ARTICLE

Progressive myoclonus epilepsies—Residual unsolved cases have marked genetic heterogeneity including dolichol-dependent protein glycosylation pathway genes

Carolina Courage,[1,2,37] Karen L. Oliver,[3,4,5,37] Eon Joo Park,[6] Jillian M. Cameron,[3] Kariona A. Grabińska,[6] Mikko Muona,[1,7] Laura Canafoglia,[8] Antonio Gambardella,[9] Edith Said,[10,11] Zaid Afawi,[12] Betul Baykan,[13] Christian Brandt,[14] Carlo di Bonaventura,[15] Hui Bein Chew,[16] Chiara Criscuolo,[17] Leanne M. Dibbens,[18] Barbara Castellotti,[19] Patrizia Riguzzi,[20] Angelo Labate,[9] Alessandro Filla,[17] Anna T. Giallonardo,[21] Geza Berecki,[22] Christopher B. Jackson,[23] Tarja Joensuu,[1] John A. Damiano,[3] Sara Kivity,[24] Amos Korczyn,[25] Aarno Palotie,[26,27,28] Pasquale Striano,[29] Davide Uccellini,[30] Loretta Giuliano,[31] Eva Andermann,[32,33] Ingrid E. Scheffer,[3,34,35,36] Roberto Michelucci,[20] Melanie Bahlo,[4,5] Silvana Franceschetti,[8] William C. Sessa,[6] Samuel F. Berkovic,[3,38,*] and Anna-Elina Lehesjoki[1,2,38,*]

Summary

Progressive myoclonus epilepsies (PMEs) comprise a group of clinically and genetically heterogeneous rare diseases. Over 70% of PME cases can now be molecularly solved. Known PME genes encode a variety of proteins, many involved in lysosomal and endosomal function. We performed whole-exome sequencing (WES) in 84 (78 unrelated) unsolved PME-affected individuals, with or without additional family members, to discover novel causes. We identified likely disease-causing variants in 24 out of 78 (31%) unrelated individuals, despite previous genetic analyses. The diagnostic yield was significantly higher for individuals studied as trios or families (14/28) versus singletons (10/50) (OR = 3.9, p value = 0.01, Fisher's exact test). The 24 likely solved cases of PME involved 18 genes. First, we found and functionally validated five heterozygous variants in *NUS1* and *DHDDS* and a homozygous variant in *ALG10*, with no previous disease associations. All three genes are involved in dolichol-dependent protein glycosylation, a pathway not previously implicated in PME. Second, we independently validate *SEMA6B* as a dominant PME gene in two unrelated individuals. Third, in five families, we identified variants in established PME genes; three with intronic or copy-number changes (*CLN6*, *GBA*, *NEU1*) and two very rare causes (*ASAH1*, *CERS1*). Fourth, we found a group of genes usually associated with developmental and epileptic encephalopathies, but here, remarkably, presenting as PME, with or without prior developmental delay. Our systematic analysis of these cases suggests that the small residuum of unsolved cases will most likely be a collection of very rare, genetically heterogeneous etiologies.

[1]Folkhälsan Research Center, Helsinki 00290, Finland; [2]Department of Medical and Clinical Genetics, Medicum, University of Helsinki, Helsinki 00290, Finland; [3]Epilepsy Research Centre, Department of Medicine, University of Melbourne, Austin Health, Heidelberg 3084, Victoria, Australia; [4]Population Health and Immunity Division, the Walter and Eliza Hall Institute of Medical Research, Parkville, VIC 3052, Australia; [5]Department of Medical Biology, the University of Melbourne, Melbourne, VIC 3010, Australia; [6]Department of Pharmacology and Vascular Biology and Therapeutics Program, Yale University School of Medicine, 10 Amistad Street, New Haven, CT 06520, USA; [7]Blueprint Genetics, Espoo 02150, Finland; [8]Neurophysiopathology, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan 20133, Italy; [9]Institute of Neurology, University Magna Græcia, Catanzaro 88100, Italy; [10]Section of Medical Genetics, Mater dei Hospital, Msida MSD2090, Malta; [11]Department of Anatomy and Cell Biology, University of Malta, Msida MSD2090, Malta; [12]Center for Neuroscience, Ben-Gurion University of the Negev, Be'er Sheva 8410402, Israel; [13]Departments of Neurology and Clinical Neurophysiology, Istanbul Faculty of Medicine, Istanbul University, Istanbul 34452, Turkey; [14]Epilepsy Center Bethel, Bielefeld 33617, Germany; [15]Department of Human Neurosciences, Sapienza University of Rome, Viale dell'Università, 30, 00185 Rome, Italy; [16]Genetics Department, Kuala Lumpur Hospital, Ministry of Health Malaysia, Jalan Pahang, 50586 Kuala Lumpur, Malaysia; [17]Department of Neuroscience, Reproductive, and Odontostomatological Sciences, University of Naples Federico II, Naples 80138, Italy; [18]Epilepsy Research Group, Australian Centre for Precision Health, UniSA Clinical and Health Sciences, University of South Australia, Adelaide, SA 5000, Australia; [19]Unit of Genetics of Neurodegenerative and Metabolic Diseases, IRCCS Istituto Neurologico Carlo Besta Milan 20133, Italy; [20]IRCCS Istituto delle Scienze Neurologiche di Bologna, Unit of Neurology, Bellaria Hospital, Bologna 40139, Italy; [21]Neurology Unit, Human Neurosciences Department, Sapienza University, Rome 00185, Italy; [22]Ion Channels and Disease Group, Florey Institute of Neuroscience and Mental Health, University of Melbourne, Parkville, VIC 3052, Australia; [23]Stem Cells and Metabolism Research Program, Faculty of Medicine, University of Helsinki, 00290 Helsinki, Finland; [24]Epilepsy Unit, Schneider Children's Medical Center of Israel, Petah Tiqvah 4922297, Israel; [25]Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 60198, Israel; [26]Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki 00290, Finland; [27]Analytic and Translational Genetics Unit, Department of Medicine, Department of Neurology and Department of Psychiatry Massachusetts General Hospital, Boston, MA 02114, USA; [28]The Stanley Center for Psychiatric Research and Program in Medical and Population Genetics, The Broad Institute of MIT and Harvard, Cambridge, Boston, MA 02142, USA; [29]Pediatric Neurology and Muscular Diseases Unit, IRCCS Istituto "G. Gaslini," Genova 16147, Italy; [30]Neurology - Neurophysiology Unit, ASST dei Sette Laghi, Galmarini Tradate Hospital, Tradate 21049, Italy; [31]Dipartimento "G.F. Ingrassia," Università degli Studi di Catania, Catania 95131, Italy; [32]Neurogenetics Unit and Epilepsy Research Group, Montreal Neurological Hospital and Institute, Montreal, QC H3A 2B4, Canada; [33]Departments of Neurology & Neurosurgery and Human Genetics, McGill University, Montreal, QC H3A 0G4, Canada; [34]Murdoch Children's Research Institute, Royal Children's Hospital, Parkville, VIC 3052, Australia; [35]Department of Paediatrics, The University of Melbourne, Royal Children's Hospital, Parkville, VIC 3052, Australia; [36]The Florey Institute, Parkville, VIC 3052, Australia
[37]These authors contributed equally
[38]These authors contributed equally
*Correspondence: s.berkovic@unimelb.edu.au (S.F.B.), anna-elina.lehesjoki@helsinki.fi (A.-E.L.)
https://doi.org/10.1016/j.ajhg.2021.03.013.

Introduction

The progressive myoclonus epilepsies (PMEs) are a group of rare clinically and genetically heterogeneous disorders that typically present in childhood or adolescence with action myoclonus, generalized tonic-clonic seizures, and progressive neurological decline.[1] The majority of PMEs follow autosomal-recessive inheritance, with rare mitochondrial causes and a small but increasing number of autosomal-dominant genes.[2,3]

Clinically, the PMEs can be categorized into two broad groups. In one group, cognition is largely preserved with clinical features dominated by severe, treatment-resistant, and physically disabling myoclonus, tonic-clonic seizures, and ataxia.[1] The most common and paradigmatic form is Unverricht-Lundborg disease (ULD, EPM1 [MIM: 254800]), which is caused by recessive mutation, most commonly a dodecamer repeat expansion, of cystatin B (*CSTB* [MIM: 601145]). The second clinical group is associated with significant cognitive impairment and decline, with the major forms including Lafora disease (EPM2A/B [MIM: 254780]) and the neuronal ceroid lipofuscinoses (NCLs) which involve a number of mostly recessive genes.

Known PME genes encode a variety of proteins, many of which have an endosomal and lysosomal function (Table S1). Despite this, there is no apparent unifying pathway leading to the phenotype.[4,5] Importantly a molecular genetic diagnosis will currently be made with an established PME gene in approximately 70% of all individuals diagnosed with PME.[2]

We previously performed a whole-exome sequencing (WES) study on a cohort of 84 molecularly unsolved and unrelated singleton cases of PME. We identified a recurrent pathogenic variant in *KCNC1* (MIM: 176258) (p.Arg320His [c.959G>A]) as a cause of PME now known as MEAK (myoclonus epilepsy and ataxia due to K+ channel mutation, EPM7 [MIM: 616187]).[3,6] This heterozygous variant not only added another autosomal-dominant gene to the list of known PME genes, but also highlighted a role for *de novo* pathogenic variants in PME.

In this study, we aimed to identify further causative genes for the unsolved PMEs by expanding our WES data analysis to include additional unsolved cases and using, where possible, a trio-design approach to enhance the detection of *de novo* pathogenic variants.

Subjects and methods

Subjects

We studied a total of 84 (78 unrelated) molecularly unsolved individuals (45 males) who had been clinically diagnosed with PME. Individuals were referred for genetic research from centers in Europe, Australia, and the USA over 25 years. Informed consent for DNA analysis was obtained from individuals in line with local institutional review board requirements at the time of collection.

The majority of the cohort had previously had extensive genetic investigations, including clinical microarray and gene panel analyses (including mitochondrial gene testing where suspected), or singleton WES as part of our earlier research (n = 57).[3] Specifically, all individuals had been screened and tested negative for recessive variants in *CSTB*, including the dodecamer repeat expansion, and for the *KCNC1* recurrent pathogenic p.Arg320His variant.

A trio-design approach was used for 22/78 unrelated individuals (28%) where DNA was available for both unaffected parents for WES. Six unrelated subjects were exome sequenced with an affected first-degree relative (parent-child, n = 2; or sibling pairs, n = 4); two of the four sibling pairs had both unaffected parents available for WES and were analyzed as a quartet. The remaining 50 affected individuals were analyzed as WES singletons (Figure S1).

Exome sequencing

This study included two sequencing cohorts (Figure S1). The first cohort comprised 57 singleton individuals with PME that remained unsolved after our initial study;[3] of these, 44 did not have parental DNA available for trio- or quartet- WES re-analysis. This cohort was exome sequenced previously at the Wellcome Sanger Institute, Cambridge, UK in 2011-2012 (details of sequencing described in Muona et al.[3]).

The second cohort comprised a total of 40 individuals with PME and 48 unaffected parents (contributing to 22 trios and 2 quartets). 27 PME-affected individuals in cohort 2 were newly referred; 13 were subjects from cohort 1 that were re-sequenced with their parents. Exome sequencing for this cohort was performed at the Broad Institute of MIT and Harvard, Cambridge, MA, USA in 2015-2016. In detail, genomic DNA (approximately 1 μg) extracted from peripheral blood for each sample was enzymatically sheared in whole-exome library preparation. In-solution hybrid exome capture was performed using the Illumina Rapid Capture Exome enrichment kit with 38 Mb target region (29 Mb baited), which includes 98.3% of the intervals in the RefSeq exome database. Sequencing was performed on either Illumina HiSeq 4000 or HiSeq X instrument with the use of 151bp paired-end reads. The mean average sequencing depth for each sample was 78-fold, with more than 80% of target bases having at least 30-fold coverage. Mitochondrial DNA (mtDNA) was not targeted in either sequencing cohort.

Variant calling

Sequence reads were processed and aligned to human genome hg19/GRCh37 as described previously.[3] Variant calling of single-nucleotide variants and indels was done by GATK (v.3.7-0) HaplotypeCaller using joint calling approach. Thirteen individuals underwent WES twice (in the previous study and here), so their sequence data were merged to maximize coverage. Variant quality scores were recalibrated jointly with GATK VariantRecalibrator. A truth sensitivity cutoff of 99.8% was used for both SNVs and indels. *De novo* variants were called by GATK GenotypeRefinement and GATK PossibleDeNovo tools.

Sex and pedigree checks

Sex and ancestry checks for all samples and relatedness checks between all sample pairs were estimated using Peddy.[7] Inbreeding coefficients for all samples were estimated using FEstim.[8]

Variant annotation

Variant consequences were annotated using Variant Effect Predictor tool.[9] *In silico* prediction of deleterious variants was

carried out by CADD,[10] SIFT,[11] PolyPhen2,[12] and, in the case of splicing variants, Transcript inferred Pathogenicity (TraP) Score.[13] Variant allele frequencies were obtained primarily from the Genome Aggregation Database (gnomADv2.1.1).[15] Gene-phenotype associations were annotated based on OMIM database and Clinical Genomics Database.

Variant filtering

To identify potentially pathogenic variants from the annotated data, all variants within 8 bp of exonic regions were filtered based on the potential modes of inheritance: X-linked, autosomal recessive, dominant, and *de novo* using a similar approach to previously.[3] In recessive filtering, the exome data were analyzed for rare (<150 heterozygous counts and no homozygotes in the gnomADv2.1.1 database)[14] homozygous or compound heterozygous variants including missense, nonsense, splice site, inframe insertion and deletion, and frameshift variants based on Variant Effect Predictor annotations in CCDS genes (Ensembl release 88). In the dominant filtering strategy (applied to both singleton cases, affected parent-child pairs, and the *de novo* variant analysis), we included heterozygous variants with <5 counts in gnomADv2.1.1.

Variant prioritization

Variants surviving the filtering steps were manually assessed and prioritized. All prioritized variants were classified according to ACMG standards and guidelines.[15] As these guidelines are not designed for novel research findings, and because they do not always capture the phenotypic subtleties, we also used a study-specific method of classification. We combined three lines of evidence: (1) at the variant level (e.g., using *in silico* prediction tools), (2) at the pedigree level (e.g., variant segregation data within families), and (3) at the gene level (e.g., prior disease phenotype associations). Each variant was given a score between 0 and 2 for the three lines of evidence making the maximum score 6 (Table S2).

We deemed variants as causative with "high confidence" if a score ≥ 5 was achieved and "moderate confidence" for variants with scores ≥ 4. Variants scoring <4 were not prioritized or reported without the support of functional data.

Variant validation and segregation

Candidate variants in known and potentially novel disease genes were confirmed by bi-directional Sanger sequencing (ABI BigDye 3.1, Applied Biosystems) on ABI3730xl DNA Analyzer. Primers were designed with Primer-BLAST.[16] The sequences were analyzed using Sequencher v.5.3 (Gene Codes Corporation).

Specific splicing *in silico* predictions were made using Human Splicing Finder v3.1.[17] Confirmation of *CLN6* (GenBank: NM_017882.3; MIM: 606725) splicing effect was performed by RT-PCR from total RNA extracted from fibroblast cells followed by sequencing of the abnormal amplicon (Figure S4).

Deletion confirmation of *NEU1* (GenBank: NM_000434.3; MIM: 608272) was performed by quantitative PCR (qPCR) (Figure S5).

Primers for Sanger sequencing and PCR are available upon request.

Copy number variant analysis from WES data

Copy number variants were called from the WES data based on relative sequencing depth. CNVkit was used to call the variants.[18] This analysis was performed separately for WES data generated in the original study[3] and in the current one owing to the different exome capture kits used. CNV analysis focused on known disease genes (annotated against Clinical Genomics Database), in particular those associated with PMEs.

Analysis of short tandem repeats

We additionally examined whether any of the probands had short tandem repeats (STRs) that were expanded at 27 known pathogenic loci (Table S3). The WES samples were examined separately using two STR detection tools, Expansion Hunter v.2.5.5[19] and exSTRa.[20] For each locus we looked for evidence of outlying samples in terms of STR length by inspecting plots of estimated STR size (ExpansionHunter), and empirical cumulative distribution function (eCDF; exSTRa) plots of the number of repeated bases observed for each sample.

Human fibroblast culture

Fibroblast cultures were established from skin biopsy samples of PME1, PME2, PME71, and PME27 as well as control subjects. Cells were cultured in DMEM (GIBCO, Thermo Scientific) plus 10% FBS, 100 units/mL penicillin, 100 μg/mL streptomycin, and 2 mM glutamine in a 37°C and 5% CO_2 humidified incubator.

