BRIEF COMMUNICATION

A recurrent, homozygous EMC10 frameshift variant is associated with a syndrome of developmental delay with variable seizures and dysmorphic features

Diane D. Shao1,2, Rachel Straussberg3,4,24, Hind Ahmed5,24, Amjad Khan5, Songhai Tian5, R. Sean Hill2,7, Richard S. Smith2, Amar J. Majmundar8, Najim Ameziane9, Jennifer E. Neji2,7, Edward Yang10, Amal Al Tenaiji11, Saumya S. Jamuar12,13, Min Dong6, Guntram Borck21,23, Christian Beetz9,25, Lihadh Al-Gazali22,25, Wafaa Eyaid5,25 and Christopher A. Walsh2,7,25

INTRODUCTION

The endoplasmic reticulum membrane complex (EMC) consists of multiple proteins that are highly conserved across eukaryotes.1 This complex has been shown to play a critical role as a transmembrane protein insertase, facilitating the proper insertion of some tail-anchored membrane proteins and multipass transmembrane proteins.2,3 Of the ten proteins that form the human EMC, only variants in EMC1 have previously been associated with a genetic syndrome that includes global developmental delay (GDD), cerebellar atrophy, seizures, microcephaly, and vision abnormalities4,5 (OMIM 616875).

In this study, we report 13 individuals from seven consanguineous nuclear families who are affected with a syndromic phenotype including GDD, intellectual disability (ID), variable seizures, and variable dysmorphic features including a long face, curly hair, cubitus valgus, and arachnodactyly. This phenotype segregated with a homozygous EMC10 frameshift variant that appears to be a mutational hotspot. Using in vitro studies, we provide additional evidence for the deleterious effect of this EMC10 variant.

MATERIALS AND METHODS

Clinical presentation was assessed by a clinical geneticist from one of the participating clinical centers, and informed consent for publication of individual photos was also obtained. Exome, genome sequencing, and/or single-nucleotide polymorphism

PURPOSE: The endoplasmic reticulum membrane complex (EMC) is a highly conserved, multifunctional 10-protein complex related to membrane protein biology. In seven families, we identified 13 individuals with highly overlapping phenotypes who harbor a single identical homozygous frameshift variant in EMC10.

METHODS: Using exome, genome, and Sanger sequencing, a recurrent frameshift EMC10 variant was identified in affected individuals in an international cohort of consanguineous families. Multiple families were independently identified and connected via Matchmaker Exchange and internal databases. We assessed the effect of the frameshift variant on EMC10 RNA and protein expression and evaluated EMC10 expression in normal human brain tissue using immunohistochemistry.

RESULTS: A homozygous variant EMC10 c.287delG (Refseq NM_206538.3, p.Gly96Alafs*9) segregated with affected individuals in each family, who exhibited a phenotypic spectrum of intellectual disability (ID) and global developmental delay (GDD), variable seizures and variable dysmorphic features (elongated face, curly hair, cubitus valgus, and arachnodactyly). The variant arose on two founder haplotypes and results in significantly reduced EMC10 RNA expression and an unstable truncated EMC10 protein.

CONCLUSION: We propose that a homozygous loss-of-function variant in EMC10 causes a novel syndromic neurodevelopmental phenotype. Remarkably, the recurrent variant is likely the result of a hypermutable site and arose on distinct founder haplotypes.

Genetics in Medicine (2021) 23:1158–1162; https://doi.org/10.1038/s41436-021-01097-x
(SNP) array, were performed either through clinical diagnostic testing at Centogene and/or through research settings. See Supplemental Material for institution-specific gene discovery methods. Collaborators were connected via Matchmaker Exchange and existing scientific networks. Genome-wide linkage analysis was performed to generate a logarithm of the odds (LOD) score from SNP array data using Merlin under a recessive mode of inheritance assuming a disease allele prevalence of 0.0001 and full penetrance. Haplotype analysis from exome and genome sequencing data considered only variants in regions that are covered by both genome and exome sequencing. A 2-Mb region up and downstream of the relevant EMC10 variant was interrogated (chr19: 48981900–52961200 [hg19]) and filtered for high quality homozygous variants.

