides of the chromosomal locus was plotted chr8:118366817-118366914 (hg38/GRCh38) using R scripts provided by fineSTRUCTURE.

Data availability

The data from this study will be made available upon request.

Results

Clinical details

A total of 456 individuals were identified in the pedigree analysis of 66 families, including 252 males and 204 females, of whom 325 are alive (172 males and 153 females) and 131 are deceased (79 males and 52 females). The 66 families identified with ADCME were stratified based on the number of generations reporting ADCME in the pedigree. Of these 19, 34 and 12 families reported patients across the 2nd, 3rd and 4th generations, respectively. The average size of members of the pedigrees was 39.93 members with 2 largest pedigrees having 81 members. The number of affected members in each family has been summarized in Table 1. We observed few associated phenotypes in these 456 individuals, namely—infertility in 24, sudden death in 2 and mental retardation in 3. Of these 66 families, 64 belonged to the same ‘Nadar’ community while 2 belonged to a different community, including 56 males and 46 females, who were sampled for genomic analysis.

The demographic, clinical, neurophysiological, imaging and treatment data of 102 patients were recorder. The cohort age ranged from 6 to 75 years with a mean age of 43.85 years (43.49 males, 44.32 females). The mean age of onset of disease was 28.58 years (25.60 in male, 30.88 in female) and a mean duration of disease at the time of diagnosis was 17.06 years (18.45 in male, 15.51 in female) with a mean duration of follow-up of 5.26 years. The details about the age distribution and the age of the onset for 102 patients are summarized in Supplementary Table 2. The major symptom at presentation in our patients was mostly seizures (59.80%) and tremors (30.39%). Myoclonic jerks were identified in all 102 patients. The tremors in these patients were predominantly in the distal portion of the upper limb and were aggravated by posture and action, while myoclonus was stimulus sensitive aggravated by photic stimuli, sleep deprivation, fatigue or emotional stress. The seizures were generalized tonic and clonic type in 89.22% and complex partial seizure type in 10.78% cases. The seizures were precipitated by sleep deprivation, fatigue, photic and sound stimulus and emotional stress in most of these patients. Of the 102 patients studied, psychiatric comorbidities were identified in 69.61% patients, including panic disorder, phobia and generalized anxiety disorder. These patients are on continuous 3 monthly follow-up and no significant cognitive decline was detected in them. The electroencephalography recordings in these patients showed generalized spike and wave pattern with photoparoxysmal responses to intermittent photic stimulation in 63.7% of patients and the rest had focal abnormalities, while the somatosensory evoked potential showed giant cortical potentials. Of the 102 patients, CT scan of the brain was non-contributory in all, MRI scan of the brain showed no structural abnormalities, and MR spectroscopy which was done in 10 patients showed a choline peak in the proximity of the dentate nuclei and subtle mineral deposits were seen in the same region in susceptibility-weighted images (SWI). All the 102 patients received treatment with valproate and clonazepam. 74 patients had complete remission of seizures and 28 had a partial response. The response to treatment was comparatively less encouraging for tremors than the seizures.

Whole-genome sequencing

Whole-genome sequencing of four individuals generated an average of 68.1 million reads, which were filtered using trimmomatic and resulted in an average of 64.8 million reads. The alignment was more than 99% i.e., ~64.7 million reads were aligned to human reference genome GRCh38/hg38 using BWA 0.7.12 with an average coverage of 30X. The data has been tabulated in Supplementary Table 3.

Table 1 Distribution of ADCME affected individuals across the pedigrees

No. of affected members in the pedigree	No. of families	No. of members alive (male)	No. of members alive (female)	Total No. of members
1	1	0	1	1
2	3	0	4	4
3	10	10	10	20
4	3	5	3	8
5	11	18	23	41
6	8	15	19	34
7	7	20	20	40
8	3	6	12	18
9	5	20	6	26
10	2	8	4	12
11	7	35	22	57
12	4	25	11	36
20	1	6	11	17
21	1	4	7	11
	66	172	153	325