Microsomal *cis*-prenyltransferase activity measurement

Crude microsomes were prepared as described[21] with minor modifications. *cis*-prenyltransferase (cisPTase) assays and activity measurements in human dermal fibroblasts were performed as described[22,23] with minor modifications. In brief, microsomal fractions from cells were prepared by centrifugation at 100,000 × *g* for 40 min at 4°C. 50 mg microsomal protein was used for cisPTase activity measurement with reaction mixture containing 45 mM FPP, 50 mM [1- 14C]- isopentenyl pyrophosphate (IPP) (55 mCi/mmol; 138,000 cpm/reaction), 25 mM Tris-HCl (pH 7.4), 5 mM $MgCl_2$, 1.25 mM DTT, 2.5 mM sodium orthovanadate, 10 mM Zaragozic acid A, and 0.35% Triton X-100. Reactions were performed at 37°C for 1 h and stopped by the addition of 4 mL of chloroform:-methanol (3:2 ratio). The protein pellet was removed by centrifugation and the supernatant was washed three times with 1/5 volume of 10 mM EDTA in 0.9% NaCl. The incorporation of radioactive IPP into organic fraction containing polyprenyl pyrophosphate was measured by scintillation counting.

Western blot analysis

Cells were washed twice with ice-cold PBS and lysed in lysis buffer (50 mM Tris-HCl, 1% NP-40, 0.1% SDS, 0.1% Deoxycholic Acid, 0.1 mM EDTA, 0.1 mM EGTA, protease and phosphatase inhibitors). Protein extracts were separated by SDS-PAGE and then transferred to nitrocellulose membrane. Primary antibodies against NUS1 (Abcam, ab168351), DHDDS (Sigma, HPA026727), ICAM1 (Santa Cruz, Sc8439), LAMP1 (BD Transduction Laboratories, 611402), and HSP90 (Cell Signaling Technology, 4877) were used. The appropriate LI-COR secondary IRDye antibodies and LI-COR Odyssey Infrared Imaging System (LI-COR, Lincoln, NE) were used for antibody detection.

Filipin staining

Cells were grown on glass cover glasses, fixed in 4% PFA for 10 min, and permeablized in 0.1% Triton X-100 for 5 min. Cells were then incubated with 50 mg/mL filipin (Sigma, F4767) for 1 h. As a positive control for induction of cholesterol accumulation, cells were treated for 16 h with 1 mM U18666A (EMD Biosciences). Relative intensity of filipin staining was quantified by calculating

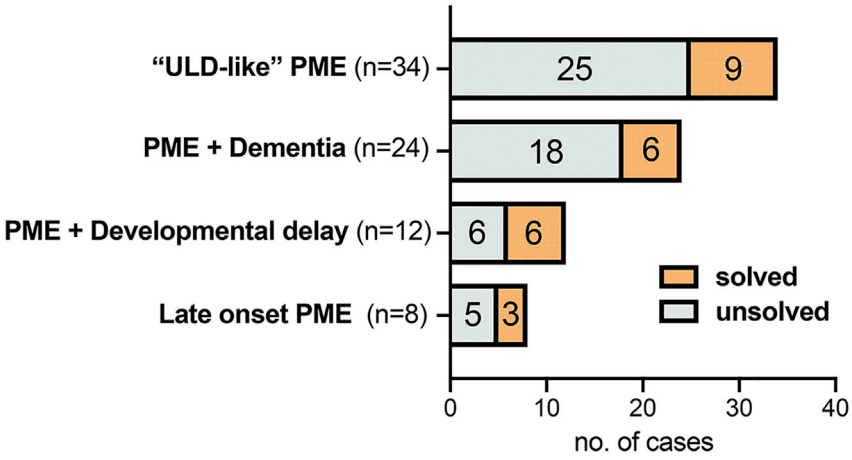


Figure 1. Proportion of all 78 unrelated individuals with PME with (solved) and without (unsolved) likely pathogenic variants by clinical group

average pixel intensity using Adobe Photoshop according to the equation: average filipin intensity = total intensity above low threshold/number of pixels above low threshold.[24]

Yeast strains and culture methods

S. cerevisiae strain *alg10*D (MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 YGR227W::kanX4, Dharmacon) and its derivatives were used. Cultures were grown at 30°C in YPD or Synthetic minimal medium made of 0.67% (wt/vol) yeast nitrogen base and 2% (wt/vol) glucose supplemented with auxotrophic requirements. For solid media, agar (Becton Dickinson) was added at a 2% (wt/vol) final concentration. Yeast transformations were performed by standard yeast genetic methods.

Functional characterization of ALG10/ALG10B variants in yeast

To examine the functionality of hALG10 proteins, the N-glycosylation status of carboxypeptidase Y (CPY) was tested in *S. cerevisiae alg10*D strain transformed with empty pKG-GW1 plasmid (2 μ, LEU2 marker[22]) (negative control) or pKG-GW1 carrying yeast (y) ALG10 ORF (positive control), human (h) ALG10, hALG10 p.Lys391Valfs*35, hALG10B or hALG10B p.Leu253Trp. Yeast transformants were inoculated from single colony and grown overnight at 30°C in synthetic medium lacking leucine. Cells from saturated overnight cultures were harvested and lysed by alkaline method. Whole-cell lysate (WCL) was subjected to SDS-PAGE (7.5% gel) and immunoblotting. Yeast CPY was detected with anti-CPY monoclonal antibodies (Fisher Scientific, clone 10A5B5).

Statistical and data analysis

Statistical analyses and graphical representation were performed with the GraphPad Prism v.7.0 software (GraphPad Software) or the R statistical programming language (v.3.6.1). Figure legends indicate the statistical test used in functional experiments. p values < 0.05 were considered significant.

Gene co-expression networks

For gene co-expression analyses, normalized brain expression values from the BrainSpan Developmental transcriptome dataset were downloaded (see web resources). Genes were removed if they had expression values missing from >50% of the 524 samples available from 42 individuals.

Using the log2 transformed expression values, a matrix of weighted correlations was generated, with weights determined as $1/\sqrt{n}$, where n is the number of samples contributed by the respective individual. Correlation plots were visualized using the corrplot R package, with genes ordered by hierarchical clustering, using the median linkage method.

To determine whether the established and candidate PME genes were more highly co-expressed than expected by chance we randomly sampled 5,000 sets of genes. We calculated the median for each random gene set and compared this to the observed median of the PME gene set.

Results

PME cohort

The study cohort included 84 individuals with PME from 78 families who did not have a known molecular basis. Genomic ancestry checks suggested 74 families (95%) were of European descent, with more than half (n = 46) referred from hospital centers in Italy. The other four families were admixed American (n = 2) and East Asian (n = 2). Inbreeding estimates using FEstim suggested 24% of families were consanguineous (19/78). This was consistent with clinical descriptions of parental relatedness in ten families; detailed pedigree histories were not available for the other nine.

Clinically, the majority of the 78 unrelated affected individuals were classified into two well-established groups: 43% (n = 34) had "ULD-like" PME (i.e., classical childhood/adolescent onset of PME; no dementia) and 31% (n = 24) had PME + dementia. Two smaller groups comprised developmental delay predating PME onset (n = 12) and a group of late-onset (>20 years) PME (n = 8) (Figure 1). Age of disease onset across the cohort ranged from late infancy to 45 years (mean 12 years) (Figure S2).

In total, we identified variants in 24 out of 78 (31%) unrelated affected individuals that we regarded with moderate-to-high confidence (see subjects and methods) as causative. Interestingly, the diagnostic success was highest in one of the two newly recognized, rarer clinical groups (PME with prior developmental delay), although the numbers were small (Figure 1).

We had the most success with individuals in whom we had sequenced additional family members (14/28); we identified a likely causal variant in 45% of trios and in 67% of subjects where another affected 1st-degree relative was sequenced. The proportion of singletons with likely causative variants was significantly less, with just 10 out of 50 cases (OR = 3.9, p value = 0.01, Fisher's exact test).

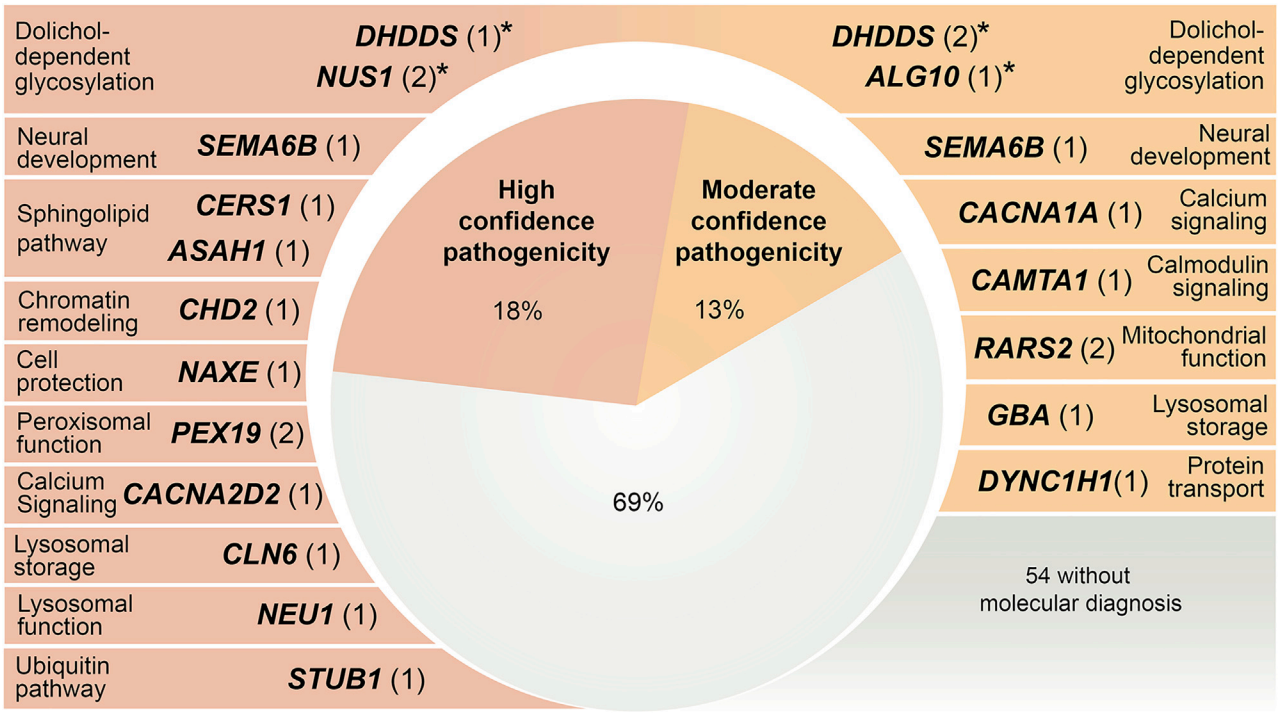


Figure 2. PME-associated genes (n = 18) with high and moderate confidence variants detected in our cohort of 78 unrelated individuals with PME that had previous extensive genetic investigations

The number of unrelated individuals with variants in each gene is shown in parentheses with the known primary function/pathway of each gene also listed. See subjects and methods for criteria followed when classifying variants as high versus moderate confidence. *Functionally validated genes in this study.

The 24 likely solved cases involved 18 genes: 1 (*ALG10* [GenBank: NM_032834.4; MIM: 618355]) has no known disease associations, 6 were known PME genes, including the very recently described *SEMA6B* (GenBank: NM_032108.4; MIM: 608873),[25] and 11 have been reported in other neurological diseases, but not previously in PME (Figure 2, Tables 1, 2, 3, S4, and S5).

Dolichol-dependent glycosylation identified as a PME pathway

In discovering variants in *NUS1* (GenBank: NM_138459.3; MIM: 610463), *DHDDS* (GenBank: NM_024887.3; MIM: 608172), and *ALG10* in a total of 6 unrelated subjects, we identified dolichol-dependent glycosylation as a disease pathway for PME (Figures 2 and S3). The age of onset and clinical features were heterogeneous (Table 1).

We subsequently functionally characterized the variants in these three related genes. *NUS1* and *DHDDS* encode two subunits of cisPTase (also known as dehydrodolichyl diphosphate synthase), the first enzyme committed to dolichol synthesis (Figure S3D).[22,23,26,27] CisPTase is located at a critical branchpoint of farnesyl diphosphate metabolism, with an alternate branch responsible for cholesterol synthesis. *ALG10* is more distal in the dolichol pathway; it is a glucosyltransferase that transfers the terminal glucose residue from dolichyl phosphate glucose (Dol-P-Glc) onto the lipid-linked oligosaccharide precursor Glc2Man9GlcNAc(2)-PP-Dol. The terminal glucose residue added is a key element required for efficient N-linked glycosylation of proteins.[28]

NUS1

Two individuals with PME had variants in *NUS1* (also termed NgBR), encoding the accessory subunit of cisPTase. PME1 carried a *de novo* frameshift variant c.740dupT (p.Asp248Glyfs*15) and PME2 a *de novo* nonsense variant c.310delG (p.Val104*) (Tables 1, S4, and S5, Figure S3A).

Initial analysis of fibroblast cells by western blotting revealed decreased amount of NUS1 in cells from individuals with PME compared to control subjects, implying the presence of nonsense-mediated mRNA decay and/or instability of the truncated proteins (Figure 3B). In PME1, also the amount of DHDDS appeared to be decreased, in line with the predicted truncated NUS1 product that is missing the interface region for heterodimerization with DHDDS and consequently DHDDS instability.[23,29,30] CisPTase activity in cells was drastically decreased, demonstrating that lower protein levels directly affect enzymatic turnover rates (Figure 3A). In order to evaluate the consequence of the reduced cisPTase activity in fibroblast cells, the glycosylation status of ICAM1 and LAMP1, established markers for N-glycosylation defects,[31,32] was examined. Altered ICAM1 and LAMP1 expression and migration were detected by western blot analysis (Figure 3B). Finally, we examined free cholesterol levels, an additional consequence of NUS1 dysfunction in cells.[33] Fibroblasts were stained with filipin

Table 1. Dolichol-dependent glycosylation genes with variants identified in the PME cohort

Patient ID country of origin	Sex	Gene (GenBank)	Variant(s) (LOVD ID)	gnomAD MAF	Inheritance	Clinical summary (onset age)	WES study design	Confidence
PME1 Italy	M	***NUS1*** (NM_138459.3)	c.740dupT (p.Asp248Glyfs*15) (het) (#00334867)	0	*de novo*	myoclonus (13y), no seizures; no ataxia, normal cognition	trio	high
PME2 Italy	F	***NUS1*** (NM_138459.3)	c.310delG (p.Val104*) (het) (#00334873)	0	*de novo*	absence with eyelid myoclonia (4y); myoclonus (8y); ataxia, moderate cognitive impairment; febrile seizures (4y) with developmental regression	trio	high
PME3 Italy	M	***DHDDS*** (NM_024887.3)	c.632G>A (p.Arg211Gln) (het) (#00334875)	0	*de novo*	myoclonus (7y); absences with eyelid myoclonia (9y); mild ataxia, moderate cognitive impairment; developmental delay	trio	high
PME71 Italy	F	***DHDDS*** (NM_024887.3)	c.614G>A (p.Arg205Gln) (het) (#00334877)	0	unknown	ataxia (late infancy); rare TCS (17y), mild action myoclonus (29y); normal cognition	singleton	moderate
PME27 Italy	F	***DHDDS*** (NM_024887.3)	c.283G>A (p.Asp95Asn) (het) (#00334878)	0	unknown	tremor (21y); myoclonus (35y), single TCS (36y); ataxia, normal cognition; bilateral deafness	singleton	moderate
PME50 Turkey	F	***ALG10*** (NM_032834.4)	c.1170_1171delAA (p.Lys391Valfs*35) (hom) (#00334880)	0	AR	frequent myoclonus (13y), rare TCS; ataxia, mild cognitive dysfunction (16y); scoliosis	trio	moderate

Abbreviations: MAF, minor allele frequency; AR, autosomal recessive; het, heterozygous; hom, homozygous; gnomAD, The Genome Aggregation Database; TCS, tonic-clonic seizure. Detailed clinical summaries can be found in Table S5. See subjects and methods for criteria for classifying variants as high versus moderate confidence.

and free cholesterol pools were examined. Both case fibroblasts exhibited increased accumulation of free cholesterol compared to controls (Figure 3C).

DHDDS

Three individuals with PME were identified with rare missense variants in *DHDDS* (also termed hCIT), encoding the catalytic subunit of cisPTase. PME3 was found to have a *de novo* missense variant c.632G>A (p.Arg211Gln) previously described in three individuals with developmental and epileptic encephalopathy (DEE).[34,35] PME71 and PME27 carried heterozygous missense variants c.614G>A (p.Arg205Gln) and c.283G>A (p.Asp95Asn), respectively. No parental samples were available for PME71 for segregation analysis (Tables 1, S4, and S5, Figure S3B). For PME27, it was possible to exclude the c.283G>A variant only in the father as maternal DNA was unavailable. PME27 was also heterozygous for a rare variant in *DNMT1* (c.1619A>G [p.Tyr540Cys]), but without functional support this variant did not meet our criteria for prioritization (Table S6).

