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.**

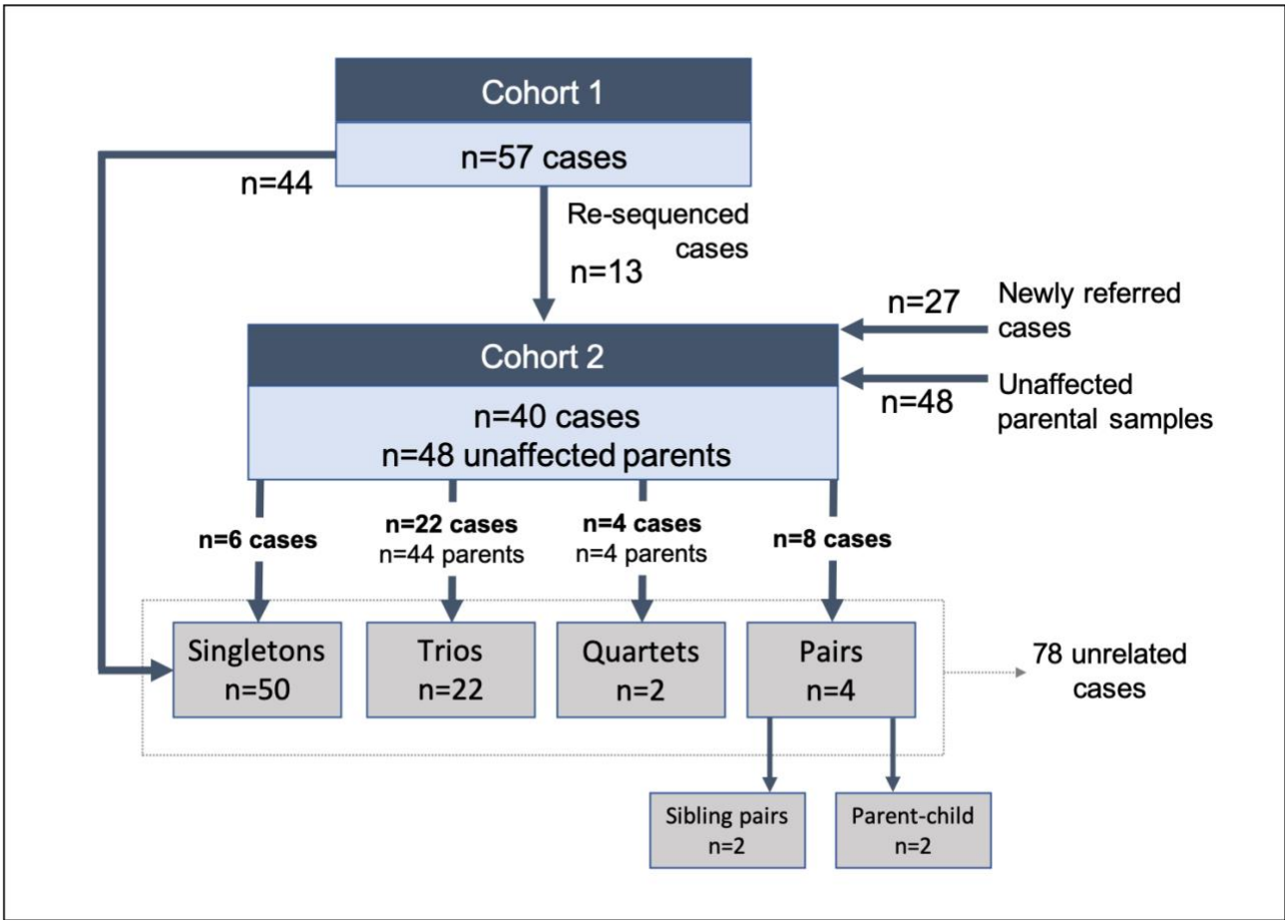


Figure S2. **Age of PME onset distribution for all 78 unrelated probands.**

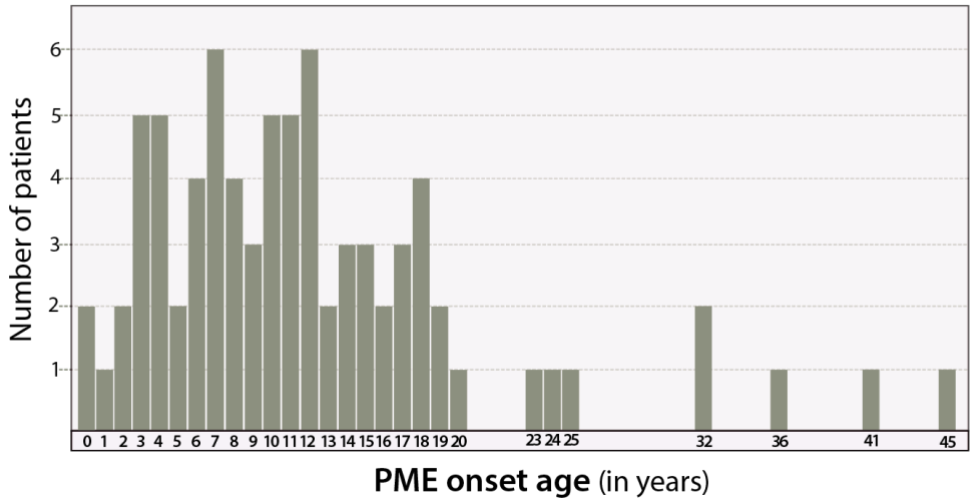
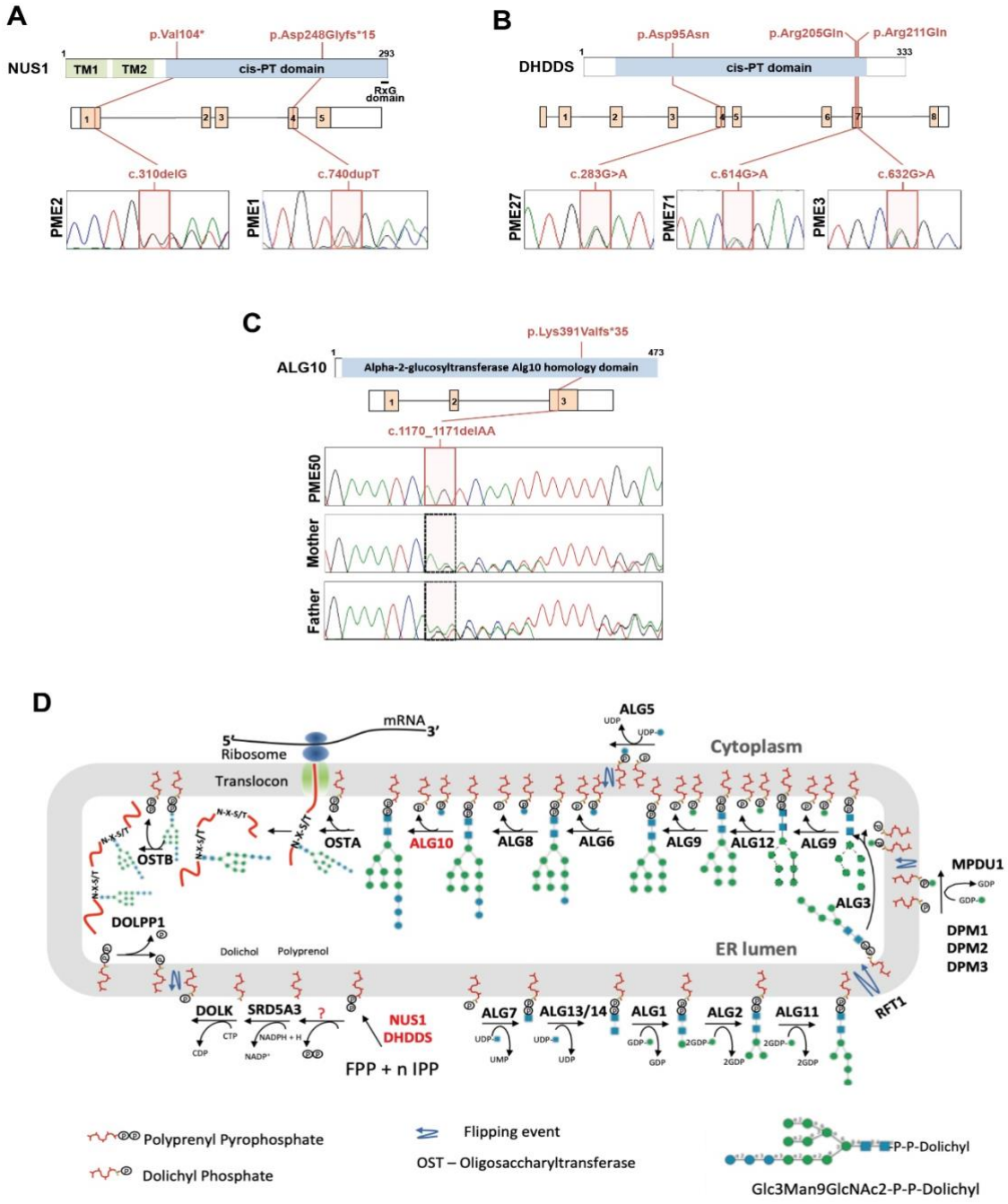