Short tandem repeat expansion detection using ExpansionHunter

In order to detect the STR expansions in the four samples (two affected and two unaffected), ExpansionHunter analysis was used with an updated repeat catalogue or json file. We found 'TTTCA' and 'TTTTA' repeats mapped at locus 8q24 in intron 4 of *SAMD12* gene encompassing chr8:118366813-118366915 (hg38/GRCh38). The pentanucleotide TTTCA repeats were present only in affected individuals, whereas the unaffected individuals were negative for the same. Similarly, TTTTA repeats were found to be highly expanded in the affected in comparison to the unaffected individuals. The average pentanucleotide repeats of TTTTA in affected individuals were two folds more than in unaffected individuals. The output variant call format (VCF) file generated by ExpansionHunter showed reads falling inside the repeat region (in-repeat reads) in the affected individuals whereas in the unaffected individuals the repeat region was spanned by the reads (spanning reads) which indicated that the number of repeats in the affected individuals were highly increased. The increase in the number of repeats has been summarized in [Supplementary Table 4](#). We have also checked the (TTTCA)_n expansion in the previously reported genes, *STARD7* at cytogenetic location 2q11.2 (chr2:96197066-96197122), *MARCH6* at 5p15.2 (chr5:10,356,347-10,356,408), *YEATS2* at 3q27.1 (chr3:183712187-183712223), *TNRC6A* at 16p12.1 (chr16:24613439-24613530) and in *RAPGEF2* gene at 4q32.1 (chr4:159342527-159342616). The Expansion Hunter and Integrative Genome Viewer (IGV) analysis targeting the above-mentioned locus were negative for the insertion and expansion of TTTCA pentanucleotide in the affected individuals ([Supplementary Fig. 1](#)).

Further, we also manually verified the paired reads from the BAM files in both affected and unaffected individuals using the IGV. We could detect 'TTTCA' repeat sequences at the genomic coordinates chr8:118366813-

118366867 (hg38/GRCh38) in intron 4 of the *SAMD12* gene in only the affected individuals ([Fig. 2](#) and [Supplementary Fig. 2](#)).

Repeat-primed PCR analysis

We performed RP-PCR analysis to screen for the TTTCA pentanucleotide repeat expansion in intron 4 of the *SAMD12* gene in 104 individuals from 66 families ([Supplementary Fig. 3](#)) including two unaffected individuals who underwent the WGS analysis ([Fig. 3](#)). The RP-PCR screening showed the presence of expansion of the TTTCA repeats in intron 4 of the *SAMD12* gene in 102 individuals, which confirmed ADCME amongst the total cohort. The length of the TTTCA repeats was found to be highly variable in the positive ADCME patients as it ranged between 7 and 80 pentanucleotide repeats. Similar variability in the TTTCA repeats was also found in the different patients from the same families. The two unaffected members did not have the TTTCA pentanucleotide repeat expansion in intron 4 of the *SAMD12* gene.

Haplotype ancestry analysis using fineSTRUCTURE

For haplotype ancestry analysis, we merged chromosome 8 variants from 2504 individuals from the 1000 Genomes Project with chromosome 8 variants from the four samples (B.III.19, B.III.17, D.III.4 and D.III.2) in our study individually, which led to 7742628, 7310484, 7319556 and 73162832 variants, respectively. We pruned the variants with low genotype call rate and retained the remaining variants. Consequently, the number of variants decreased to 185472, 184335, 184581 and 185850 for B.III.19, B.III.17, D.III.4 and D.III.2 respectively. [Figure 4](#) represents the painted 5 MB upstream and downstream regions across chr8:118366817-118366914 (hg38/GRCh38) locus. After phasing and chromosomal painting