Functional studies in fibroblasts from PME71 and PME27 showed apparently normal amounts of both DHDDS and NUS1 (Figure 3B) in line with the preserved capacity for heterodimerization, decreased cisPTase activity in isolated membranes (Figure 3A), and altered levels and migration of ICAM1 and LAMP1 proteins indicating protein N-glycosylation defect (Figure 3B). Furthermore, consistent with reduced cisPTase activity and protein glycosylation defect, increased cholesterol accumulation was detected in both fibroblast cells (Figure 3C). Fibroblasts were not available from PME3 but, because the variant was *de novo* and previously reported, we regarded it as disease causing with high confidence.

ALG10

PME50 was included in the first exome study and identified to carry the homozygous frameshift variant c.1170_1171delAA (p.Lys391Valfs*35) (Tables 1, S4, and S5, Figure S3C) in *ALG10*,[3] encoding a putative alpha-1,2-glucosyltransferase. At that time, with no prior disease association for *ALG10* and with no functional studies performed, the variant was regarded as of uncertain significance. Here, we now provide evidence for its pathogenicity.

Hypo-glycosylation of reporter proteins ICAM1 and LAMP1, identified in western blot analysis of fibroblasts from PME50 (Figure 4A), predicts a defect in alpha-1,2-glucosyltransferase activity. The heterozygous carrier parents of PME50 also showed abnormal glycoslation pattern (Figure 4A). In the sixth decade of their life, both parents

Table 2. High and moderate confidence variants identified in established PME genes

Patient ID country of origin	Sex	Gene (GenBank)	Variant(s) (LOVD ID)	gnomAD MAF	Inheritance	Clinical summary (onset age)	WES study design	Confidence
PME83 Australia	M	***SEMA6B*** (NM_032108.4)	c.1993delC (p.Arg665Glyfs*20) (het) (#00334899)	0	AD	developmental delay and regression; ataxia, tremor (2.5y); drop attacks and absence seizures (4y), TCS (11y), wheelchair (11y), multifocal myoclonus (15 y); severe ID	singleton[a]	high
PME25 Canada	F	***SEMA6B*** (NM_032108.4)	c.2032delG (p.Glu678Argfs*7) (het) (#00334902)	0	AD	developmental delay; ataxia (2.5y); TCS (5y), resting and action myoclonus (10y), possible absence and focal seizures, tremor, wheelchair (14y); moderate ID	singleton	moderate
PME15 Italy	F	***CLN6*** (NM_017882.3)	c.486+28T>C (splicing) (hom)[b] (#00334904)	0	AR	ataxia (14y); severe myoclonus (32y), TCS, dementia, pyramidal signs, psychiatric co-morbidities	singleton	high
PME26 (dec.) Germany	M	***GBA*** (NM_001005742.2)	c.761+4A>G (splicing) (hom) (#00334906)	0	AR	myoclonus (8y); ataxia, ophthalmoplegia, mild cognitive impairment, splenomegaly	singleton	moderate
PME10 Malaysia	M	***NEU1*** (NM_000434.3)	c.544A>G (p.Ser182Gly); deletion of *NEU1* (comp het) (#00334907)	0.001; 0	AR	occasional TCS (12y); frequent myoclonus (14y), ataxia, normal cognition, visual deterioration (20y), cherry-red spots (21y)	trio	high
PME7 Israel	F	***CERS1*** (NM_021267.4)	c.210G>A (p.Trp70*); c.202C>A (p.Leu68Met) (both hom) (#00334908), (#00334910)	0; 0	AR	action myoclonus (11/16yr); ataxia, occasional TCS, mild cognitive impairment	sibling pair; quartet	high
PME8 Israel	F							
PME9 (dec.) Australia	M	***ASAH1*** (NM_004315.4)	c.966−2A>G, splicing; c.504A>C (p.Lys168Asn) (comp het) (#00334911)	0.000004; 0.00006	AR	multifocal myoclonus (10y). TCS, progressive limb and bulbar weakness (16y); hearing impairment (4y); deceased (19y)	trio	high

Abbreviations: MAF, minor allele frequency; comp het, compound heterozygous; hom, homozygous; AR, autosomal recessive; gnomAD, The Genome Aggregation Database; dec., deceased; TCS, tonic-clonic seizure. Detailed clinical summaries can be found in Table S5. See subjects and methods for criteria followed when classifying variants as high versus moderate confidence.
[a]Variant subsequently confirmed *de novo* by Sanger sequencing; maternal DNA did not meet quality control requirements for WES.
[b]Splicing effect of intronic variant confirmed by RT-PCR (see Figure S4).

were neurologically normal. The mother, like PME50 (Table S5), was morbidly obese, while the father was of normal weight.

To confirm the predicted function of ALG10 as an alpha-1,2-glucosyltransferase and to model the *ALG10* variant, we used a yeast alg10 deletion strain to re-express human wild-type and mutant ALG10 proteins for functional complementation. In the absence of dolichyl-phosphoglucose-dependent alpha-1,2-glucosyltransferase activity, the lipid-linked oligosaccharide (N-glycan precursor) lacking terminal glucose is less efficiently transferred to glycoprotein,[28] resulting in the reporter protein for N-glycosylation, CPY, being hypo-glycosylated and running more quickly on SDS-polyacrylamide gel electrophoresis. Both yeast and human ALG10 did complement the yeast deletion strain (Figure 4B), as judged by the presence of mainly the mature form of CPY. Alg10 deletion strains transformed either with empty vector or with mutated ALG10 (Figure 4B) showed multiple bands of CPY corresponding to hypo-glycosylated forms of the protein, thus supporting the pathogenicity of the *ALG10* p.Lys391Valfs*35 variant. Given that the *ALG10* variant was only detected in one individual with PME and has no established disease association, we classified it as disease causing with moderate confidence, despite the functional evidence for its pathogenicity.

PME50, however, is also homozygous for a missense variant c.758T>G (p.Leu253Trp) in the highly homologous *ALG10B* (MIM: 603313) gene encoding alpha-1,2-glucosyltransferase B.[36] The missense variant is reported in gnomAD with an allele frequency of 0.004 with 5 homozygous individuals so it is unlikely to be pathogenic on its own. However, while human ALG10B complemented

Table 3. High and moderate confidence variants identified in established disease genes (not PME)

Patient ID country of origin	Sex	Gene (GenBank)	Disease previously associated with gene (MIM ID)	Variant(s) (LOVD ID)	gnomAD MAF	Inheritance	Clinical presentation	WES study design	Confidence
PME11 Italy	M	***CHD2*** (NM_001271.3)	epileptic encephalopathy, childhood-onset (MIM: 615369)	c.532A>T (p.Arg178*) (het) (#00334913)	0	*de novo*	frequent absence seizures and rare TCS (6y), severe myoclonus (14y); ataxia, dementia; developmental delay	trio	high
PME19 Italy	M	***CACNA2D2*** (NM_001174051.2)	cerebellar atrophy with seizures and variable developmental delay (MIM: 618501)	c.1260G>A (p.Thr420=) (het, *de novo*); c.1112A>G (p.Tyr371Cys) (het, pat inherited) (#00334914)	0; 0	AR	myoclonus, absence and tonic seizures (4y); dementia, no ataxia; developmental delay	trio	high
PME4 (dec.) Italy	F	***STUB1*** (NM_005861.4)	autosomal-recessive spinocerebellar ataxia 16 (MIM: 615768); spinocerebellar ataxia 48 (MIM: 618093)	c.169C>T (p.Pro57Ser) (hom) (#00334915)	0	AR	ataxia (12y); myoclonus, TCS (30y); dementia; tetraparesis	trio	high
PME16 Italy	F	***CACNA1A*** (NM_001127222.2)	early infantile epileptic encephalopathy (MIM: 617106); spinocerebellar ataxia 6 (MIM: 183086); episodic ataxia type 2 (MIM: 108500); familial heiplegic migraine 1 (MIM: 141500); familial hemiplegic migraine 1 with progressive cerebellar ataxia (MIM: 141500)	c.4897G>A (p.Asp1633Asn) (het)[a] (#00334917)	0	unknown	ataxia, myoclonus (30y); cognitive impairment; sensorineural hearing impairment	singleton	moderate
PME17 Italy	F	***CAMTA1*** (NM_015215.4)	non-progressive cerebellar ataxia with mental retardation (MIM: 614756)	c.4418G>C (p.Ser1473Thr) (het) (#00334922), (#00334925)	0.000004	AD	myoclonus (18y), no TCS; no ataxia or dementia	parent-child	moderate
PME18 Italy	M						myoclonus, rare TCS (25y); no ataxia or dementia		
PME21 Malta	M	***PEX19*** (NM_001193644.1)	peroxisome biogenesis disorder 12A (Zellweger) (MIM: 614886)	c.254C>T (p.Ala85Val) (hom) (#00334927), (#00334928)	0.0009	AR	progressive ataxia (7yr); myoclonus, TCS (9y), dementia (10y); limb spasticity	sibling pair	high
PME22 Malta	M						progressive ataxia (8y); myoclonus, TCS (9y); dementia (10y); limb spasticity		

(*Continued on next page*)

Table 3. ***Continued***

Patient ID country of origin	Sex	Gene (GenBank)	Disease previously associated with gene (MIM ID)	Variant(s) (LOVD ID)	gnomAD MAF	Inheritance	Clinical presentation	WES study design	Confidence
PME60 (dec.) Malta	F	***PEX19*** (NM_001193644.1)	Peroxisome biogenesis disorder 12A (Zellweger) (MIM: 614886)	c.254C>T (p.Ala85Val) (hom) (#00334930)	0.0009	AR	progressive severe ataxia (8y); TCS (12y); hypertonia	singleton	high
PME5 (dec.) Italy	F	***NAXE*** (NM_144772.2)	encephalopathy, progressive early-onset, with brain edema and/or leukoencephalopathy (MIM: 617186)	c.128C>A (p.Ser43*) (hom) (#00334932)	0.00003	AR	versive motor seizures (12y), daily absence (13y) and myoclonus (15y), rare TCS (21y); slowly progressive ataxia (19y) dementia and pyramidal signs; developmental delay	singleton	high
PME12 Italy	M	***RARS2*** (NM_020320.3)	pontocerebellar hypolasia type 6 (MIM: 611523)	c.943C>T (p.Arg315*); c.425T>C (p.Val142Ala) (comp het) (#00334933), (#00334935)	0.00004; 0.00005	AR	mild ataxia (childhood), moderate cognitive impairment; rare TCS and absence seizures (9y), mild myoclonus (11y)	sibling pair; quartet	moderate
PME13 Italy	F						ataxia (childhood), moderate cognitive impairment; rare TCS and absence seizures (9), myoclonus (11y)		
PME14 Italy	F	***RARS2*** (NM_020320.3)	pontocerebellar hypolasia type 6 (MIM: 611523)	c.1026G>A (p.Met342Ile); c.3G>A (p.Met1Ile) (comp het) (#00334936)	0.0002; 0	AR	prominent progressive action myoclonus (25y); no TCS, no ataxia, no dementia	trio	moderate
PME64 Italy	M	***DYNC1H1*** (NM_001376.4)	Charcot-Marie-Tooth disease axonal type 20 (MIM: 614228); mental retardation, autosomal-dominant 13 (MIM: 614563); spinal muscular atrophy lower extremity-predominant (MIM: 158600)	c.7828delC (p.Arg2610Glyfs*23) (het) (#00334938)	0	*de novo*	myoclonus (12y), refractory TCS and absence seizures (22y); no ataxia or dementia	trio	moderate

Abbreviations: MAF, minor allele frequency; AR, autosomal recessive; comp het, compound heterozygous; het, heterozygous; hom, homozygous; AD, autosomal dominant; gnomAD, The Genome Aggregation Database; dec., deceased; TCS, tonic-clonic seizure. Detailed clinical summaries can be found in Table S5. Please see subjects and methods for criteria followed when classifying variants as high versus moderate confidence.
[a]See Figure S7 for molecular modeling that supports a loss-of-function effect for this *CACNA1A* variant.

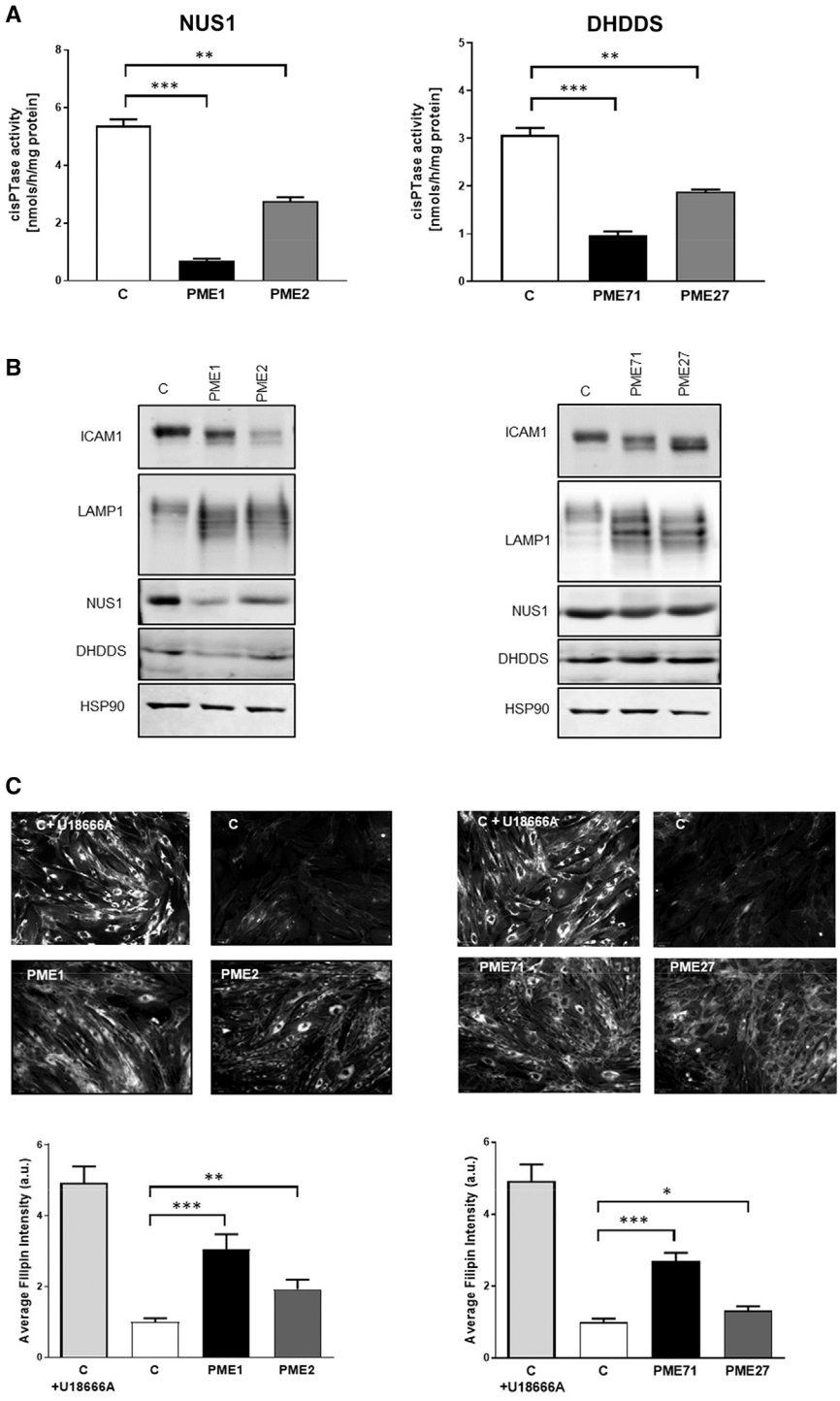


Figure 3. The *NUS1* and *DHDDS* variants cause defects in protein glycosylation due to reduced cisPTase activity in patient-derived fibroblast cell lines

(A) Reduced microsomal cisPTase activity in isolated membranes from PME (*NUS1*: PME1, PME2 and *DHDDS*: PME71, PME27) compared to control (C) fibroblasts. **p < 0.005, ***p < 0.001. Data presented as mean ± SEM of at least three independent measurements.

(B) Affected protein glycosylation in PME fibroblasts. Western blot analysis of NUS1, DHDDS, LAMP1, and ICAM1 levels. HSP90 was used as loading control.

(C) Increased cholesterol accumulation in PME fibroblasts. Filipin staining and quantitative representation from PME and control cells. U18666A was used as a positive control for inhibition of cholesterol trafficking. *p < 0.05, **p < 0.005, ***p < 0.001, a.u., arbitrary units. Data are representative of at least three experiments.

glycosylation in the yeast assay, the variant ALG10B was not quite as effective (Figure 4B), so we could not rule out a contribution of the homozygous *ALG10B* variant to the phenotype.

Likely causative variants in established PME genes

In seven families, we identified likely causative variants in six established PME genes (*SEMA6B, CLN6, GBA* [GenBank: NM_001005742.2; MIM: 606463], *NEU1, CERS1* [GenBank: NM_021267.4; MIM: 606919], *ASAH1* [GenBank: NM_004315.4; MIM: 613468]) (Table 2). These cases all defied diagnosis earlier because of unusual genetic mechanisms or very rare or newly recognized causes.