Further details of genetic and experimental methods can be found in Supplemental Material.

RESULTS
Genetic findings
Using exome or genome sequencing, we identified a biallelic EMC10 frameshift variant at RefSeq NM_206538.3: c.287delG (p. Gly96Alafs*9), in all affected individuals with shared phenotype in seven consanguineous families of Bedouin, Saudi Arabia, and United Arab Emirates origin (Fig. 1a). None of the individuals had other rare variants predicted to alter gene function in any previously reported genes associated with neurodevelopmental conditions. There are no individuals who are homozygous for the EMC10 c.287delG variant in human reference databases such as gnomAD, GenomAsia 100K Project, and the Greater Middle East Variome. No individuals with biallelic loss-of-function variants in EMC10 were identified by comprehensively searching Centogene’s disease-associated variant database CentoMD, which contains data from >80,000 individuals with hereditary disorders analyzed by exome or genome sequencing. Sanger sequencing confirmed segregation of the EMC10 variant with disease in all families, including the extended family tree of affected individuals in families 1 and 2 (Fig. S1).

Genome-wide linkage analysis to the phenotype of intellectual disability determined a maximum LOD score of 6.49 on chromosome 19, consistent with the location of EMC10 (Fig. 1b). There were no other genomic regions that exhibited significant linkage. Linkage analysis using only array data for affected individuals (thus removing the assumption that siblings, who were not directly assessed, are truly unaffected) showed linkage to the same region. Regions of homozygosity (ROH) were also reviewed from SNP array, exome, or genome sequencing data, and ranged from 1.1 Mb to 2.8 Mb. The consensus ROH was ~225 kB in size (hg19 chr19: 50789967–51015404), in which the only shared coding variant was in EMC10 (Fig. 1c).

Because the variant was identical in multiple unrelated families, we looked specifically at whether there was a founder effect (i.e., a single shared haplotype) or a potential hotspot for genetic variation (i.e., a variant that arose in multiple haplotypes). Haplotype analysis clearly showed two distinct haplotypes based on SNPs from exome sequencing (Fig. 1d). SNP array data also supported the presence of two distinct haplotypes, and included family 6 for whom exome data was unavailable (Fig. S2). Haplotypes were shared by affected individuals of families 1 and 2 who are second cousins, and a separate haplotype was identified in families 3 through 7.

EMC10 c.287delG variant in human reference databases such as gnomAD, GenomAsia 100K Project, and the Greater Middle East Variome. No individuals with biallelic loss-of-function variants in EMC10 were identified by comprehensively searching Centogene’s disease-associated variant database CentoMD, which contains data from >80,000 individuals with hereditary disorders analyzed by exome or genome sequencing. Sanger sequencing confirmed segregation of the EMC10 variant with disease in all families, including the extended family tree of affected individuals in families 1 and 2 (Fig. S1).

Genome-wide linkage analysis to the phenotype of intellectual disability determined a maximum LOD score of 6.49 on chromosome 19, consistent with the location of EMC10 (Fig. 1b). There were no other genomic regions that exhibited significant linkage. Linkage analysis using only array data for affected individuals (thus removing the assumption that siblings, who were not directly assessed, are truly unaffected) showed linkage to the same region. Regions of homozygosity (ROH) were also reviewed from SNP array, exome, or genome sequencing data, and ranged from 1.1 Mb to 2.8 Mb. The consensus ROH was ~225 kB in size (hg19 chr19: 50789967–51015404), in which the only shared coding variant was in EMC10 (Fig. 1c).

Because the variant was identical in multiple unrelated families, we looked specifically at whether there was a founder effect (i.e., a single shared haplotype) or a potential hotspot for genetic variation (i.e., a variant that arose in multiple haplotypes). Haplotype analysis clearly showed two distinct haplotypes based on SNPs from exome sequencing (Fig. 1d). SNP array data also supported the presence of two distinct haplotypes, and included family 6 for whom exome data was unavailable (Fig. S2). Haplotypes were shared by affected individuals of families 1 and 2 who are second cousins, and a separate haplotype was identified in families 3 through 7.