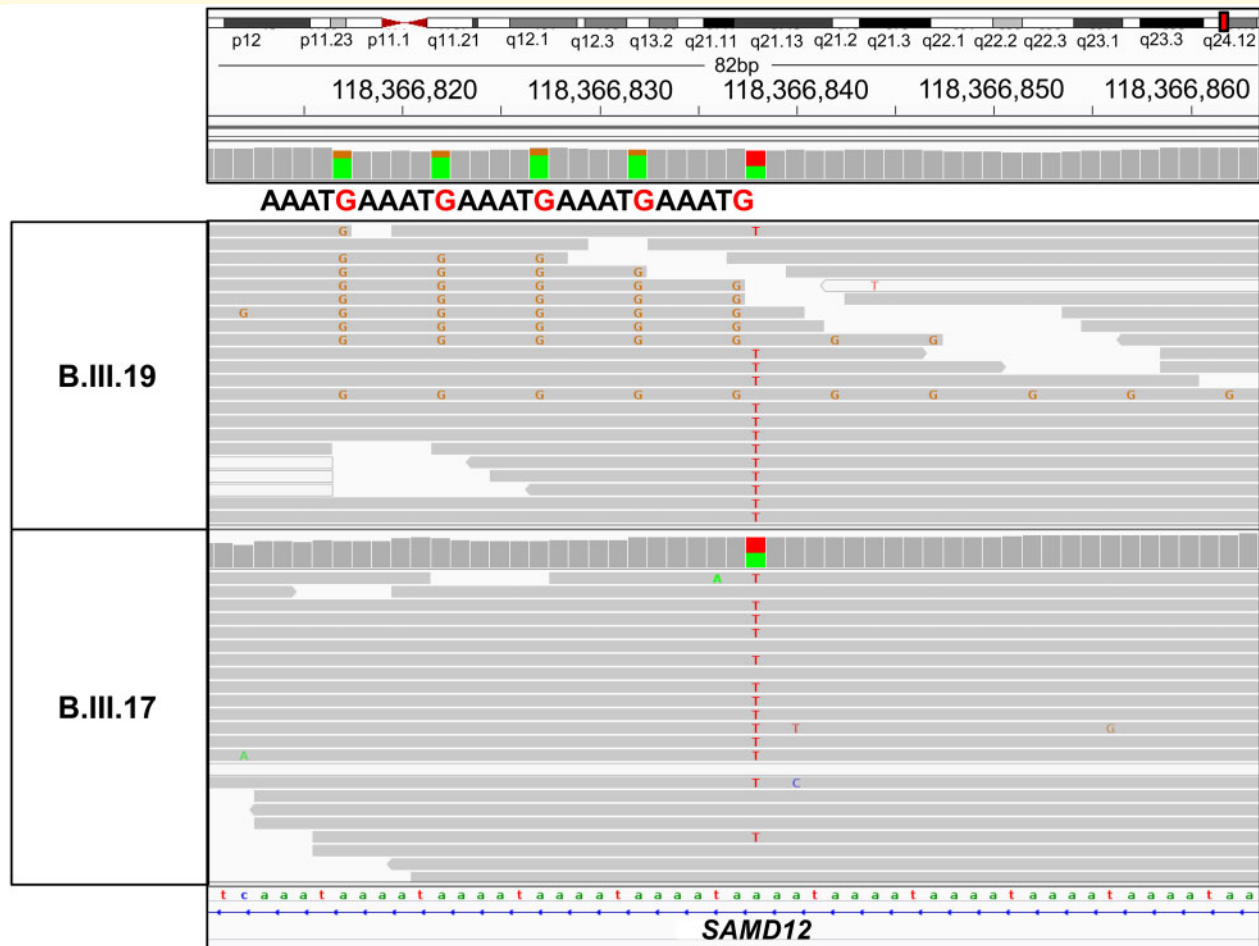


Figure 2 Visualization of *SAMD12* TTTCA intronic repeats in the Integrative Genome Viewer (IGV). WGS reads visualization in IGV revealed an intronic repeat “TTTCA” in *SAMD12* gene at coordinate chr8:118366817-118366914 (hg38/GRCh38) in the affected individual (B.III.19) whereas it was absent in the unaffected individual (B.III.17).

using fineSTRUCTURE, we found that all four samples had a South Asian ancestry.

Discussion

Autosomal dominant cortical, myoclonus and epilepsy (ADCME) is a well-delineated clinical syndrome, reported worldwide. Due to the undetermined nature of its nosology, it is not officially recognized under the International League Against Epilepsy (ILAE) classification. The prevalence of this disease is not well documented due to a lack of systematic population-based studies (Ikeda *et al.*, 1990; Coppola *et al.*, 2011; Striano and Zara, 2016). According to a report by Uyama and colleagues, the estimated prevalence of ADCME in the Japanese population was found to be approximately 1 in 35 000 (Uyama *et al.*, 2005). ADCME is unique in its presentation, varying from other idiopathic generalized epilepsies, in its age of

onset, benign clinical course, associated cortical tremors and good response to treatment.

In this study, the pedigree charts (Supplementary Fig. 3) documented the clinical evidence of an autosomal dominant transmission of the ADCME with a high degree of penetrance. In an earlier study, we had reported 48 index patients who were domiciled in two southern districts of the state of Tamil Nadu in India, with 241 affected members. The patients are presently domiciled in various urban and rural locations across Tamil Nadu. These 48 families belonged to a community called ‘Nadar’, a Tamil speaking Dravidian group domiciled predominantly in the districts of Virudhunagar, Tirunelveli, Thoothukudi and Nagercoil in Tamil Nadu (Mahadevan *et al.*, 2016). Individuals in the Nadar community have unique phenotype including shiny skin (black or brown) with sharp angulated facial features. Individuals of this group have a tradition of marrying within their community. In terms of occupation, they were regarded initially as ‘toddy-tappers’, and presently,