SEMA6B was recently published as a dominant PME gene with *de novo* variants in four individuals.[26] We independently validate this finding with an additional two affected individuals (PME83, PME25). Both of our subjects had frameshift variants in the last exon of *SEMA6B* (Table 2) within very close proximity to the published series.[26] Low coverage, due to high GC content, of this exon meant that only one of the two variants were initially called by our bioinformatics pipeline and thus both variants escaped detection until targeted *SEMA6B* reanalysis. Clinically, PME83 and PME25 were classified as PME with developmental delay, consistent with the published cases (Table S5). We confirmed the *de novo* status for PME83 by subsequent Sanger sequencing of the parents, but parental DNA was unavailable for PME25.

In the case of *CLN6* and *GBA*, the putative causative variants (Table 2) are both intronic and were not prioritized by initial filtering strategies. Prior to genetic testing, PME15 and PME26 were clinically suspected of having Kufs Type A (MIM: 601780) and Gaucher disease type III (MIM: 231000), respectively.[37,38] Both variants are homozygous and inbreeding coefficient estimates were consistent with

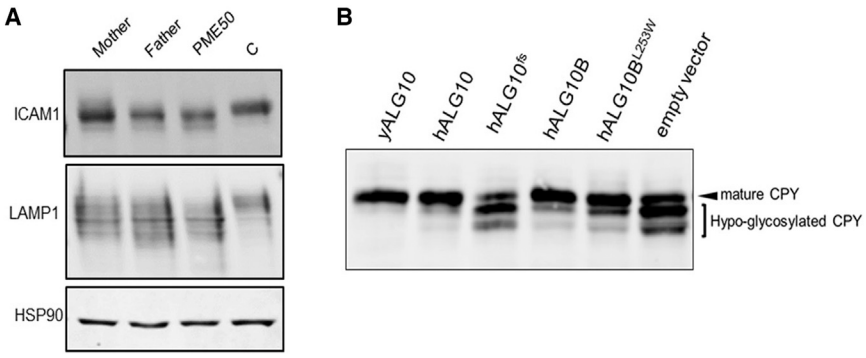


Figure 4. The *ALG10* frameshift mutation causes defects in protein N-glycosylation due to a predicted defect in alpha-1,2-glucosyltransferase activity
(A) Affected protein glycosylation in fibroblasts carrying the *ALG10* and *ALG10B* variants. Western blot analysis of ICAM1 and LAMP1 expression. HSP90 was used as loading control.
(B) Protein N-glycosylation of CPY shows multiple hypo-glycosylated bands in a yeast alg10 deletion strain transformed with mutated human (h) ALG10 (hALG10fs) or empty vector. N-glycosylation deficiency is rescued when transformed with either wild-type yeast ALG10 (yALG10), hALG10, or hALG10B.

parental consanguinity for the two families. Predictions for the *GBA* splice-site variant having an effect on mRNA splicing was consistent across all splicing *in silico* tools; however, without the ability to confirm this experimentally (PME26 deceased), we classified the variant as likely causative with moderate confidence. The deep intronic *CLN6* variant was predicted *in silico* to create an intronic exonic splicing enhancer (ESE) site[39] and RT-PCR from PME15 fibroblast cells confirmed aberrant mRNA splicing (Figure S4A). Sanger sequencing of the aberrant product revealed inclusion of 119 nucleotides of intronic sequence downstream of the 3′ end of exon 4 (Figure S4B). These data are compatible with the homozygous variant in PME15 causing activation of a non-canonical splice site through creation of an intronic ESE site (Figure S4C). The intronic inclusion creates a premature stop codon after 60 nucleotides of open reading frame in the intronic sequence. This is predicted to result in nonsense-mediated decay with partial loss-of-functional protein, compatible with the late-onset CLN6 disease in PME15. As such we classified this variant as likely causative with high confidence (Figure 2; Tables S4 and S5).

Our single CNV finding was at the *NEU1* locus. In PME10, WES data initially suggested a homozygous c.544A>G (p.Ser182Gly) *NEU1* variant (Table 2). Validation by Sanger sequencing showed that only the mother was a heterozygous carrier of the missense variant. Reanalysis of the WES data for a potential CNV in the region indicated the presence of a deletion on the paternal allele, confirmed by quantitative PCR (Figure S5). Subsequently, PME10's younger brother developed symptoms and genetic analysis confirmed his compound heterozygous status for the same *NEU1* variants. Clinically the presentation for both brothers was consistent with sialidosis (MIM: 256550)[40] (Tables S4 and S5).

Recessive variants in very rare PME genes involved in the sphingolipid pathway, *CERS1*[41] and *ASAH1*,[42] were identified in one family each. Siblings PME7 and PME8 were homozygous for two variants in *CERS1*, a nonsense and a missense variant (Table 2). Segregation analysis confirmed heterozygosity for both variants in one of the parents, respectively. The parents were known to be related, with consistent inbreeding F estimates. In PME9, WES revealed compound heterozygous variants in *ASAH1*, one splice-site and one missense variant (Tables 2, S4, and S5); at diagnosis the individual had PME but not spinal muscular atrophy although this subsequently developed.

Likely causative variants in other known disease genes

An additional 11 likely causative variants were identified in genes not previously associated with PME, but recognized in neurological phenotypes including seizures or ataxia (Table 3). *CHD2* (GenBank: NM_001271.3; MIM: 602119), *CACNA2D2* (GenBank: NM_001174051.2; MIM: 607082), and *CACNA1A* (GenBank: NM_001127222.2; MIM: 601011) are established DEE genes, as are *NUS1* and *DHDDS* involved in dolichol metabolism (see above). *CACNA1A* is also associated with ataxia syndromes as are *STUB1* (GenBank: NM_005861.4; MIM: 607207) and *CAMTA1* (GenBank: NM_015215.4; MIM: 611501).

PEX19 (GenBank: NM_001193644.1; MIM: 600279), *NAXE* (GenBank: NM_144772.2; MIM: 608862), *RARS2* (GenBank: NM_020320.3; MIM: 611524), and *DYNC1H1* (GenBank: NM_001376.4; MIM: 600112) are currently associated with more complex neurological phenotypes (Table 3). These variants all met our criteria for moderate to high confidence in causation based on both the genetic data and phenotypic overlap (Tables S4 and S5) (see subjects and methods).

In the case of *PEX19*, this is a well-established gene for peroxisome biogenesis disorders (MIM: 614886). We identified three individuals (PME21, PME22, PME60) from two unrelated families of Maltese origin, with the same homozygous missense variant c.254C>T (p.Ala85Val). All three individuals shared a similar phenotype with onset around age 9 years involving myoclonus, tonic-clonic seizures, ataxia, cognitive decline, and marked photosensitivity (Tables 3 and S5). PME60 had a clinically similarly affected brother who was deceased and not tested. This variant is not present in the Maltese Genome project (J. Vella et al., 2013, European Society of Human Genetics, conference) with 400 individuals; however, haplotype analysis results were consistent with a distant founder effect (Figure S6). Further, independent studies have identified two additional Maltese individuals with the same homozygous variant and similar clinical phenotype (data not shown).

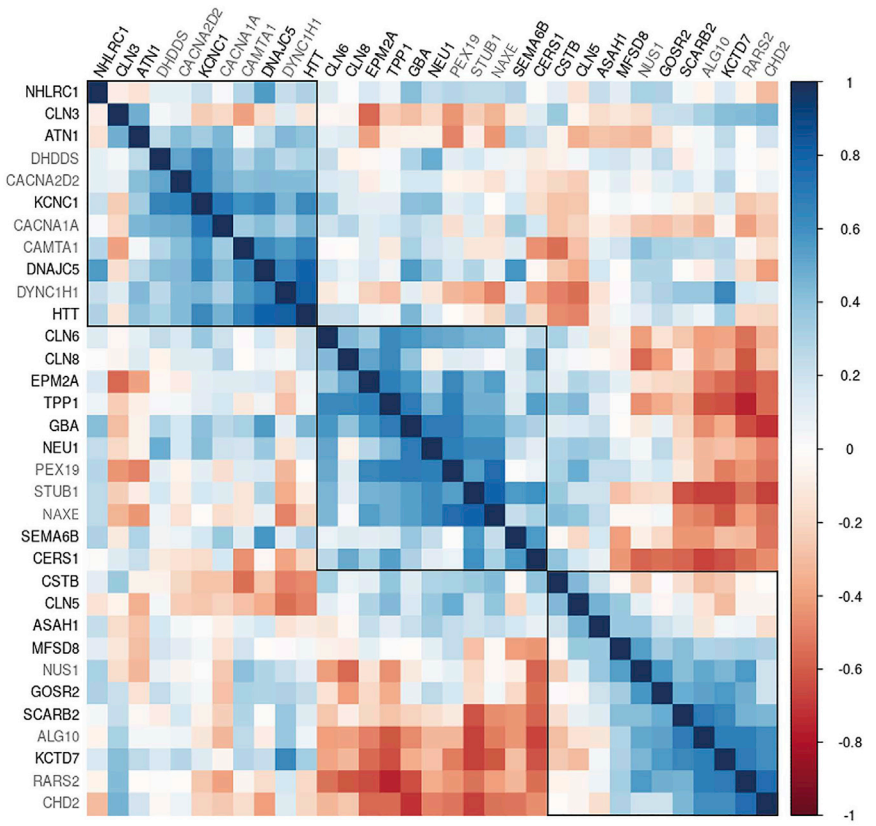


Figure 5. Gene co-expression matrix for 33 known (black) and candidate (gray) PME genes

Pairwise Spearman correlations between genes shown, based on 524 samples from 42 individuals from the BrainSpan resource. Genes are ordered and grouped with hierarchical clustering, using the median linkage method.

Filtered variants that did not meet our criteria for prioritization can be found in Table S6. Our short tandem repeat analyses did not detect any expansions at the known pathogenic loci (Table S3).

PME gene brain co-expression networks

Using brain expression data from BrainSpan, we examined the co-expression between the major established PME genes (Table S1) and all genes we report here with likely causative variants (Tables 1, 2, and 3). Expression data were not available for *MT-TK* (MIM: 590060) responsible for myoclonus epilepsy associated with ragged-red fibers (MERRF [MIM: 545000]); this mitochondrial gene was therefore excluded from the analysis.

The ordered correlation matrices revealed some striking patterns (Figure 5). We observed 3 large clusters of 11 positively correlated gene sets that accounted for all candidate and established (**bold**) PME genes. Cluster one contains ***NHLRC1*** (MIM: 608072), ***CLN3*** (MIM: 607042), ***ATN1*** (MIM: 607462), *DHDDS*, *CACNA2D2*, ***KCNC1***, *CACNA1A*, *CAMTA1*, ***DNAJC5*** (MIM: 611203), *DYNC1H1*, ***HTT*** (MIM: 613004). Cluster two contains ***CLN6***, ***CLN8*** (MIM: 607837), ***EPM2A*** (MIM: 607566), ***TPP1*** (MIM: 607998), ***GBA***, ***NEU1***, *PEX19*, *STUB1*, *NAXE*, ***SEMA6B***, ***CERS1***. Cluster three contains ***CSTB***, ***CLN5*** (MIM: 608102), ***ASAH1***, ***MFSD8*** (MIM: 611124), *NUS1*, ***GOSR2*** (MIM: 604027), ***SCARB2*** (MIM: 602257), *ALG10*, ***KCTD7*** (MIM: 611725), *RARS2*, *CHD2*.

Using a Monte Carlo sampling approach, we found evidence that the established and candidate PME genes were more highly co-expressed than would be expected by chance ($p < 0.05$). These results suggest that overall these genes have similar brain gene expression signatures. Shared biological networks are further supported by the observation that clusters 2 and 3 are negatively correlated.

Discussion

Our data uncovered dolichol-dependent protein glycosylation as an important pathway underlying PME. Additional findings were the confirmation of *SEMA6B* as a cause of PME and that PME can sometimes be a rare manifestation of variants in genes associated with developmental and epileptic encephalopathy or ataxia syndromes. Finally, our results suggest that there is unlikely to be a major shared genetic basis to the remaining unsolved cases, but rather the answer will most likely be a heterogenous mix of rare disorders. However, rare variants in a novel gene, particularly in the introns, and regions of low coverage cannot be excluded.

Overall, we identified plausible pathogenic variants in 24 out of 78 (31%) unrelated affected individuals. This cohort had been extensively studied for known genetic causes previously, so, it is notable that our diagnostic yield was this high. As *de novo* dominant mutations were recently established as an important alternative cause of PME,[3] we pursued a trio-design WES analysis where possible. Overall, we had significantly greater success identifying plausible pathogenic variants in subjects that had been sequenced with other family members (i.e., as an affected trio- or quartet- with unaffected parents or part of an affected sibling or parent-offspring pair). This was driven in part by the importance of *de novo* variants in dominant genes, that has previously been under-appreciated for this disease group, but also the ability to confirm compound heterozygosity and/or homozygosity for variants under a recessive model. Clinically, the two primary categories of PME have historically been separated according to the presence (PME with dementia) or absence ("ULD-like") of cognitive decline. In this analysis of cases

defying molecular diagnosis, two additional clinical groups were apparent: PME with prior developmental delay and a late-onset group. Our success rate in diagnosis was highest for one of the newly recognized, albeit smaller, clinical groups: 50% for PME with prior developmental delay (Figure 1).

We associate dolichol-dependent glycosylation with the PME phenotype through the identification of variants in *NUS1*, *DHDDS*, and *ALG10*, supported by demonstrating glycosylation defects in fibroblast cell lines and/or in yeast assays. Protein glycosylation is a ubiquitous post-translational modification that contributes to several crucial biological and physiological processes within cells. Given that variants in *NUS1*, *DHDDS*, and *ALG10* were associated with altered expression and migration of ICAM1 and LAMP1, and since ALG10 is specifically linked to N-glycosylation, it is plausible that variants in these genes result in N-glycosylation defects in cells.[23,29–32] N-glycosylation followed by oligomannose phosphorylation of the N-glycated protein is pivotal for lysosomal targeting of enzymes.[43] Given that defects in many lysosomal enzymes have been associated with PME, hypoglycosylation caused by impaired N-glycosylation of such proteins may be contributing to the phenotype in individuals with mutations in *NUS1*, *DHDDS*, and *ALG10*. However, the exact mechanisms would need to be explored in further functional studies. Of note, dolichol metabolism was first associated with PME more than 30 years ago with the observation that dolichol content was significantly increased in the brains and urinary sediment of individuals with NCL.[44,45] The reason for this observation remained unknown but was postulated to be caused by a possible defect in dolichol recycling or metabolism. PME now joins the expanding list of phenotypes included under the rubric of congenital disorders of glycosylation, which are quite clinically heterogeneous (see GeneReviews in web resources). Unlike most of the established PME genes where the clinical presentation is somewhat characteristic for each gene, the clinical picture of the dolichol pathway genes (Table 1) is more reminiscent of *TBC1D24* (MIM: 613577) where the clinical spectrum is much wider.[46]

A handful of pathogenic variants in *NUS1* and *DHDDS* have previously been associated with various phenotypes. Bi-allelic mutations in both genes have been reported in single families with congenital disorders of glycosylation showing severe, multiorgan manifestations,[22,47] and in *DHDDS* additionally with retinitis pigmentosa.[48] More recently, heterozygous *de novo* variants in both *NUS1* and *DHDDS* were reported in individuals with DEE.[34,35] Interestingly, one of these *DHDDS* variants was identified in PME3 in our cohort and recently the recurrent p.Arg37His variant was reported in a case with mild intellectual disability, rare generalized seizures, and a stable myoclonic tremor.[49] *NUS1* variants have also been associated with early-onset Parkinson disease with an increase in rare variant burden in PD-affected individuals versus control subjects.[50] Remarkably, variants in two established recessive PME genes, *GBA* and *SCARB2*, are also risk alleles for Parkinson disease.[51–53] Finally, the recent *NUS1* reports of a recurrent heterozygous *de novo* variant in two unrelated individuals with epilepsy, myoclonus, and ataxia[54] and an autosomal-dominant family with epilepsy, ataxia, and tremor segregating a heterozygous frameshift variant,[55] support our conclusion that *NUS1* is a PME gene. However, it is clear that the phenotypic spectrum for both *NUS1* and *DHDDS* is broad.

The majority of proteins involved in the N-glycosylation pathway (like *NUS1* and *DHDDS*) have been associated with mainly autosomal-recessive congenital disorders of glycosylation.[56] *ALG10* is a rare exception as it has previously not been associated with any clinical phenotype, the only exception being our report of it as a candidate gene for PME based on the identification of a homozygous frameshift variant in PME50.[3] Here, through functional characterization of this variant, we give further support for *ALG10* being a PME gene. However, despite functional evidence implying pathogenicity of the reported *ALG10* variant, further individuals with PME should be identified to establish *ALG10* as a disease gene. Interestingly, PME50 was also homozygous for a missense variant in the highly homologous *ALG10B* gene (also known as *KCR1*),[36] that has not been previously associated with any human recessive disorder. Our yeast complementation data imply that the *ALG10B* variant may be a hypomorph with attenuated ability for transferring the glucose residue to the lipid-linked oligosaccharide precursor. In the absence of ALG10 activity, this may not be enough to maintain a proper level of cellular transferase activity. It is therefore possible that compromised function of both genes is required for an *ALG10*-related disease to manifest.