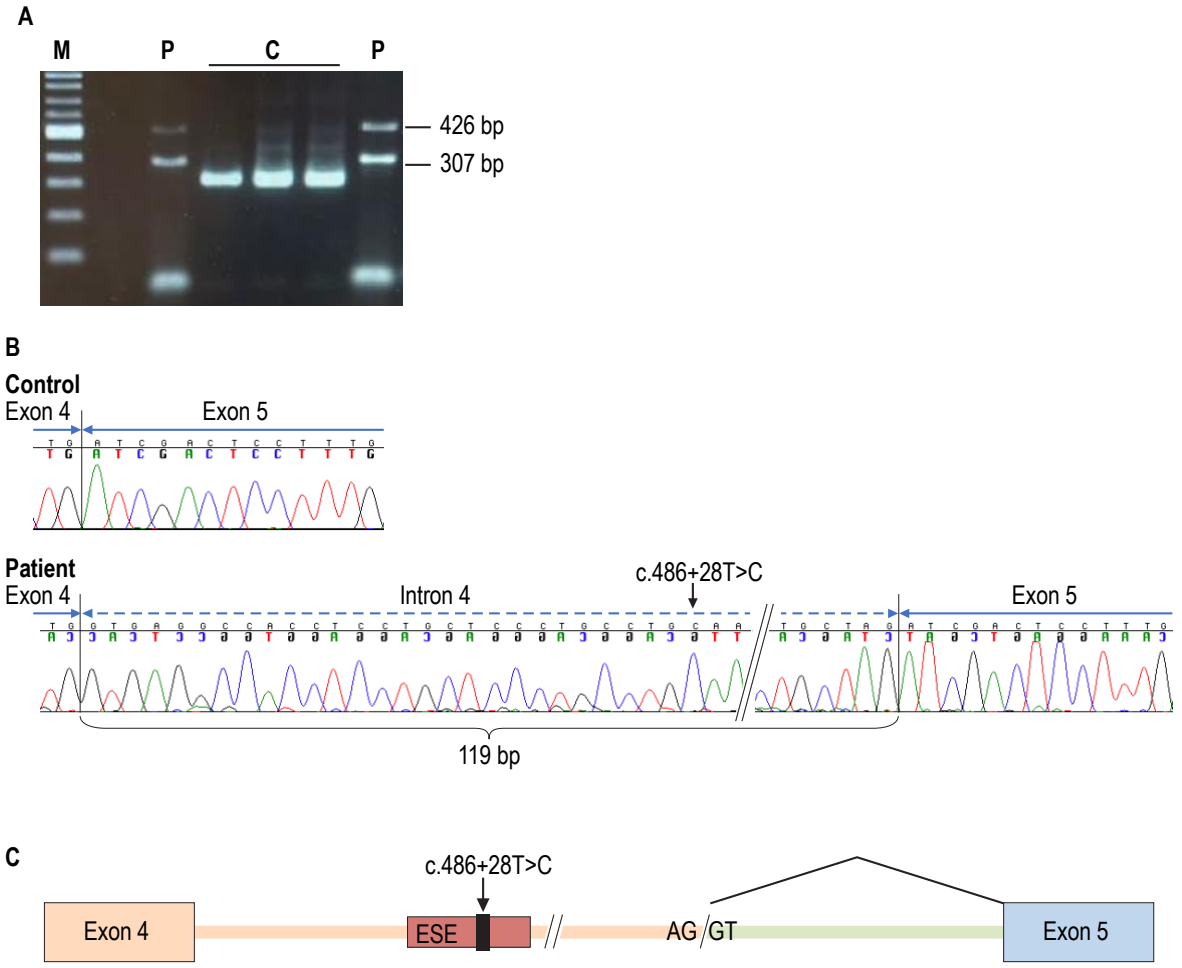
Figure S3. **Pathogenic variants in *NUS1*, *DHDDS* and *ALG10* and dolichol-dependent glycosylation pathway.**



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

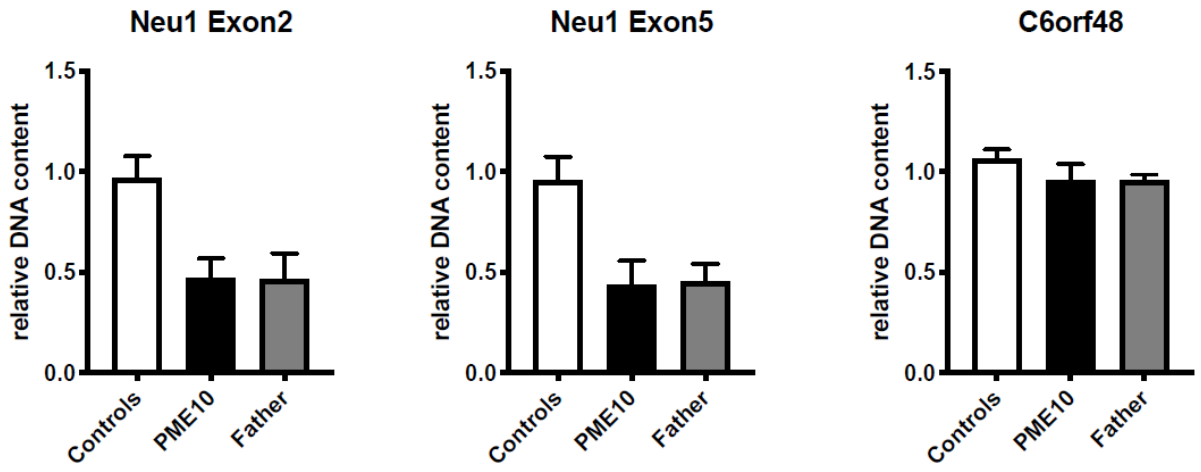
(**A**) Locations of variants in *NUS1*. (**B**) Locations