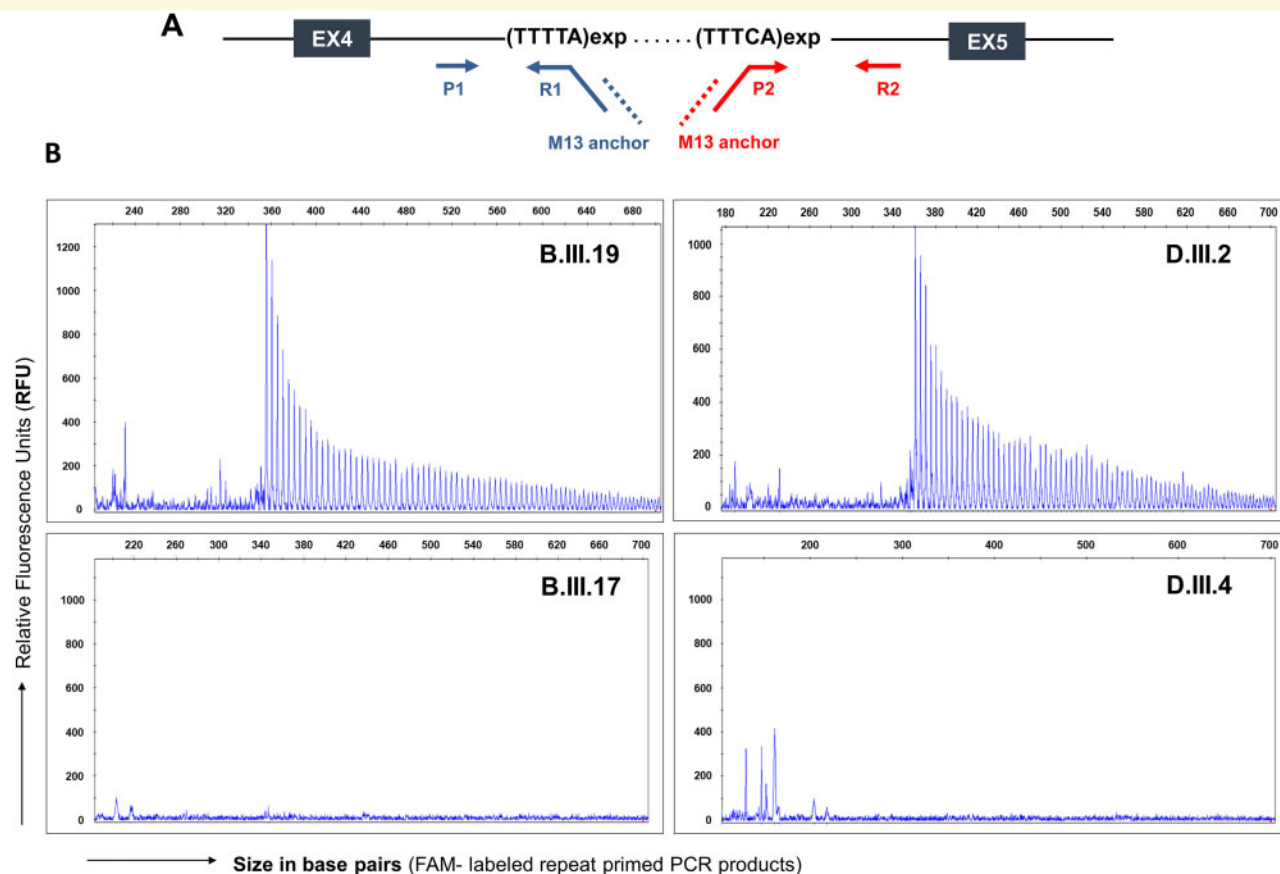


Figure 3 Identification of the TTTCA expanded repeats in families affected with ADCME. (A) Schematic representation of exons 4, 5 and intron 4 of the *SAMD12* gene. The primer sets P1, R1 and M13 anchor were designed to amplify the TTTTA repeats, whereas, P2, R2 and M13 targeted the TTTCA repeats. (B) Representative images of RP-PCR analysis targeting $(TTTCA)_n$ repeats. The RP-PCR analysis showed the TTTCA repeat expansion in ADCME affected individuals (B.III.19 and D.III.2) whereas, the TTTCA repeats were absent in unaffected individuals from both the families (B.III.17 and D.III.4).

adapted as merchants, professionals and agriculturalists. They are considered to be one of the primal colonists of South India and are believed to have their origin in a southern India known as 'Komari Land', from where the great Mediterranean culture extending to Africa, Australia and the Middle East originates (Mencher, 1972; Sridharan and Shankarkumar, 2004).

In the recent report, Ishiura and group (Ishiura *et al.*, 2018) identified the TTTCA pentanucleotide insertion and expansion in intron four of the *SAMD12* gene as the causative mutation in the Japanese FCMTE1 pedigree. The long read sequencing using Nanopore of the FCMTE1 patients showed two repeat configurations $[(TTTTA)_n (TTTCA)_n]$ and $[(TTTTA)_n (TTTCA)_n (TTTTA)_n]$. In our study, we performed a whole-genome sequencing analysis of the four individuals, two reported with ADCME and two were unaffected. This was followed by the ExpansionHunter and IGV analysis that showed TTTTA and TTTCA expansion in intron 4 of *SAMD12* in the two patients only, whereas the two unaffected individuals were negative for the TTTTA and

TTTCA expansion. Further, the RP-PCR analysis of 102 patients showed that the size of both TTTTA and TTTCA pentanucleotide repeat expansions was variable in the affected and unaffected individuals. The identification of pentanucleotide $(TTTCA)_n$ insertion in the *SAMD12* gene in patients with ADCME, replicated the findings of Ishiura and group (Ishiura *et al.*, 2018). However, in one of the unaffected individuals (D.III.4), we observed the TTTCA expansion in the *STARD7* gene. At this point, we are unable to comment on the significance of these repeats in *STARD7* gene and its contribution to epilepsy. Currently, the individual (D.III.4) is apparently healthy without any signs or symptoms of epilepsy and is under follow-up.