Individuals with PME are typically cognitively normal prior to epilepsy onset. Here we highlight a rare group with prior developmental delay (n = 12); six have plausible genetic findings. Importantly, two of the six had heterozygous frameshift variants in *SEMA6B*. *SEMA6B* was recently discovered as a rare PME gene, with frameshift variants all occurring in the GC-rich last exon of this gene[26] in four subjects. They had mild initial developmental delay, seizure onset between 11 months and 6 years with subsequent cognitive and motor regression, needing assistance with ambulation by the early second decade. Microcephaly and spasticity were present in some. All were regarded as having severe intellectual disability and they were all alive at ages 12–28 years. Our cases had a similar course (Table S5), but did not have microcephaly or significant spasticity and the level of intellectual disability was moderate-severe. To date they have survived until 38 and 39 years without further deterioration, unlike the pattern seen in some PMEs due to storage disorders and those with mutations in *SCARB2* or *GOSR2* with prominent early-adult deterioration and often early death.

Traditionally, PME and DEE are regarded as distinct syndromic groups; this distinction continues to be practically

useful. However, it is now clear that the boundary between these groups is blurred, both from a genetic and phenotypic view point. Three of the other four developmentally delayed PME-affected individuals with molecular findings had variants in established DEE genes. This included *de novo* dominant variants in *DHDDS* and *CHD2* and a recessive *CACNA2D2* mutation. Here, we associate these three DEE genes with PME, building on our initial study where we expanded the *TBC1D24* phenotypic spectrum to PME.[3] Similarly, *KCNA2* (MIM: 176262), another established DEE gene, was recently reported in a single case of PME.[57] In the reverse direction, after we discovered *KCNC1* as a causative *de novo* dominant PME gene, it has now also been established as an important DEE gene, although the causative mutations differ.[58]

We also report putative pathogenic variants in a handful of known ataxia genes (Table 3), both recessive and dominant. These genes join *AFG3L2* (MIM: 604581) and *SACS* (MIM: 604490) reported in our initial study[3] as known ataxia genes with pathogenic variants in individuals with PME. We had a small number of individuals in our PME cohort of 84 who had no reported tonic-clonic seizures making their clinical presentation more consistent with progressive myoclonic ataxia (PMA [MIM: 159700]). This clinical overlap, with both PME and PMA presentations, is well established for genes such as *GOSR2* and *KCNC1*.[6,59] We also identified interesting variants in other known neurological disease genes (Table 3) that not only significantly broaden the genetic basis to the PMEs, but also highlight the need for further functional studies and larger patient numbers to fully understand genotype-phenotype correlations.

The brain gene co-expression analysis uncovered some potentially important relationships between established PME and newly reported PME genes. The advantage of using a brain-specific resource for this analysis, such as BrainSpan, is the detection of brain-specific signatures. An additional advantage of the brain gene co-expression approach is that it is not biased against genes with little known about their function or limited by published material as can be the case for other network generating data sources (e.g., protein-protein interactions or text-mining). The observation that three PME genes associated with the sphingolipid pathway (i.e., *CERS1*, *NEU1*, *GBA*) cluster together in gene set 2 (Figure 5) is proof of principle for the unbiased gene co-expression approach. As such, the clustering of genes that have not previously been biologically associated may indeed be highlighting biologically relevant pathways.

Future perspectives

The PMEs are the genetically best-characterized group of epilepsies. They are highly genetically heterogeneous and there are founder effects, resulting in a different distribution of particular types of PME in various populations. The most comprehensive study of ~200 affected individuals from Italy reached a diagnosis in ~70% of cases although not all were fully investigated.[2] A number of the residual cases have been diagnosed subsequently, including via this study. Future whole-genome sequencing approaches such as long read sequencing, as well as improved bioinformatic software (e.g., for structural variant calling, repeat expansion detection) likely hold the key to uncovering the elusive genetic basis to these remaining rare genetic disorders.

We only report one pathogenic CNV in this study, but we cannot rule out CNVs as a more important genetic factor due again to exome-sequencing data being limited in its ability to detect such genetic variants. The same argument is true for the detection of repeat expansions. Over half of the known disease-causing repeat expansions are located in intronic and UTR gene regions that are not well captured by exome-sequencing data, so it is not perhaps surprising we had no positive results from this analysis. The recent discovery of pathogenic intronic pentanucleotide expansions in familial adult myoclonic epilepsy (FAME), a dominant disorder that is on the mild end of severity of the PME spectrum, reinforces the relevance and importance of searching for known and novel repeat expansions in genetically unsolved individuals with PME.[60–63] Finally, it is also possible that there are important mtDNA pathogenic variants that remain undetected as we did not systematically study the mitochondrial genome.

Our experience with this cohort has highlighted just how genetically heterogeneous the residuum of unsolved individuals with PME are. Of the remaining unsolved cases, they are unlikely to include another gene affecting a large proportion of cases like *KCNC1*,[3] but rather they are probably a collection of multiple rare genetic causes. Collectively, we estimate that it is now less than 20% of individuals with PME that cannot be attributed to known disease genes with intronic variants possibly going undetected in previous analyses as was the case for our *CLN6*- and *GBA*-positive cases. The detection and interpretation of such variants will only improve as the field transitions from exome sequencing to whole-genome sequencing.

Data and code availability

All variants reported in this study have been deposited to the LOVD database; individual LOVD identifiers are provided in Tables 1, 2, 3, and S4. WES data have not been deposited in a public repository due to privacy and ethical restrictions. R code used in brain co-expression analysis is available as supplemental methods.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2021.03.013.

Acknowledgments

The authors are indebted to the families participating in this study. We thank Paula Hakala, Katri Aksentjeff, Saara Tegelberg, Simona

Allievi, and Marta Bayly for technical support and Michael Hildebrand for molecular analysis. The following funding bodies are acknowledged: Swiss National Foundation (Early Postdoc Mobility Grant [to C. Courage]), Folkhälsan Research Foundation (to A.-E.L.), NIH grant R35 HL139945 (to W.C.S.), Australian National Health and Medical Research Council (NHMRC) Program Grants GNT1054618 (to M.B.) and GNT1091593 (to S.F.B. and I.E.S.), NHMRC Senior Research Fellowship (GNT1102971) and Independent Research Institute Infrastructure Support Scheme (IRIISS) (to M.B.), Victorian Government's Operational Infrastructure Support Program (to M.B.), Istanbul University Scientific Research Fund-BAP-2019K12-149071 (to B.B.), NHMRC Senior Research Fellowship (GNT1104718) (to L.M.D.); and NHMRC Practitioner Fellowship (GNT1104831) (to I.E.S.). A.-E.L. is a HiLIFE Fellow at the University of Helsinki.

Declaration of interests

M.M. is employed by Blueprint Genetics. All other authors declare no competing interests.

Received: June 29, 2020
Accepted: March 5, 2021
Published: April 1, 2021

Web resources

BrainSpan – Atlas of the Developing Human Brain, http://www.brainspan.org/
GenBank, https://www.ncbi.nlm.nih.gov/genbank/
GeneReviews, Sparks, S.E., and Krasnewich, D.M. (1993). Congenital Disorders of N-Linked Glycosylation and Multiple Pathway Overview, https://www.ncbi.nlm.nih.gov/books/NBK1332/
gnomAD Browser v.2.1.1, https://gnomad.broadinstitute.org/
Human Splicing Finder, http://www.umd.be/HSF3/
LOVD, https://www.lovd.nl/3.0/home
OMIM, https://www.omim.org/
TraP, http://trap-score.org

References

1. Berkovic, S.F., Andermann, F., Carpenter, S., and Wolfe, L.S. (1986). Progressive myoclonus epilepsies: specific causes and diagnosis. N. Engl. J. Med. *315*, 296–305.
2. Franceschetti, S., Michelucci, R., Canafoglia, L., Striano, P., Gambardella, A., Magaudda, A., Tinuper, P., La Neve, A., Ferlazzo, E., Gobbi, G., et al.; Collaborative LICE study group on PMEs (2014). Progressive myoclonic epilepsies: definitive and still undetermined causes. Neurology *82*, 405–411.
3. Muona, M., Berkovic, S.F., Dibbens, L.M., Oliver, K.L., Maljevic, S., Bayly, M.A., Joensuu, T., Canafoglia, L., Franceschetti, S., Michelucci, R., et al. (2015). A recurrent de novo mutation in KCNC1 causes progressive myoclonus epilepsy. Nat. Genet. *47*, 39–46.
4. Ramachandran, N., Girard, J.M., Turnbull, J., and Minassian, B.A. (2009). The autosomal recessively inherited progressive myoclonus epilepsies and their genes. Epilepsia *50* (*Suppl 5*), 29–36.
5. Kollmann, K., Uusi-Rauva, K., Scifo, E., Tyynelä, J., Jalanko, A., and Braulke, T. (2013). Cell biology and function of neuronal ceroid lipofuscinosis-related proteins. Biochim. Biophys. Acta *1832*, 1866–1881.
6. Oliver, K.L., Franceschetti, S., Milligan, C.J., Muona, M., Mandelstam, S.A., Canafoglia, L., Boguszewska-Chachulska, A.M., Korczyn, A.D., Bisulli, F., Di Bonaventura, C., et al. (2017). Myoclonus epilepsy and ataxia due to KCNC1 mutation: Analysis of 20 cases and K^+ channel properties. Ann. Neurol. *81*, 677–689.
7. Pedersen, B.S., and Quinlan, A.R. (2017). Who's Who? Detecting and Resolving Sample Anomalies in Human DNA Sequencing Studies with Peddy. Am. J. Hum. Genet. *100*, 406–413.
8. Leutenegger, A.L., Labalme, A., Genin, E., Toutain, A., Steichen, E., Clerget-Darpoux, F., and Edery, P. (2006). Using genomic inbreeding coefficient estimates for homozygosity mapping of rare recessive traits: application to Taybi-Linder syndrome. Am. J. Hum. Genet. *79*, 62–66.
9. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biol. *17*, 122.
10. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J., and Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. *47* (D1), D886–D894.
11. Sim, N.L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., and Ng, P.C. (2012). SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res. *40*, W452-7.
12. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat. Methods *7*, 248–249.
13. Gelfman, S., Wang, Q., McSweeney, K.M., Ren, Z., La Carpia, F., Halvorsen, M., Schoch, K., Ratzon, F., Heinzen, E.L., Boland, M.J., et al. (2017). Annotating pathogenic non-coding variants in genic regions. Nat. Commun. *8*, 236.
14. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al.; Genome Aggregation Database Consortium (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443.
15. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al.; ACMG Laboratory Quality Assurance Committee (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet. Med. *17*, 405–424.
16. Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T.L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics *13*, 134.
17. Desmet, F.O., Hamroun, D., Lalande, M., Collod-Béroud, G., Claustres, M., and Béroud, C. (2009). Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Research *37*, e67.
18. Talevich, E., Shain, A.H., Botton, T., and Bastian, B.C. (2016). CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. PLoS Comput. Biol. *12*, e1004873.
19. Dolzhenko, E., van Vugt, J.J.F.A., Shaw, R.J., Bekritsky, M.A., van Blitterswijk, M., Narzisi, G., Ajay, S.S., Rajan, V., Lajoie,

B.R., Johnson, N.H., et al.; US–Venezuela Collaborative Research Group (2017). Detection of long repeat expansions from PCR-free whole-genome sequence data. Genome Res. *27*, 1895–1903.
20. Tankard, R.M., Bennett, M.F., Degorski, P., Delatycki, M.B., Lockhart, P.J., and Bahlo, M. (2018). Detecting Expansions of Tandem Repeats in Cohorts Sequenced with Short-Read Sequencing Data. Am. J. Hum. Genet. *103*, 858–873.
21. Rush, J.S., Matveev, S., Guan, Z., Raetz, C.R., and Waechter, C.J. (2010). Expression of functional bacterial undecaprenyl pyrophosphate synthase in the yeast rer2Delta mutant and CHO cells. Glycobiology *20*, 1585–1593.
22. Park, E.J., Grabińska, K.A., Guan, Z., Stránecký, V., Hartmannová, H., Hodaňová, K., Barešová, V., Sovová, J., Jozsef, L., Ondrušková, N., et al. (2014). Mutation of Nogo-B receptor, a subunit of cis-prenyltransferase, causes a congenital disorder of glycosylation. Cell Metab. *20*, 448–457.
23. Harrison, K.D., Park, E.J., Gao, N., Kuo, A., Rush, J.S., Waechter, C.J., Lehrman, M.A., and Sessa, W.C. (2011). Nogo-B receptor is necessary for cellular dolichol biosynthesis and protein N-glycosylation. EMBO J. *30*, 2490–2500.
24. Pipalia, N.H., Huang, A., Ralph, H., Rujoi, M., and Maxfield, F.R. (2006). Automated microscopy screening for compounds that partially revert cholesterol accumulation in Niemann-Pick C cells. J. Lipid Res. *47*, 284–301.
25. Hamanaka, K., Imagawa, E., Koshimizu, E., Miyatake, S., Tohyama, J., Yamagata, T., et al. (2020). De Novo Truncating Variants in the Last Exon of SEMA6B Cause Progressive Myoclonic Epilepsy. Amer. J. Hum. Genet. *106*, 549–558.
26. Grabińska, K.A., Edani, B.H., Park, E.J., Kraehling, J.R., and Sessa, W.C. (2017). A conserved C-terminal RXG motif in the NgBR subunit of *cis*-prenyltransferase is critical for prenyltransferase activity. J. Biol. Chem. *292*, 17351–17361.
27. Grabińska, K.A., Park, E.J., and Sessa, W.C. (2016). cis-Prenyltransferase: New Insights into Protein Glycosylation, Rubber Synthesis, and Human Diseases. J. Biol. Chem. *291*, 18582–18590.
28. Burda, P., and Aebi, M. (1998). The ALG10 locus of Saccharomyces cerevisiae encodes the α-1,2 glucosyltransferase of the endoplasmic reticulum: the terminal glucose of the lipid-linked oligosaccharide is required for efficient N-linked glycosylation. Glycobiology *8*, 455–462.
29. Edani, B.H., Grabińska, K.A., Zhang, R., Park, E.J., Siciliano, B., Surmacz, L., Ha, Y., and Sessa, W.C. (2020). Structural elucidation of the *cis*-prenyltransferase NgBR/DHDDS complex reveals insights in regulation of protein glycosylation. Proc. Natl. Acad. Sci. USA *117*, 20794–20802.
30. Bar-El, M.L., Vaňková, P., Yeheskel, A., Simhaev, L., Engel, H., Man, P., Haitin, Y., and Giladi, M. (2020). Structural basis of heterotetrameric assembly and disease mutations in the human cis-prenyltransferase complex. Nat. Commun. *11*, 5273.
31. Medina-Cano, D., Ucuncu, E., Nguyen, L.S., Nicouleau, M., Lipecka, J., Bizot, J.C., Thiel, C., Foulquier, F., Lefort, N., Faivre-Sarrailh, C., et al. (2018). High N-glycan multiplicity is critical for neuronal adhesion and sensitizes the developing cerebellum to N-glycosylation defect. eLife *7*, 7.
32. He, P., Ng, B.G., Losfeld, M.E., Zhu, W., and Freeze, H.H. (2012). Identification of intercellular cell adhesion molecule 1 (ICAM-1) as a hypoglycosylation marker in congenital disorders of glycosylation cells. J. Biol. Chem. *287*, 18210–18217.
33. Harrison, K.D., Miao, R.Q., Fernandez-Hernándo, C., Suárez, Y., Dávalos, A., and Sessa, W.C. (2009). Nogo-B receptor stabilizes Niemann-Pick type C2 protein and regulates intracellular cholesterol trafficking. Cell Metab. *10*, 208–218.
34. Hamdan, F.F., Myers, C.T., Cossette, P., Lemay, P., Spiegelman, D., Laporte, A.D., Nassif, C., Diallo, O., Monlong, J., Cadieux-Dion, M., et al.; Deciphering Developmental Disorders Study (2017). High Rate of Recurrent De Novo Mutations in Developmental and Epileptic Encephalopathies. Am. J. Hum. Genet. *101*, 664–685.
35. Lelieveld, S.H., Reijnders, M.R., Pfundt, R., Yntema, H.G., Kamsteeg, E.J., de Vries, P., de Vries, B.B., Willemsen, M.H., Kleefstra, T., Löhner, K., et al. (2016). Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. Nat. Neurosci. *19*, 1194–1196.
36. Nakajima, T., Hayashi, K., Viswanathan, P.C., Kim, M.-Y., Anghelescu, M., Barksdale, K.A., Shuai, W., Balser, J.R., and Kupershmidt, S. (2007). HERG is protected from pharmacological block by α-1,2-glucosyltransferase function. J. Biol. Chem. *282*, 5506–5513.
37. Arsov, T., Smith, K.R., Damiano, J., Franceschetti, S., Canafoglia, L., Bromhead, C.J., Andermann, E., Vears, D.F., Cossette, P., Rajagopalan, S., et al. (2011). Kufs disease, the major adult form of neuronal ceroid lipofuscinosis, caused by mutations in CLN6. Am. J. Hum. Genet. *88*, 566–573.
38. Park, J.K., Orvisky, E., Tayebi, N., Kaneski, C., Lamarca, M.E., Stubblefield, B.K., Martin, B.M., Schiffmann, R., and Sidransky, E. (2003). Myoclonic epilepsy in Gaucher disease: genotype-phenotype insights from a rare patient subgroup. Pediatr. Res. *53*, 387–395.
39. Vaz-Drago, R., Custódio, N., and Carmo-Fonseca, M. (2017). Deep intronic mutations and human disease. Hum. Genet. *136*, 1093–1111.
40. Canafoglia, L., Robbiano, A., Pareyson, D., Panzica, F., Nanetti, L., Giovagnoli, A.R., Venerando, A., Gellera, C., Franceschetti, S., and Zara, F. (2014). Expanding sialidosis spectrum by genome-wide screening: NEU1 mutations in adult-onset myoclonus. Neurology *82*, 2003–2006.
41. Vanni, N., Fruscione, F., Ferlazzo, E., Striano, P., Robbiano, A., Traverso, M., Sander, T., Falace, A., Gazzerro, E., Bramanti, P., et al. (2014). Impairment of ceramide synthesis causes a novel progressive myoclonus epilepsy. Ann. Neurol. *76*, 206–212.
42. Zhou, J., Tawk, M., Tiziano, F.D., Veillet, J., Bayes, M., Nolent, F., Garcia, V., Servidei, S., Bertini, E., Castro-Giner, F., et al. (2012). Spinal muscular atrophy associated with progressive myoclonic epilepsy is caused by mutations in ASAH1. Am. J. Hum. Genet. *91*, 5–14.
43. Kim, J.-J.P., Olson, L.J., and Dahms, N.M. (2009). Carbohydrate recognition by the mannose-6-phosphate receptors. Curr. Opin. Struct. Biol. *19*, 534–542.
44. Wolfe, L.S., Ng Ying Kin, N.M., Palo, J., and Haltia, M. (1983). Dolichols in brain and urinary sediment in neuronal ceroid lipofuscinosis. Neurology *33*, 103–106.
45. Wolfe, L.S., Palo, J., Santavuori, P., Andermann, F., Andermann, E., Jacob, J.C., and Kolodny, E. (1986). Urinary sediment dolichols in the diagnosis of neuronal ceroid-lipofuscinosis. Ann. Neurol. *19*, 270–274.
46. Balestrini, S., Milh, M., Castiglioni, C., Lüthy, K., Finelli, M.J., Verstreken, P., Cardon, A., Stražišar, B.G., Holder, J.L., Jr., Lesca, G., et al. (2016). TBC1D24 genotype-phenotype correlation: Epilepsies and other neurologic features. Neurology *87*, 77–85.