of variants in *DHDDS*. (**C**) Locations of variant in *ALG10.* (**D**) Glycosylation pathway showing involvement of NUS1, DHDDS and ALG10 (in red) (adapted from Stanley P, Taniguchi N, Aebi M. N-Glycans. In: Varki A, Cummings RD, Esko JD, Stanley 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.**



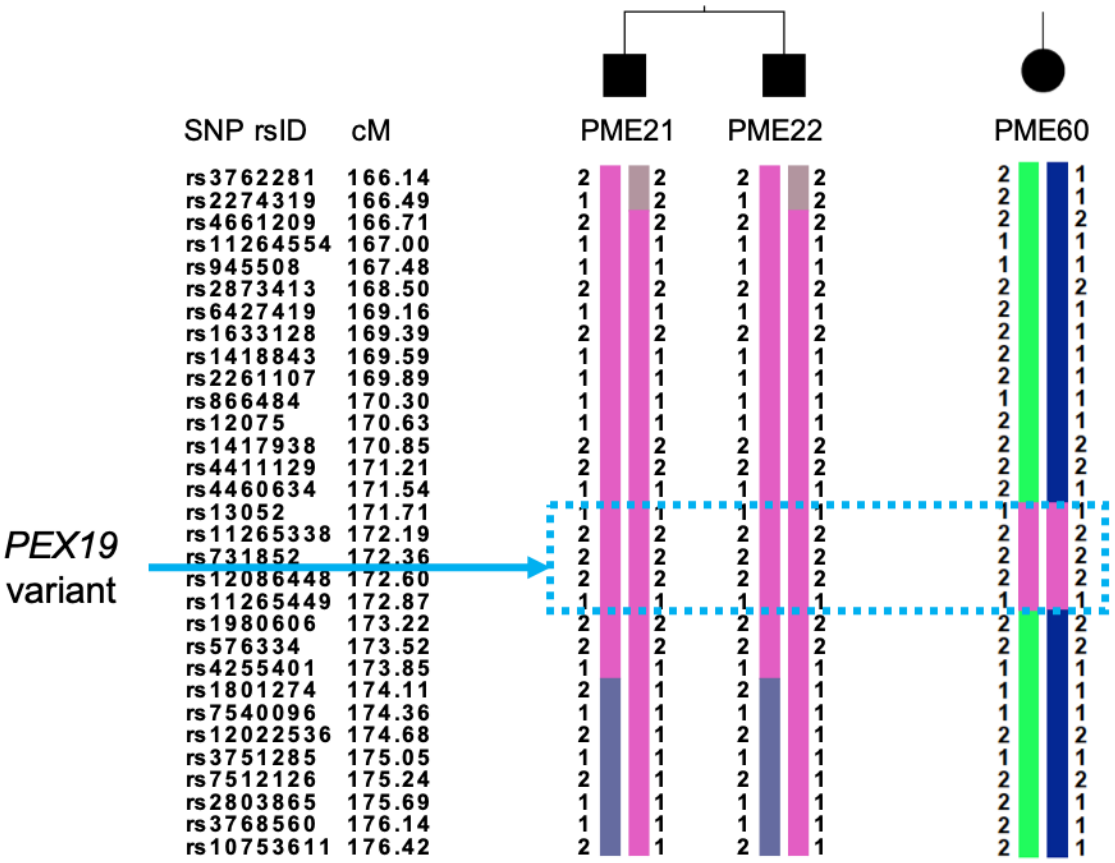
(**A**) Agarose gel electrophoresis showing the migration of RT-PCR products amplified from total RNA extracted from patient fibroblast cells using primers from exons 4 and 6 of *CLN6*. From patient (P) cells