The molecular aetiology of the TTTCA repeat expansion in the ADCME pathogenesis has not been elucidated before. However, according to some of the hypotheses, the repeat expansion can result in RNA foci formation by RNA aggregation leading to neurotoxicity, or the translation of expanded transcript into neurotoxic peptides, or it may contribute to the loss or the

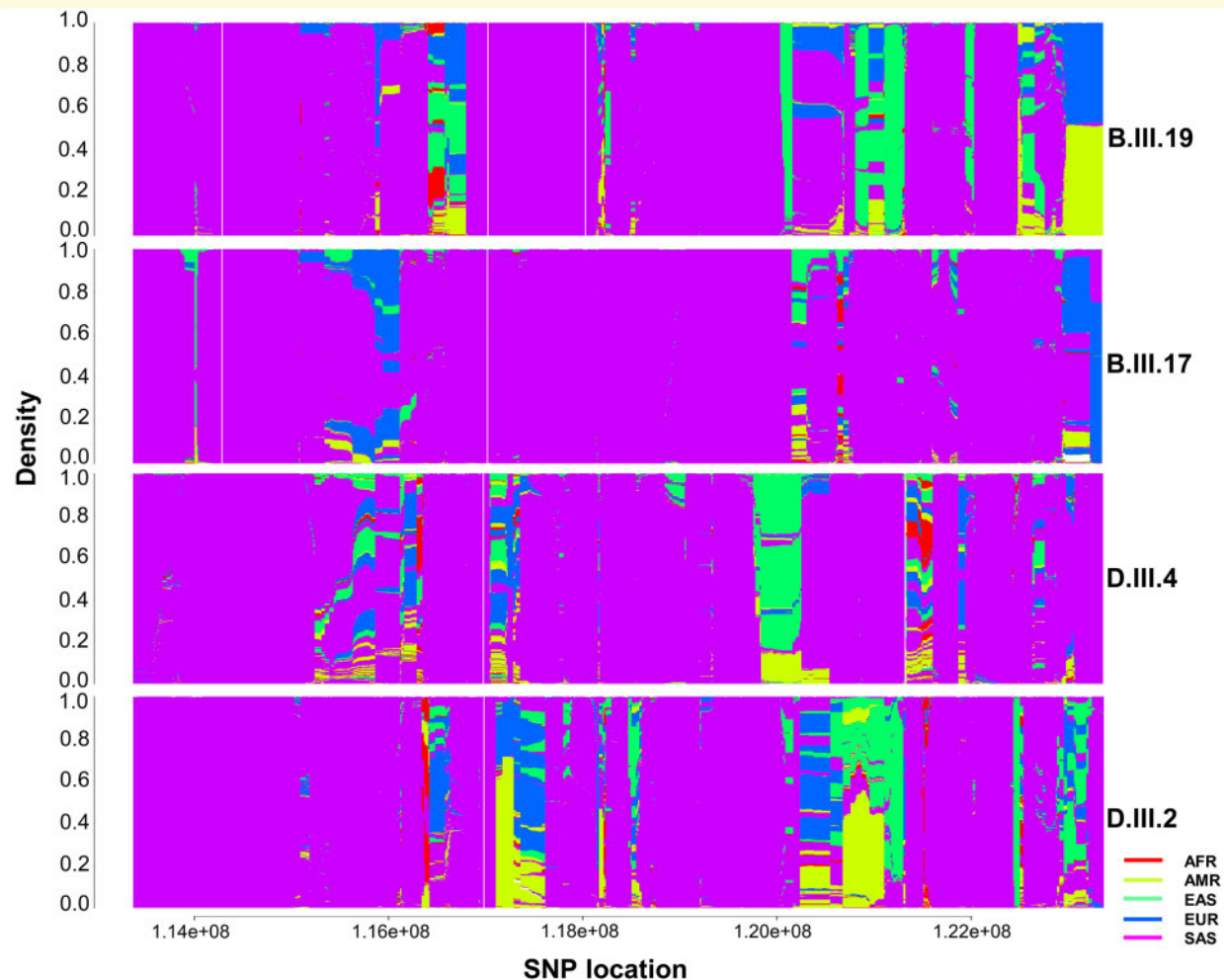


Figure 4 Haplotype ancestry prediction using fineSTRUCTURE. Ancestry prediction of two affected (B.III.19 and D.III.2) and two unaffected (B.III.17 and D.III.4) individuals was done for locus 5MB upstream and 5MB downstream to chr8:118366817-118366914 (hg38/GRCh38) using variant of 2504 individuals from five major population (AFR-African, AMR- American, EAS- East Asian, EUR-European, and SAS-South Asian) of 1000 Genomes Project.