47. Sabry, S., Vuillaumier-Barrot, S., Mintet, E., Fasseu, M., Valayannopoulos, V., Héron, D., Dorison, N., Mignot, C., Seta, N., Chantret, I., et al. (2016). A case of fatal Type I congenital disorders of glycosylation (CDG I) associated with low dehydrodolichol diphosphate synthase (DHDDS) activity. Orphanet J. Rare Dis. *11*, 84.
48. Züchner, S., Dallman, J., Wen, R., Beecham, G., Naj, A., Farooq, A., Kohli, M.A., Whitehead, P.L., Hulme, W., Konidari, I., et al. (2011). Whole-exome sequencing links a variant in DHDDS to retinitis pigmentosa. Am. J. Hum. Genet. *88*, 201–206.
49. Togashi, N., Fujita, A., Shibuya, M., Uneoka, S., Miyabayashi, T., Sato, R., Okubo, Y., Endo, W., Inui, T., Jin, K., et al. (2020). Fifteen-year follow-up of a patient with a DHDDS variant with non-progressive early onset myoclonic tremor and rare generalized epilepsy. Brain Dev. *42*, 696–699.
50. Guo, J.F., Zhang, L., Li, K., Mei, J.P., Xue, J., Chen, J., Tang, X., Shen, L., Jiang, H., Chen, C., et al. (2018). Coding mutations in *NUS1* contribute to Parkinson's disease. Proc. Natl. Acad. Sci. USA *115*, 11567–11572.
51. Gan-Or, Z., Dion, P.A., and Rouleau, G.A. (2015). Genetic perspective on the role of the autophagy-lysosome pathway in Parkinson disease. Autophagy *11*, 1443–1457.
52. Michelakakis, H., Xiromerisiou, G., Dardiotis, E., Bozi, M., Vassilatis, D., Kountra, P.M., Patramani, G., Moraitou, M., Papadimitriou, D., Stamboulis, E., et al. (2012). Evidence of an association between the scavenger receptor class B member 2 gene and Parkinson's disease. Mov. Disord. *27*, 400–405.
53. Aharon-Peretz, J., Rosenbaum, H., and Gershoni-Baruch, R. (2004). Mutations in the glucocerebrosidase gene and Parkinson's disease in Ashkenazi Jews. N. Engl. J. Med. *351*, 1972–1977.
54. Den, K., Kudo, Y., Kato, M., Watanabe, K., Doi, H., Tanaka, F., Oguni, H., Miyatake, S., Mizuguchi, T., Takata, A., et al. (2019). Recurrent NUS1 canonical splice donor site mutation in two unrelated individuals with epilepsy, myoclonus, ataxia and scoliosis - a case report. BMC Neurol. *19*, 253.
55. Araki, K., Nakamura, R., Ito, D., Kato, K., Iguchi, Y., Sahashi, K., Toyama, M., Hamada, K., Okamoto, N., Wada, Y., et al. (2020). *NUS1* mutation in a family with epilepsy, cerebellar ataxia, and tremor. Epilepsy Res. *164*, 106371.
56. Ng, B.G., and Freeze, H.H. (2018). Perspectives on glycosylation and its congenital disorders. Trends Genet. *34*, 466–476.
57. Canafoglia, L., Castellotti, B., Ragona, F., Freri, E., Granata, T., Chiapparini, L., Gellera, C., Scaioli, V., Franceschetti, S., and DiFrancesco, J.C. (2019). Progressive myoclonus epilepsy caused by a gain-of-function KCNA2 mutation. Seizure *65*, 106–108.
58. Cameron, J.M., Maljevic, S., Nair, U., Aung, Y.H., Cogné, B., Bézieau, S., Blair, E., Isidor, B., Zweier, C., Reis, A., et al. (2019). Encephalopathies with KCNC1 variants: genotype-phenotype-functional correlations. Ann. Clin. Transl. Neurol. *6*, 1263–1272.
59. Boissé Lomax, L., Bayly, M.A., Hjalgrim, H., Møller, R.S., Vlaar, A.M., Aaberg, K.M., Marquardt, I., Gandolfo, L.C., Willemsen, M., Kamsteeg, E.J., et al. (2013). 'North Sea' progressive myoclonus epilepsy: phenotype of subjects with GOSR2 mutation. Brain *136*, 1146–1154.
60. Ishiura, H., Doi, K., Mitsui, J., Yoshimura, J., Matsukawa, M.K., Fujiyama, A., Toyoshima, Y., Kakita, A., Takahashi, H., Suzuki, Y., et al. (2018). Expansions of intronic TTTCA and TTTTA repeats in benign adult familial myoclonic epilepsy. Nat. Genet. *50*, 581–590.
61. Corbett, M.A., Kroes, T., Veneziano, L., Bennett, M.F., Florian, R., Schneider, A.L., Coppola, A., Licchetta, L., Franceschetti, S., Suppa, A., et al. (2019). Intronic ATTTC repeat expansions in STARD7 in familial adult myoclonic epilepsy linked to chromosome 2. Nat. Commun. *10*, 4920.
62. Florian, R.T., Kraft, F., Leitão, E., Kaya, S., Klebe, S., Magnin, E., van Rootselaar, A.F., Buratti, J., Kühnel, T., Schröder, C., et al.; FAME consortium (2019). Unstable TTTTA/TTTCA expansions in MARCH6 are associated with Familial Adult Myoclonic Epilepsy type 3. Nat. Commun. *10*, 4919.
63. Yeetong, P., Pongpanich, M., Srichomthong, C., Assawapitaksakul, A., Shotelersuk, V., Tantirukdham, N., Chunharas, C., Suphapeetiporn, K., and Shotelersuk, V. (2019). TTTCA repeat insertions in an intron of YEATS2 in benign adult familial myoclonic epilepsy type 4. Brain *142*, 3360–3366.

The American Journal of Human Genetics, Volume *108*

Supplemental information

Progressive myoclonus epilepsies—Residual unsolved cases have marked genetic heterogeneity including dolichol-dependent protein glycosylation pathway genes

Carolina Courage, Karen L. Oliver, Eon Joo Park, Jillian M. Cameron, Kariona A. Grabińska, Mikko Muona, Laura Canafoglia, Antonio Gambardella, Edith Said, Zaid Afawi, Betul Baykan, Christian Brandt, Carlo di Bonaventura, Hui Bein Chew, Chiara Criscuolo, Leanne M. Dibbens, Barbara Castellotti, Patrizia Riguzzi, Angelo Labate, Alessandro Filla, Anna T. Giallonardo, Geza Berecki, Christopher B. Jackson, Tarja Joensuu, John A. Damiano, Sara Kivity, Amos Korczyn, Aarno Palotie, Pasquale Striano, Davide Uccellini, Loretta Giuliano, Eva Andermann, Ingrid E. Scheffer, Roberto Michelucci, Melanie Bahlo, Silvana Franceschetti, William C. Sessa, Samuel F. Berkovic, and Anna-Elina Lehesjoki

Supplemental Figures

Figure S1. **Flowchart describing sequencing cohorts and case collection.**

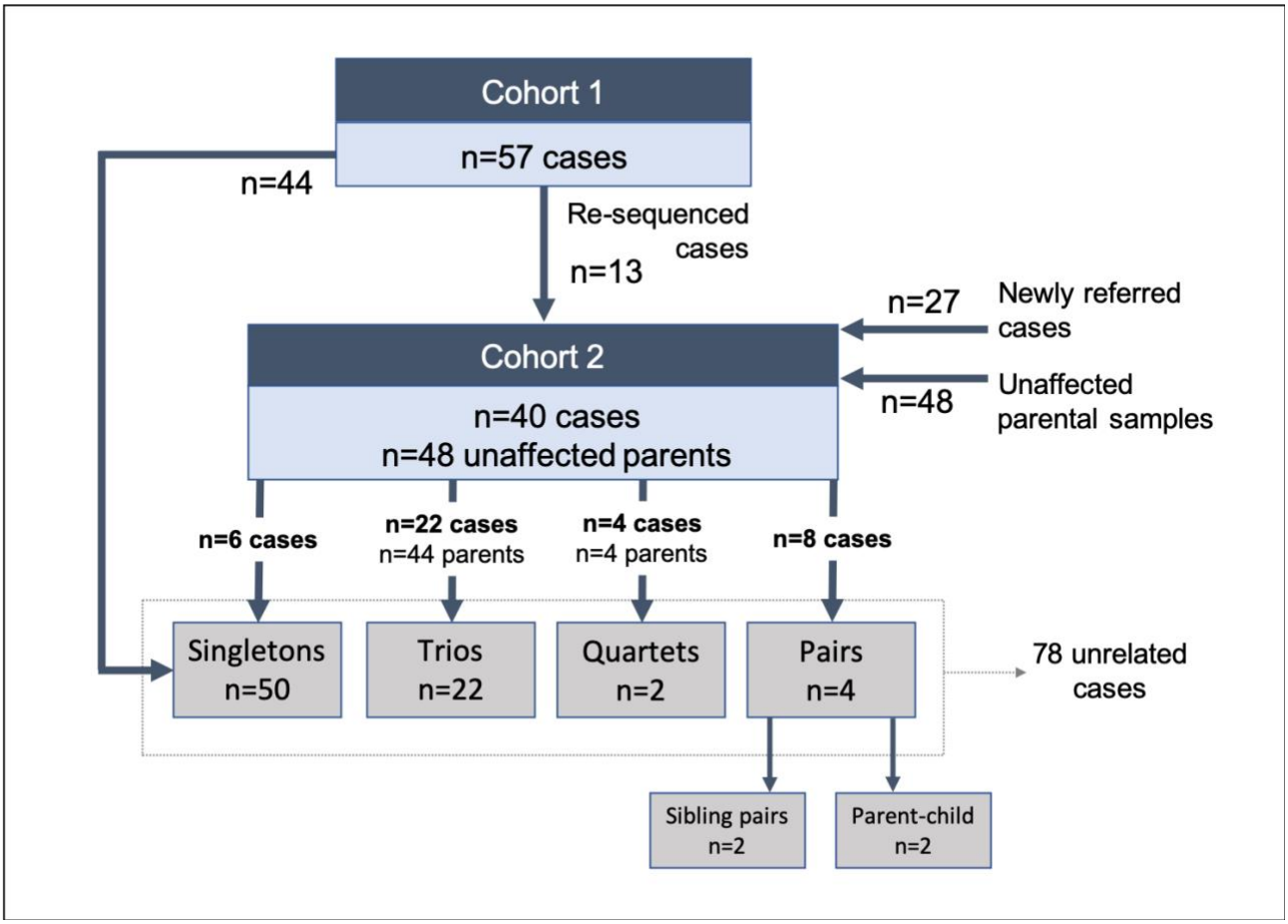


Figure S2. **Age of PME onset distribution for all 78 unrelated probands.**

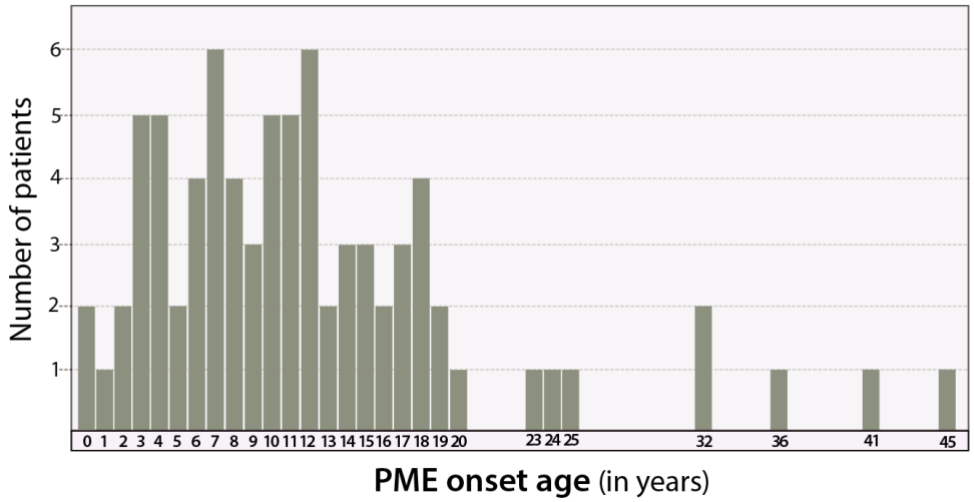
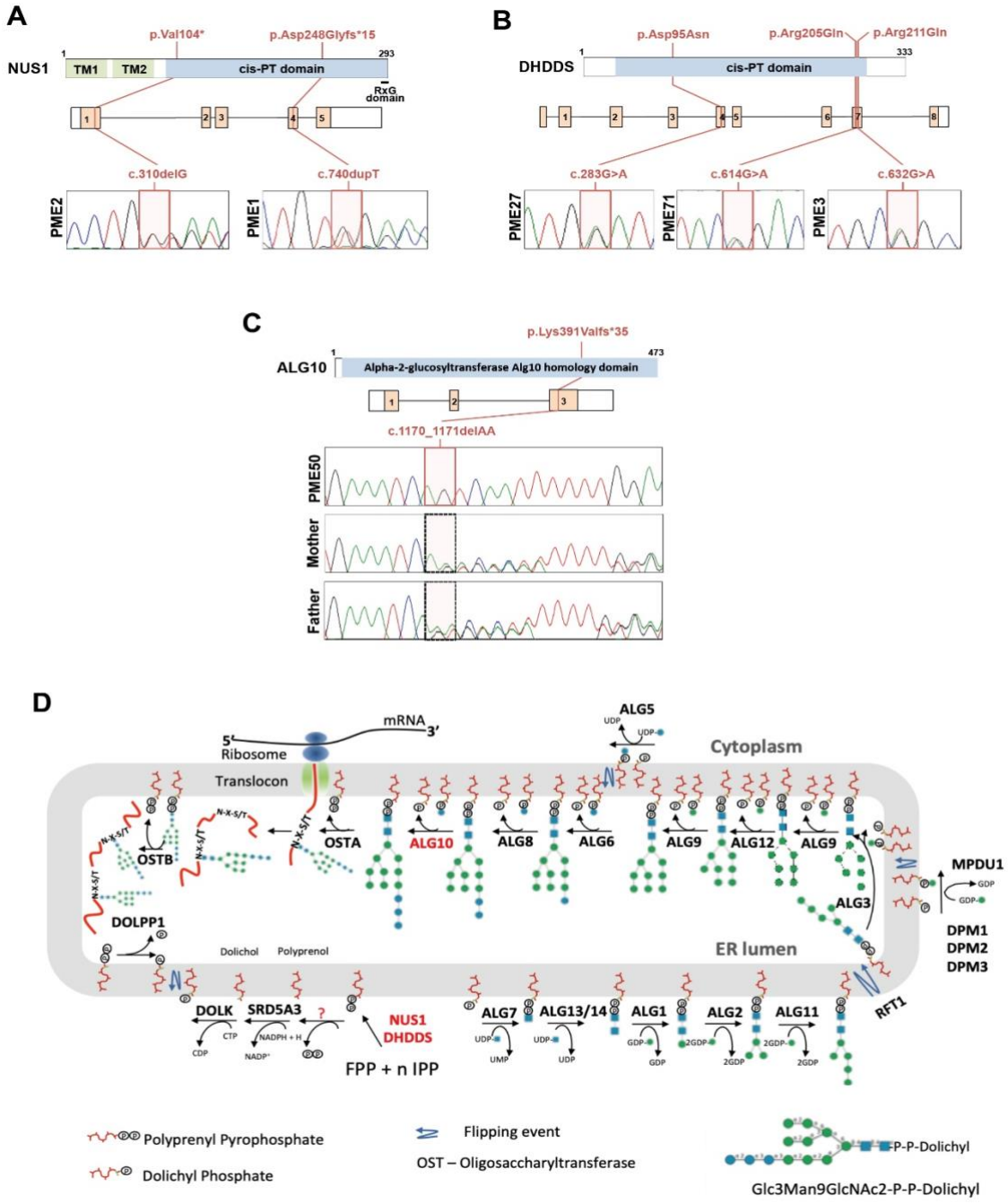