two fragments are amplified. Controls (C) show one strong amplicon. The sizes of the two fragments 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.**



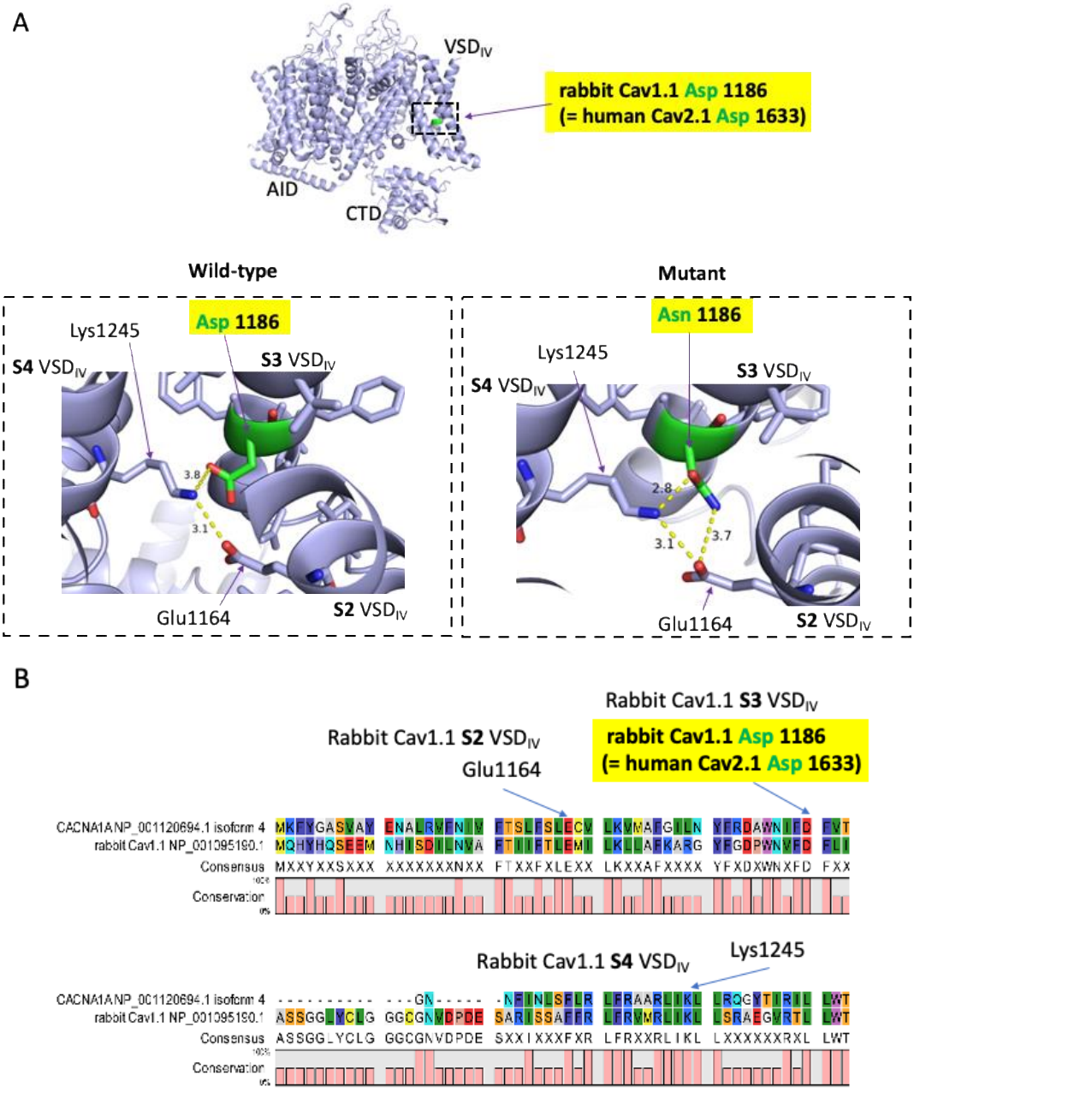
Primers for *NEU1* in exon 2 and exon 5 as well as adjacent non-deleted control gene C6orf48 were normalized to the single-copy gene β-microglobulin (B2M) using the ΔΔCt method in DNA from patient PME10, his affected brother and carrier father compared to controls. qPCR was performed using the IQ SybrGreen 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).