overexpression of the linked genes (Gatchel and Zoghbi, 2005; Todd and Paulson, 2010; Loureiro et al., 2016; Zhang and Ashizawa, 2017). However, in the report by Ishiura and group, the RNA foci were observed in the cortical neurons and Purkinje cells in the brain of patients using a Cy3-(TGAAA)₁₂ probe, which strongly supports the hypothesis that RNA-mediated toxicity could be the underlying mechanism in the pathogenesis of ADCME (Ishiura et al., 2018).

A recent study in five Chinese families showed that the size of the TTTCA expansion is highly unstable over generations (Lei et al., 2019) indicating intergenerational instability. In FAME3 the TTTCA repeat length in *MARCH6* and the age of onset has been found to be inversely associated, a larger size of the expansion leads to the earlier manifestation of the dominant trait (Florian et al., 2019). Our ExpansionHunter analysis estimated the TTTCA repeat size to be 70 and 64 for two affected

individuals. Previous reports from Ishiura et al. (2018) found 458 and 225 TTTCA repeat units in two affected individuals. Similarly, Cen et al. (2018) estimated the size of the TTTCA repeats to be larger than 105 units by RP-PCR method. Our RP-PCR results for TTTCA repeat expansion in 102 ADCME affected individuals have shown that the length of the pentanucleotide expansion varied greatly between the different patients in a range of 7–80 repeat units. The number of repeat units was found to be variable even among the patients from the same families and did not show any genetic anticipation.

The Nadars have been documented to express unique haplotypes such as A3-B44, A24-B51, B44-Cw1 and DR6-DQ1 compared to other Indian populations (Sridharan and Shankarkumar, 2004). This suggests that the Nadars have been influenced by genetic drift caused by genetic selection, demographic and culture characteristics and also indicate a lower proportion of racial

admixture (Sridharan and Shankarkumar, 2004). Our haplotype ancestry analysis showed that all the four sequenced individuals had South Asian ancestry. In this study, we have identified the pentanucleotide (TTTCA)_n insertional expansion in intron 4 of the *SAMD12* gene as a causal genetic defect underlying ADCME in Nadar community. The report describes racial clustering of this rare disease in a unique ethnic group and is the largest report till date on ADCME.

Identification of the gene and the causative mutation for ADCME opens the opportunity to explore the origin and the prevalence of the disease across the population. Short-read next-generation sequencing (NGS) data, such as those generated by the Illumina sequencing platform coupled with the bioinformatics tools are currently predominant in the discovery of the STRs in the genome. However, for population-scale screening, the existing WGS techniques are not cost-effective. The current RP-PCR assay targeting the TTTTA/TTTCA repeats in the intronic region of the *SAMD12* gene showed 100% concordance with those obtained by WGS analysis. RP-PCR allowed us to identify the presence of expanded alleles or to indicate their absence in the given samples. As discussed, in the ‘Nadar’ community, there may be additional undiagnosed individuals with the spectrum of ADCME whose genetic cause may be due to the TTTCA pentanucleotide expansion. The RP-PCR can serve as a sensitive and cost-effective technique to do community level genetic screening. Screening for such repeat expansions associated with epileptic disorders across diverse communities in India will provide further understanding of the molecular basis of the disease and help in bridging the gap for developing therapeutic strategies (Batra *et al.*, 2017; Zain and Smith, 2019).

Supplementary material

Supplementary material is available at *Brain Communications* online.

Acknowledgements

We are indebted to all the patients and family members for their generous participation in this study. We would like to acknowledge Mr. Ankit Verma for his inputs in whole-genome sequencing experiments, Mr. Vigneshwar Senthivel for creating the graphical abstract and Ms. Anjali Bajaj for the editorial assistance.

Funding

This study was supported by (1) Department of Health Research, Government of India, under the Multi-Disciplinary Research Unit, Tirunelveli Medical College, Tirunelveli and (2) Council of Scientific and Industrial Research (CSIR,

India) funded RareGen (MLP1801) and GOMED (MLP1601) projects.

Competing interests

The authors report no competing interests.