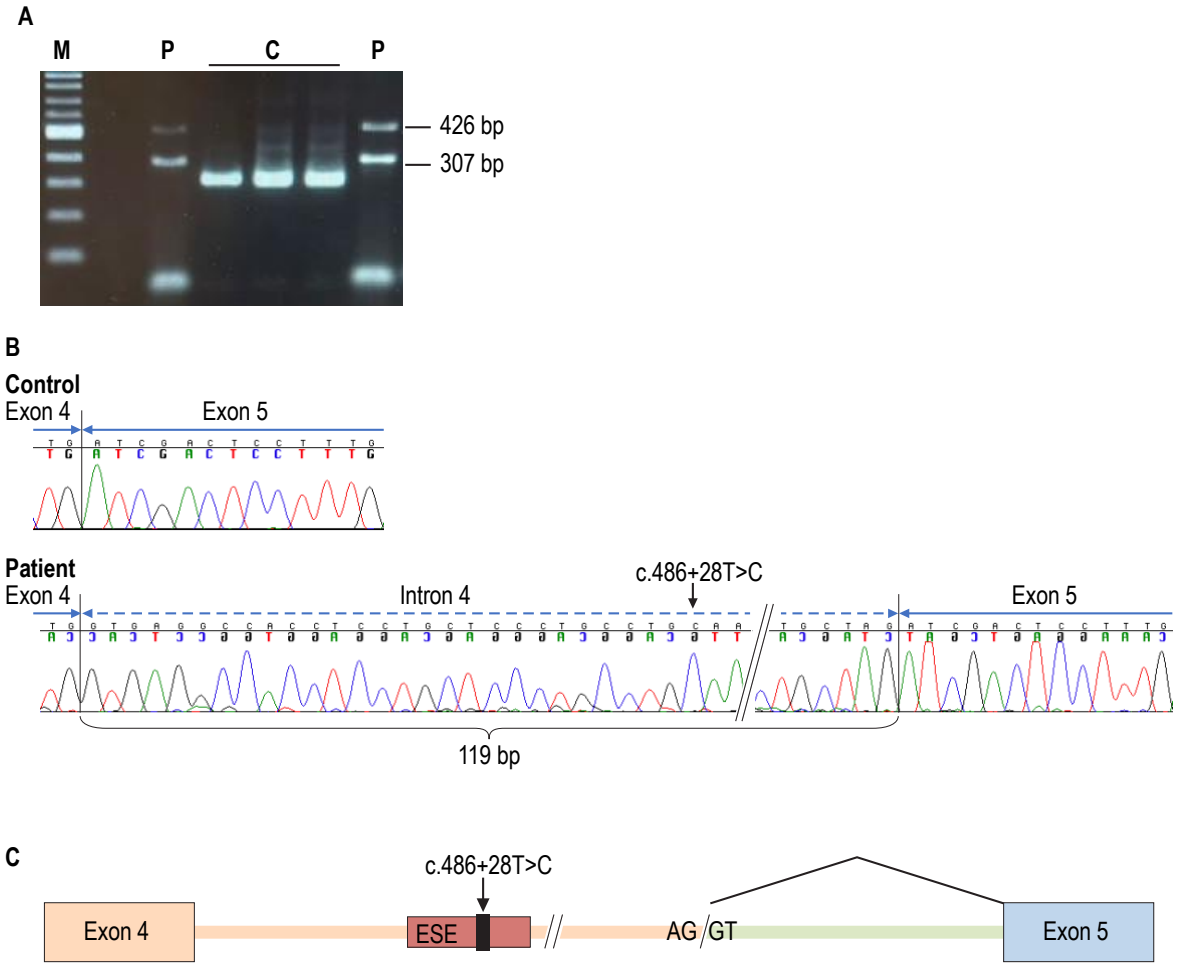
Figure S3. **Pathogenic variants in *NUS1*, *DHDDS* and *ALG10* and dolichol-dependent glycosylation pathway.**



Abbreviations: *DHDDS* - Dehydrodolichyl Diphosphate Synthase Subunit; *FPP* - farnesyl pyrophospatase domain; *IPP* - isopentenyl pyrophosphatase domain; *NPC2* - Intracellular cholesterol transporter 2; *NUS1* - Nuclear Undecaprenyl Pyrophosphate Synthase 1 (Nogo-B Receptor), TM - transmembrane domain

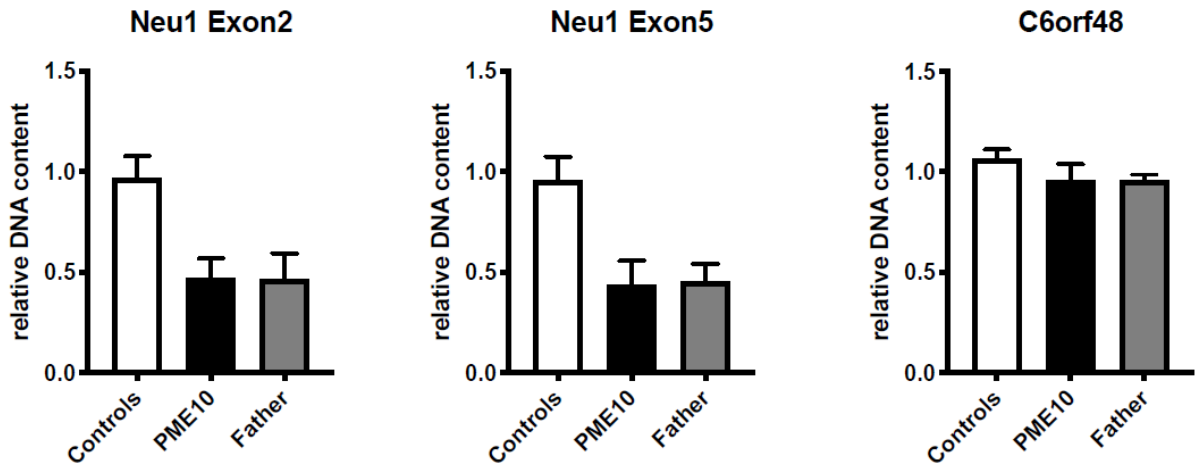
(**A**) Locations of variants in *NUS1*. (**B**) Locations of variants in *DHDDS*. (**C**) Locations of variant in *ALG10.* (**D**) Glycosylation pathway showing involvement of NUS1, DHDDS and ALG10 (in red) (adapted from Stanley P, Taniguchi N, Aebi M. N-Glycans. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW*, et al.*, editors. Essentials of Glycobiology. Cold Spring Harbor (NY); 2015. p. 99-111.)[1]

Figure S4: **Aberrant splicing caused by the deep intronic *CLN6* variant.**



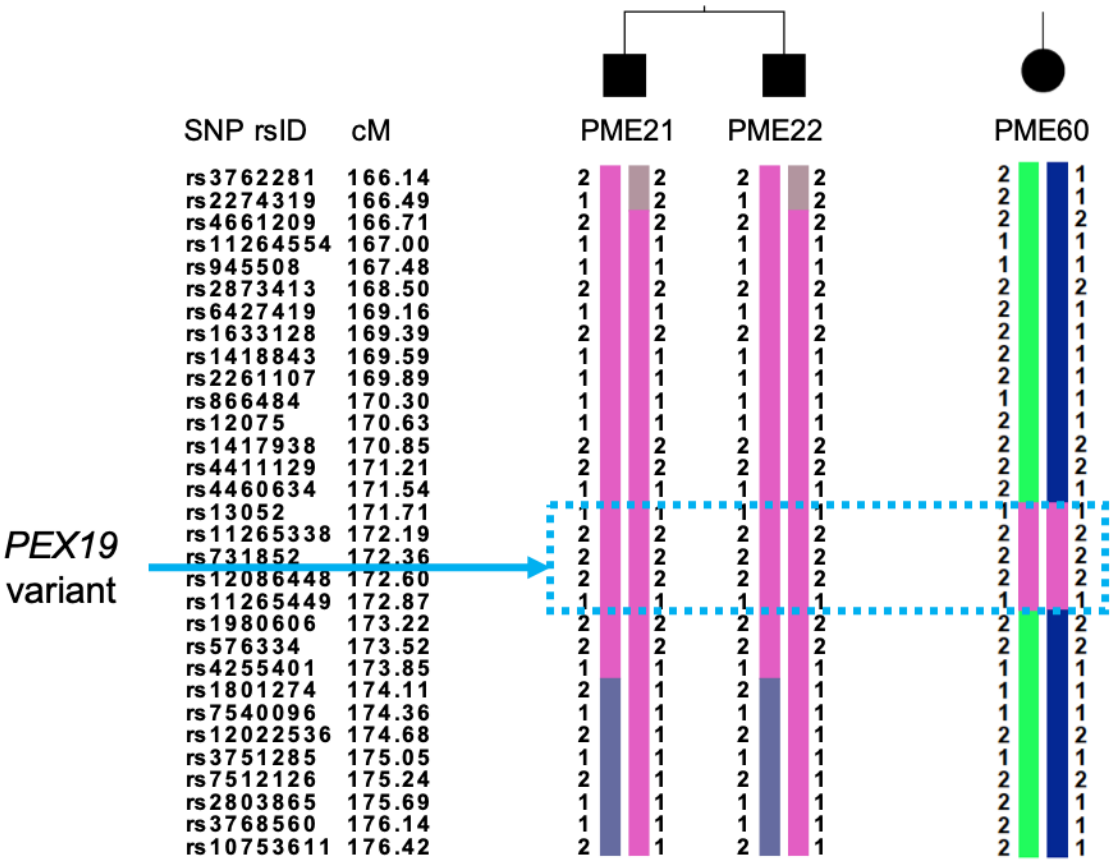
(**A**) Agarose gel electrophoresis showing the migration of RT-PCR products amplified from total RNA extracted from patient fibroblast cells using primers from exons 4 and 6 of *CLN6*. From patient (P) cells two fragments are amplified. Controls (C) show one strong amplicon. The sizes of the two fragments identified in the patient samples, based on sequence analysis, are shown on the right. The lower fragment corresponds to the expected product. The fragments seen in controls also correspond to the expected product, based on sequence analysis, even if the fragments run differently from those in the patient samples. (**B**) Partial sequence chromatogram of a control individual sample shows expected sequence in the exon 4-exon 5 boundary in the 307-bp amplicon. Partial sequence chromatogram from the 426-bp amplicon in the patient sample. The exon 4 sequence is followed by 119 bp of intronic sequence (shown only in part) before beginning of the exon 5 sequence. The position of the homozygous c.486+28T>C variant is pointed by an arrow. (**C**) Schematic representation of intron 4 of *CLN6* showing the position of the c.486+28T>C patient, the intronic ESE created by the variant and the non-canonical splice site (AG/GT) activated. The intronic sequence included in the 426-bp amplicon is shown in pink color and the intronic sequence excluded from the mRNA is shown in green.

Figure S5. **Deletion confirmation of NEU1 was performed by quantitative PCR.**



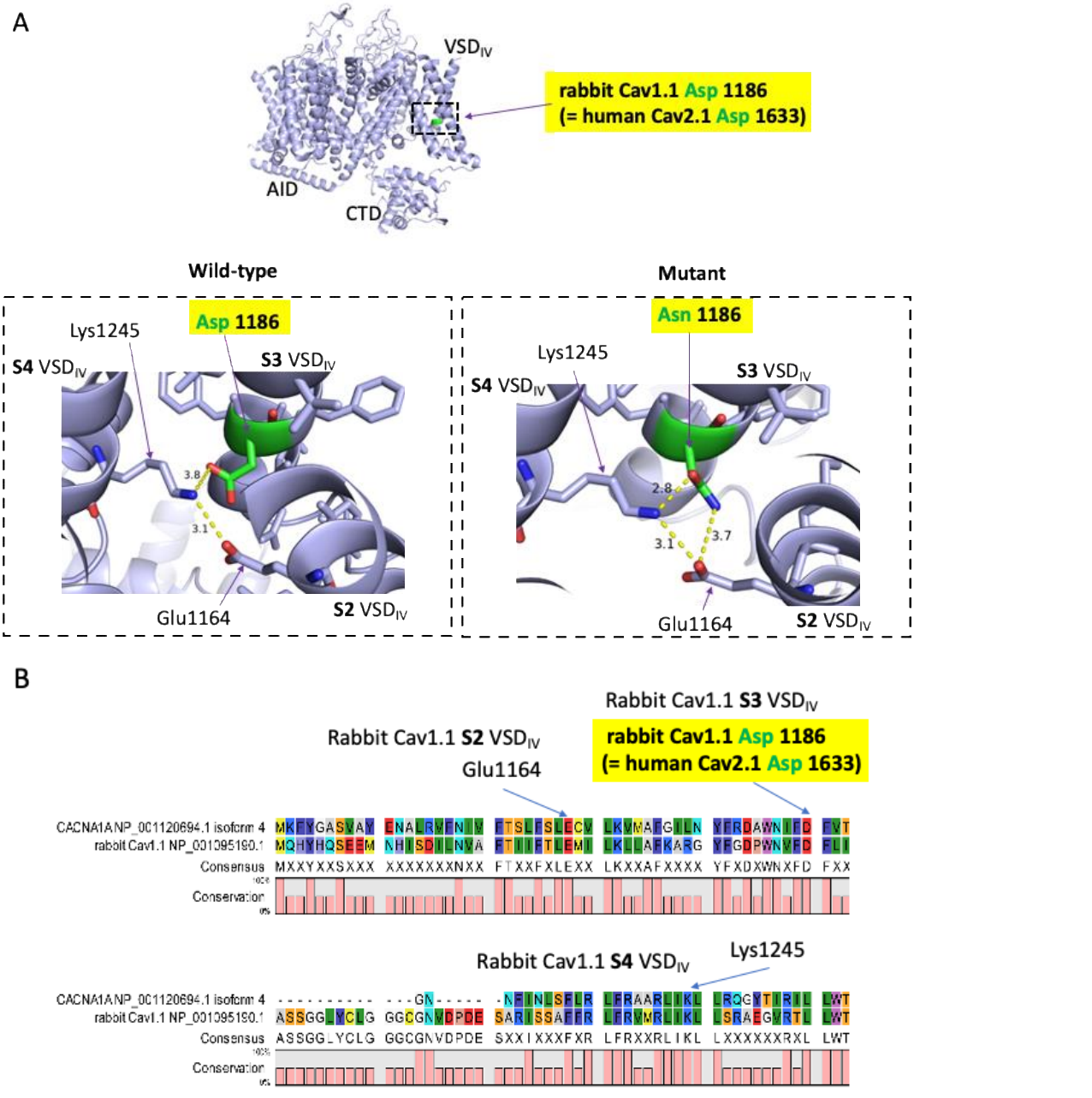
Primers for *NEU1* in exon 2 and exon 5 as well as adjacent non-deleted control gene C6orf48 were normalized to the single-copy gene β-microglobulin (B2M) using the ΔΔCt method in DNA from patient PME10, his affected brother and carrier father compared to controls. qPCR was performed using the IQ SybrGreen kit (Bio-Rad) on a CFX96 Touch qPCR system (Bio-Rad). Primer efficiencies and their linear range were determined by serially diluted genomic DNA and the presence of any unspecific amplification was excluded by melting curve analysis and agarose gel electrophoresis. All reactions were performed in triplicates.