References

- 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM *et al.* A global reference for human genetic variation. *Nature* 2015; 526: 68–74.
- Batra R, Nelles D, Pirie E, Blue S, Marina R, Wang H, *et al.* Elimination of toxic microsatellite repeat expansion RNA by RNA-targeting Cas9. *Cell* 2017; 170: 899–912.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014; 30: 2114–20.
- Cen Z, Jiang Z, Chen Y, Zheng X, Xie F, Yang X, *et al.* Intronic pentanucleotide TTTCA repeat insertion in the *SAMD12* gene causes familial cortical myoclonic tremor with epilepsy type 1. *Brain* 2018; 141: 2280–8.
- Cen Z-D, Xie F, Lou D-N, Lu X-J, Ouyang Z-Y, Liu L, *et al.* Fine mapping and whole-exome sequencing of a familial cortical myoclonic tremor with epilepsy family. *Am J Med Genet B Genet* 2015; 168: 595–9.
- Coppola A, Caccavale C, Santulli L, Balestrini S, Cagnetti C, Licchetta L, *et al.* Psychiatric comorbidities in patients from seven families with autosomal dominant cortical tremor, myoclonus, and epilepsy. *Epilepsy Behav* 2016; 56: 38–43.
- Coppola A, Santulli L, Del Gaudio L, Minetti C, Striano S, Zara F, *et al.* Natural history and long-term evolution in families with autosomal dominant cortical tremor, myoclonus, and epilepsy. *Epilepsia* 2011; 52: 1245–50.
- Corbett MA, Kroes T, Veneziano L, Bennett MF, Florian R, Schneider AL, *et al.* Intronic ATTTC repeat expansions in *STARD7* in familial adult myoclonic epilepsy linked to chromosome 2. *Nat Commun* 2019; 10: 4920.
- De Fusco M, Vago R, Striano P, Di Bonaventura C, Zara F, Mei D, *et al.* The $\alpha 2B$ -adrenergic receptor is mutant in cortical myoclonus and epilepsy. *Ann Neurol* 2014; 75: 77–87.
- Delaneau O, Marchini J, Zagury J-F. A linear complexity phasing method for thousands of genomes. *Nat Methods* 2012; 9: 179–81.
- Depienne C, Magnin E, Bouteiller D, Stevanin G, Saint-Martin C, Vidailhet M, *et al.* Familial cortical myoclonic tremor with epilepsy: the third locus (FCMTE3) maps to 5p. *Neurology* 2010; 74: 2000–3.
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011; 43: 491–8.
- Dolzhenko E, van Vugt JJFA, Shaw RJ, Bekritsky MA, van Blitterswijk M, Narzisi G, *et al.* The US–Venezuela Collaborative Research Group. Detection of long repeat expansions from PCR-free whole-genome sequence data. *Genome Res* 2017; 27: 1895–903.
- Exome Variant Server [Internet]. Evs.gs.washington.edu 2020. <https://evs.gs.washington.edu/EVS/> (9 April 2020, date last accessed).
- Florian R, Kraft F, Leitão E, Kaya S, Klebe S, Magnin E, *et al.*; FAME Consortium. Unstable TTTTA/TTTCA expansions in *MARCH6* are associated with Familial Adult Myoclonic Epilepsy type 3. *Nat Commun* 2019; 10.
- Gao L, Li L, Ye J, Zhu X, Shen N, Zhang X, *et al.* Identification of a novel mutation in *PLA2G6* gene in a Chinese pedigree with familial cortical myoclonic tremor with epilepsy. *Seizure* 2016; 41: 81–5.
- Gatchel JR, Zoghbi HY. Diseases of unstable repeat expansion: mechanisms and common principles. *Nat Rev Genet* 2005; 6: 743–55.

- Guerrini R, Bonanni P, Patrignani A, Brown P, Parmeggiani L, Grosse P, et al. Autosomal dominant cortical myoclonus and epilepsy (ADCME) with complex partial and generalized seizures: a newly recognized epilepsy syndrome with linkage to chromosome 2p11.1-q12.2. *Brain* 2001; 124: 2459–75.
- Ikeda A, Kakigi R, Funai N, Neshige R, Kuroda Y, Shibasaki H. Cortical tremor: a variant of cortical reflex myoclonus. *Neurology* 1990; 40: 1561–5.
- ILLUMINA/ExpansionHunter [Internet]. GitHub 2020. <https://github.com/illumina/ExpansionHunter> (9 April 2020, date last accessed).
- Ishiura H, Doi K, Mitsui J, Yoshimura J, Matsukawa MK, Fujiyama A, et al. Expansions of intronic TTTCA and TTTTA repeats in benign adult familial myoclonic epilepsy. *Nat Genet* 2018; 50: 581–90.
- Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020; 581: 434–443.
- Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* 2018; 46: D1062–D1067.
- Lawson D, Hellenthal G, Myers S, Falush D. Inference of population structure using dense haplotype data. *PLoS Genet* 2012; 8: e1002453.
- Lei XX, Liu Q, Lu Q, Huang Y, Zhou XQ, Sun HY, et al. TTTCA repeat expansion causes familial cortical myoclonic tremor with epilepsy. *Eur J Neurol* 2019; 26: 513–8.
- Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016; 536: 285–91.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; 25: 1754–60.
- Loureiro JR, Oliveira CL, Silveira I. Unstable repeat expansions in neurodegenerative diseases: nucleocytoplasmic transport emerges on the scene. *Neurobiol. Aging* 2016; 39: 174–83.
- Madia F, Striano P, Di Bonaventura C, de Falco A, de Falco FA, Manfredi M, et al. Benign adult familial myoclonic epilepsy (BAFME): evidence of an extended founder haplotype on chromosome 2p11.1-q12.2 in five Italian families. *Neurogenetics* 2008; 9: 139–42.
- Mahadevan R, Viswanathan N, Shanmugam G, Sankaralingam S, Essaki B, Chelladurai R. Autosomal dominant cortical tremor, myoclonus, and epilepsy (ADCME) in a unique south Indian community. *Epilepsia* 2016; 57: e56–e59.
- Martí-Massó J, Bergareche A, Makarov V, Ruiz-Martínez J, Gorostidi A, de Munain A, et al. The *ACMSD* gene, involved in tryptophan metabolism, is mutated in a family with cortical myoclonus, epilepsy, and parkinsonism. *J Mol Med* 2013; 91: 1399–406.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; 20: 1297–303.
- Mencher J. The Nadars of Tamilnad: the political culture of a community in change. *Am Anthropologist* 1972; 74: 51–3.
- Miller S, Dykes D, Polesky H. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 1988; 16: 1215.
- Picard Tools - By Broad Institute [Internet]. Broadinstitute.github.io 2020. <http://broadinstitute.github.io/picard/> (9 April 2020, date last accessed).
- Pruitt K, Tatusova T, Maglott D. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 2007; 35: D61–D65.
- Sherry S. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001; 29: 308–311.
- Sridharan B, Shankarkumar U. HLA antigens in Nadars a Native Dravidian Caste Group of Tamil Nadu, South India. *Int J Hum Genetics* 2004; 4: 119–124.
- Striano P, Caranci F, Di Benedetto R, Tortora F, Zara F, Striano S. MR spectroscopy indicates prominent cerebellar dysfunction in benign adult familial myoclonic epilepsy. *Epilepsia* 2009; 50: 1491–1497.
- Striano P, Louis ED, Manto M. Autosomal dominant cortical tremor, myoclonus, and epilepsy: is the origin in the cerebellum? *Cerebellum* 2013; 12: 145–146.
- Striano P, Zara F, Striano S. Autosomal dominant cortical tremor, myoclonus and epilepsy: many syndromes, one phenotype. *Acta Neurol Scand* 2005; 111: 211–217.
- Striano P, Zara F. Autosomal dominant cortical tremor, myoclonus and epilepsy. *Epileptic Disorders* 2016; 18: 139–144.
- Todd PK, Paulson HL. RNA-mediated neurodegeneration in repeat expansion disorders. *Ann Neurol* 2010; 67: 291–300.
- Uyama E, Fu Y-H, Ptáček LJ. Familial adult myoclonic epilepsy (FAME). *Adv Neurol* 2005; 95: 281–288.
- van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 2013; 43: 11.10.1–11.10.33.
- van Rootselaar A, Groffen A, de Vries B, Callenbach P, Santen G, Koelewijn S, et al. δ -Catenin (CTNND2) missense mutation in familial cortical myoclonic tremor and epilepsy. *Neurology* 2017; 89: 2341–2350.
- Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010; 38: e164–e164.
- Yeetong P, Ausavarat S, Bhidayasiri R, Piravej K, Pasutharnchat N, Desudchit T, et al. A newly identified locus for benign adult familial myoclonic epilepsy on chromosome 3q26.32-3q28. *Eur J Hum Genet* 2013; 21: 225–228.
- Yeetong P, Pongpanich M, Srichomthong C, Assawapitaksakul A, Shotelersuk V, Tantirukdham N, et al. TTTCA repeat insertions in an intron of *YEATS2* in benign adult familial myoclonic epilepsy type 4. *Brain* 2019; 142: 3360–3366.
- Zain R, Smith CIE. Targeted oligonucleotides for treating neurodegenerative tandem repeat diseases. *Neurotherapeutics* 2019; 16: 248–262.
- Zhang N, Ashizawa T. RNA toxicity and foci formation in microsatellite expansion diseases. *Curr Opin Genet Dev* 2017; 44: 17–29.