Figure S6. **Chr1q23.2 haplotype encompassing *PEX19* c.254C>T (p.A85V) variant.**



Shared homozygous-by-descent haplotype (pink) found in the three patients of Maltese origin with *PEX19* variants. The haplotype length shared between the two unrelated families is much smaller (~1cM) consistent with a distant common ancestor.

Figure S7: **Molecular modelling supports CACNA1A p.Asp1633Asn variant loss-of-function effect**.



Symbols and abbreviations: Yellow dashed lines with number: distance between residues in Å; Red sticks: oxygen atoms; Blue sticks: nitrogen atoms; VSD: voltage sensor domain; CTD: C-terminal domain; AID: α1-interacting domain; S2, S3, S4: segments 2, 3, and 4. Residues involved in the interactions shown in panel (A) are marked by arrows in panel (B).

CACNA1A p.Asp1633 represents a conserved residue. The human sequence either side of Asp1633 is homologous with the rabbit Cav1.1 channel, enabling Pymol modelling of the structural impact of the p.Asp1633Asn variant identified in patient PME16. **(A)** Homology modelling of the human Cav2.1 Asp1633Asn mutation using the cryo-EM structure of the rabbit Cav1.1 channel Wu et al (2015) Science 350: aad2395-aad2395, and (2016) Nature 537: 191-196 - PDB accession number 3JBR **(B)** Amino acid sequence alignment of the of the human Cav2.1 channel (GenBank NM_001127222.1; Protein ID = NP_001120694.1) and the rabbit Cav1.1 channel (protein ID = NP_001095190.1), using CLC sequence Viewer 7.7 (Qiagen, Aarhus, Denmark).

In the wild type channel, Asp 1633 is located in segment 2 of the voltage sensor domain IV (VSD_{IV}). In the human Cav2.1 channel, Asp 1633 corresponds to Asp 1186 of the rabbit Cav1.1 channel. Asp 1186 has a negatively charged sidechain, which interacts with the positively charged sidechain of Lysine 1245; Lys 1245 may also interact with the Glutamic acid (Glu) 1164; repulsion may occur between Asp 1186 and Glu1164.

In the mutant channel, the acidic Asp residue (with negatively charged sidechain) is replaced by the polar/neutral Asn 1186 residue. Asn 1186 may interact with both Lys 1245 and Glu 1164; whereas the polar interaction between Lys 1245 and Glu 1164 (that exists also in the wild-type channel) should remain unaffected. It's likely that the Asp1186Asn mutation (equivalent with Asp1633Asn) stabilises the interaction between the S4 and the S3 segments in VSD_{IV}. Because of the increased interaction between S3-S4, the mutation may compromise activation gating. As a result, the typical vertical (outward) movements of the S4 segment during activation may be impeded, leading to loss-of-function.

Consistent with the above structural modelling, a Web-based machine learning model, capable of predicting loss-of-function (LoF) or gain-of-function effects in voltage gated calcium channels (Heyne HO *et al*. *Sci Transl Med*, 2020)[2], predicted loss-of-function with a probability of 0.77, and pathogenicity with a probability of 0.87, for the p.Asp1633Asn variant.

Supplemental Tables

Table S1. **Major forms of PME with known genetic etiology.**

PME subtype	**Inheritance pattern**	**Gene(s)**	**Protein function / molecular pathway**	**Published >20 independent cases**
ULD (EPM1)	AR	*CSTB*	Inhibitor of lysosomal cysteine proteases	Yes
Lafora disease (EPM2A/B)	AR	*EPM2A, NHLRC1*	Glycogen metabolism	Yes
NCLs	AR	*TPP1, CLN3, CLN5, CLN6, MFSD8, CLN8*	Lysosomal enzymes or membrane proteins	Yes
	AD	*DNAJC5*		
AMRF (EPM4)	AR	*SCARB2*	Lysosomal membrane protein	Yes
North Sea PME (EPM6)	AR	*GOSR2*	Golgi vesicle transport	Yes
MERRF	Mitochondrial	*MT-TK^*	Mitochondrial transfer-RNA	Yes
PME (EPM3)	AR	*KCTD7*	Interaction with potassium ion channels	Yes
Sialidosis type 1	AR	*NEU1*	Lysosomal enzyme which breaks down oligosaccharides	Yes
DRPLA	AD	*ATN1*	Accumulation of ATN1 in neurons due to repeat expansion	Yes
MEAK (EPM7)	AD	*KCNC1*	Neuronal voltage-gated potassium ion channel	Yes
Juvenile Huntingtons	AD	*HTT*	Transcription regulation	Yes
Gaucher disease type 3	AR	*GBA*	Lysosomal enzyme which breaks down glycolipid glucosylceramide	Yes

^pathogenic variants in this gene accounting for ~90% of MERRF patients

Table S2. **Research variant prioritization score.**

Variant level	a) Null variant (nonsense, frameshift, canonical +/- 1 or 2 splice sites, initiation codon, deletion) b) Damaging missense (all *in silico* tools predict damaging effect)	**2**
	c) Conflicting missense (at least 1, but not all *in silico* tools predict damaging effect) d) Splicing variant (all *in silico* tools predict a splicing effect, but variant not at canonical +/- 1 or 2 sites) e) Inframe deletion	**1**
	f) Benign missense (all *in silico* tools predict benign effect) g) Conflicting or benign splicing variant (at least 1 *in silico* tool predicts no splicing effect)	**0**
Pedigree level	a) Heterozygous *de novo* variant in established dominant disease gene (i.e. parental DNA available) b) Comp het variant in established recessive disease gene (i.e. two variants *in trans*) c) Homozygous variant in established recessive disease gene with pedigree segregation and/or linkage data to support inheritance model	**2**
	d) Homozygous variant in established recessive disease gene (+/- support with F>0 / variant located in runs of homozygosity RoH) e) Heterozygous variant in established dominant disease gene inherited from affected parent f) Heterozygous *de novo* variant in gene with no established disease association g) Comp het or homozygous variant in gene with no established disease association	**1**
	h) Heterozygous variant with undetermined parental inheritance (0.5 if segregation known in single parent)	**0**
Gene level	a) Established PME gene b) Established neurological gene (e.g., epilepsy, ataxia) with clear patient phenotypic match on clinical review c) Established neurological gene with overlapping PME features with variants in multiple unrelated patients	**2**
	d) Established neurological gene (e.g., epilepsy, ataxia) with some patient phenotypic overlap on clinical review e) Gene has established biological overlap with known PME genes with variants in multiple unrelated patients (0.5 if single patient) f) Uncertain clinical/biological match with multiple unrelated patients	**1**
	g) Uncertain clinical/biological match in single patient	**0**

Table S3. **Catalogue of short tandem repeats searched for across PME cohort.**

locus	long name	OMIM	inheritance	gene	location	gene region	motif
DM1	Myotonic dystrophy 1	160900	AD	*DMPK*	19q13	3'UTR	CTG
DM2	Myotonic dystrophy 2	602668	AD	*ZNF9/CNBP*	3q21.3	intron	CCTG
DRPLA	Dentatorubral-pallidoluysian atrophy	125370	AD	*DRPLA/ATN1*	12p13.31	coding	CAG
EPM1A	Myoclonic epilepsy of Unverricht and Lundborg	254800	AR	*CSTB*	21q22.3	promotor	CCCCGCC CCGCG
FRAXA	Fragile-X site A	309550	X	*FMR1*	Xq27.3	5'UTR	CGG
FRAXE	Fragile-X site E	309548	X	*FMR2*	Xq28	5'UTR	CCG
FRDA	Friedreich ataxia	229300	AR	*FXN*	9q13	intron	GAA
FTDALS1	Amyotrophic lateral sclerosis-frontotemporal dementia	105550	AD	*C9orf72*	9p21	intron	GGGGCC
HD	Huntington disease	143100	AD	*HTT*	4p16.3	coding	CAG
HDL2	Huntington disease-like 2	606438	AD	*JPH3*	16q24.3	exon	CTG
SBMA	Kennedy disease	313200	X	*AR*	Xq12	coding	CAG
SCA1	Spinocerebellar ataxia 1	164400	AD	*ATXN1*	6p23	coding	CAG
SCA2	Spinocerebellar ataxia 2	183090	AD	*ATXN2*	12q24	coding	CAG
SCA3	Machado-Joseph disease	109150	AD	*ATXN3*	14q32.1	coding	CAG
SCA6	Spinocerebellar ataxia 6	183086	AD	*CACNA1A*	19p13	coding	CAG
SCA7	Spinocerebellar ataxia 7	164500	AD	*ATXN7*	3p14.1	coding	CAG
SCA8	Spinocerebellar ataxia 8	608768	AD	*ATXN8OS/ATXN8*	13q21	utRNA	CTG
SCA10	Spinocerebellar ataxia 10	603516	AD	*ATXN10*	22q13.31	intron	ATTCT
SCA12	Spinocerebellar ataxia 12	604326	AD	*PPP2R2B*	5q32	promotor	CAG
SCA17	Spinocerebellar ataxia 17	607136	AD	*TBP*	6q27	coding	CAG
SCA36	Spinocerebellar ataxia 36	614153	AD	*NOP56*	20p13	intron	GGCCTG
FECD3	Fuchs endothelial corneal dystrophy 3	613267	AD	*TCF4*	18q21.2	intron	CTG
FAME1	Familial adult myoclonic epilepsy 1	601068	AD	*SAMD12*	8q24	intron	TTTCA
FAME2	Familial adult myoclonic epilepsy 2	607876	AD	*STARD7*	2q11.2	intron	TTTCA
FAME3	Familial adult myoclonic epilepsy 3	613608	AD	*MARCHF6*	5p15.31-p15.1	intron	TTTCA
FAME6	Familial adult myoclonic epilepsy 6	618074	AD	*TNRC6A*	16p12.1	intron	TTTCA
FAME7	Familial adult myoclonic epilepsy 7	618075	AD	*RAPGEF2*	4q32.1	intron	TTTCA

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; X, X-linked

Supplemental Methods

```
## Brain co-Expression gene analysis ##
## .R code
# load required R packages
library(corrplot)
library(ggrepel)
library(RColorBrewer)
library(magrittr)
library(tidyverse)
library(dynamicTreeCut)
library(DescTools)
library(data.table)
library(WGCNA)
library(dendextend)
library(gProfileR)
library(RUVcorr)
library(qgraph)
options(stringsAsFactors = FALSE)
# read in matrices
brainSpan_dir <- " "
samps <- fread(paste0(brainSpan_dir, "/genes_matrix_csv/columns_metadata.csv"))
genes <- fread(paste0(brainSpan_dir, "/genes_matrix_csv/rows_metadata.csv"))
matrix <- fread(paste0(brainSpan_dir,
                       "/genes_matrix_csv/expression_matrix.csv")) %>%
  as.matrix(., rownames=1) %>%
  t

colnames(matrix) <- genes[,gene_symbol]
rownames(matrix) <- samps[,column_num]
# Cleaning data
## Identify genes and samples with an excess of missing data
## identify time points of interest.
samp.interest <- samps[,age] %like any% c("%pcw", "%mos", "%yrs")
## remove genes with excess of missing data and select samples of interest
include.matrix <- matrix[samp.interest, ]
gsg<- WGCNA::goodSamplesGenes(include.matrix,
                              minNSamples = nrow(samps)/2)
                              #tol = 1)
## summarise
gsg$goodSamples %>% table
gsg$goodGenes %>% table

## remove genes with excess of missing data and select samples of interest
clean.matrix <- include.matrix[ , gsg$goodGenes]
## Remove genes with  expresion == 0 for >50% samples
n.expressed.gt0 <- apply(clean.matrix, 2, function(x) length(x[x>0]))
keep.expressed.gt0.50pc <- n.expressed.gt0 >= nrow(clean.matrix)*0.5
keep.unique <- !duplicated(colnames(clean.matrix))
## collate list of genes to keep.
table( keep.expressed.gt0.50pc)
keep.genes <- keep.expressed.gt0.50pc & keep.unique
## filter for genes to keep and log2 transform
temp.matrix <- clean.matrix[, keep.genes]
c <-  matrix(1, nrow = nrow(temp.matrix), ncol=ncol(temp.matrix))
temp.matrix.c <- temp.matrix + c
clean.log.matrix <- log(temp.matrix.c, 2)
clean.log.matrix  %>% hist
samps.matrix  <- samps[samp.interest]
## calculate weights
samps.matrix <- samps.matrix[ , weights:=sapply(samps.matrix[,donor_id],
                                                function(x)
1/sqrt(sum(samps.matrix[,donor_id]==x)))]
# Define known and candidate PME genes
PME.genes <- c("CSTB", "KCNC1", "EPM2A", "NHLRC1", "NEU1", "GBA", "CLN6",
               "DNAJC5",
               "SCARB2", "GOSR2", "ASAH1", "KCTD7", "CERS1", "ATN1", "CLN3",
               "CLN5",
               "HTT", "TPP1", "MFSD8", "CLN8")
# MT-TK expression data not available in resource
cand.genes <- c("STUB1", "CHD2", "NUS1", "DYNC1H1", "DHDDS", "CACNA1A",
                "CAMTA1",
                "PEX19", "APOA1BP", "ALG10", "CACNA2D2", "RARS2")
all.genes <- c(PME.genes, cand.genes)
# Generate Correlation matrix
weights <- samps.matrix %>%
  .[match(rownames(clean.log.matrix), column_num)] %>%
  .[, weights]
cor.matrix <- cov.wt(clean.log.matrix,
             wt = weights,
             cor = TRUE)$cor
```

```
# Function to plot heatmap matrix
plotCandidatesCorr <- function(cor.matrix = cor.matrix,
                               candidates = all.genes,
                               colours = colours,
                               order = "hclust",
                               title = "") {
  candidates.corr <- cor.matrix[rownames(cor.matrix) %in% candidates,
                                colnames(cor.matrix) %in% candidates]
    if(order=="hclust") {
    genes.corr <- data.table(gene = colnames(candidates.corr))
    label.cols <- data.table(gene = candidates,
                             col = colours) %>%
      .[genes.corr, on = "gene", col] %>%
      .[corrMatOrder(candidates.corr, order= order, hclust.method= "median")]
    corrplot(candidates.corr,
             order = "hclust",
             hclust.method= "median",
             method = "color",
             addrect = 6,
             tl.col = label.cols,
             tl.cex = 0.8,
             tl.srt = 60)
  } else  {
    if(order=="original") {
      genes.corr <- candidates1[candidates1 %in% colnames(candidates.corr)]
      candidates.corr <- candidates.corr[genes.corr , genes.corr]

      label.cols <- data.table(gene = candidates,
                               col = colours) %>%
        .[gene %in% genes.corr, col]
  } else {
    genes.corr <- data.table(gene = colnames(candidates.corr))
    label.cols <- data.table(gene = candidates,
                             col = colours) %>%
      .[genes.corr, on = "gene", col] %>%
      .[corrMatOrder(candidates.corr, order= order)]
    }
    corrplot(candidates.corr,
             order = order,
             method = "color",
             tl.col = label.cols,
             tl.cex = 0.8,
             tl.srt = 60,
             title = title)
}
}
# Generate heatmap figure
plotCandidatesCorr(cor.matrix = cor.matrix,
                   candidates = unique(c(PME.genes, cand.genes)),
                   colours = c(rep("black", times=length(PME.genes)),
                               rep("dimgrey", times=length(cand.genes))))
# Test for evidence of excessive co-expression
## P-value
calcCoExpPval <- function(candidates=all.genes, cor.matrix=cor.matrix) {
    candsVal <- cor.matrix[rownames(cor.matrix) %in% candidates,
                           colnames(cor.matrix) %in% candidates] %>%
    replace(., lower.tri(., TRUE), NA) %>%
    melt %>%
    as.data.table %>%
    setnames(., c("gene1", "gene2", "corr")) %>%
    .[gene1 != gene2] %>%
    .[,corr] %>%
    na.omit %>%
    abs %>%
    median
randomVals <- list()
  for (i in 1:5000) {
    random <- sample(colnames(cor.matrix),
                     length(candidates[candidates %in%
                                         colnames(cor.matrix)]))
    randomVals[[i]] <- cor.matrix[rownames(cor.matrix) %in% random,
                                  colnames(cor.matrix) %in% random] %>%
      replace(., lower.tri(., TRUE), NA) %>%
      melt %>%
      as.data.table %>%
      setnames(., c("gene1", "gene2", "corr")) %>%
      .[gene1 != gene2] %>%
      .[,corr] %>%
      na.omit %>%
      abs %>%
      median
    }
```

```
    sortedVals <- randomVals %>%
    unlist %>%
    sort
  pVal <- 1-ecdf(sortedVals)(candsVal)
  return(pVal)
}
calcCoExpPval(candidates=all.genes, cor.matrix=cor.matrix)
```

Supplemental References

1. Stanley P, Taniguchi N, Aebi M. N-Glycans. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, *et al.*, editors. Essentials of Glycobiology. Cold Spring Harbor (NY); 2015. p. 99-111.
2. Heyne HO, Baez-Nieto D, Iqbal S, Palmer DS, Brunklaus A, *et al*. Predicting functional effects of missense variants in voltage-gated sodium and calcium channels. Sci Transl Med. 2020;12(556).