Fig. 1 EMC10 variant segregates with disease phenotype in multiple affected families. (a) Pedigrees of affected families. Affected individuals in families 1 and 2 are second cousins. Affected individuals in families 5 and 6 are first cousins. Solid black, affected. Genotypes, where indicated, represent results of evaluation for the EMC10 c.287delG variant by Sanger sequencing. (b) Genome-wide logarithm of the odds (LOD) score distribution. (c) Affected individuals share a region of homozygosity on chromosome 19 (boxed), overlapping the location of EMC10 variant. Single-nucleotide polymorphism (SNP) array data for affected individuals in families 1, 2, 3, and 6 are shown. Homozygous SNPs are displayed in red or blue. Heterozygous SNPs are displayed in green. (d) Haplotypes based on SNPs determined from sequencing data in the consensus region of homozygosity (chr19: 50789967–51015404) indicate that EMC10 variant arose on two distinct haplotypes. SNP array independently confirmed two haplotypes (Fig. S2).
Clinical characteristics

Clinical findings for all 13 affected individuals show a core phenotype of GDD/ID and, to a lesser extent, dysmorphic features and seizures (Table S1). Facial dysmorphisms described in multiple individuals include a long face, pointed chin, and curly hair, although evaluation by several dysmorphologists did not concur on a consistent facial gestalt (Fig. 2a). Limb anomalies included cubitus valgus (6/13), arachnodactyly (3/13), and bilateral 5th digit clinodactyly (1/13). Most individuals exhibited GDD in domains including social, motor, language, and cognitive, and/or ID (11/12). Individual II-1 in family 7 was age 3 months at ascertainment; thus most milestones could not be assessed. Seizures were noted in 6/13 individuals, typically during childhood or in the neonatal period, and included multifocal as well as generalized tonic–clonic seizures. The majority of affected individuals who underwent brain magnetic resonance imaging (MRI) had abnormal findings (9/10); however, findings were individually nonspecific, including cerebellar tonsillar ectopia or Chiari I (4/12), a thin corpus callosum (3/10), and white matter signal abnormalities (3/10) (Fig. 2b; Table S2). Neurologic symptoms appeared to be static or nonprogressive.

Additional minor features included failure to thrive (4/13), umbilical and inguinal hernias (5/13), and ventricular septal
disruption. Changes in EMC function are likely to decrease, but not abolish, the insertion and proper function of multiple transmembrane proteins. Indeed, a survey of 61 proteins dependent upon the EMC in proteomic studies (Tian et al.20 and Shurtleff et al.21) revealed that many have been independently implicated in human neurodevelopmental diseases (Table S5). EMC10 is ubiquitously expressed in the body including in the brain, kidney, gastrointestinal tract, and musculoskeletal tissue22 thus it is not surprising that the disease phenotype also involves multiple organs.

In summary, we implicate EMC10 as a gene whose disruption leads to a human neurodevelopmental syndrome. The systemic nature of the phenotype highlights the pleiotropic roles of the EMC. Open questions remain in terms of the variability of the phenotype despite a single recurrent variant and the functional role of EMC10 in different organ systems.

**DATA AVAILABILITY**

De-identified materials, data sets, and protocols are available upon request. The reported variant was submitted to ClinVar: [S18862021].

Received: 7 August 2020; Revised: 28 December 2020; Accepted: 4 January 2021.
Published online: 2 February 2021
REFERENCES

1. Wideman, J. G. The ubiquitous and ancient ER membrane protein complex (EMC): tether or not? F1000Research. 4, 624 (2015).
2. Chitwood, P. J., Juszkiewicz, S., Guna, A., Shao, S. & Hegde, R. S. EMC is required to initiate accurate membrane protein topogenesis. Cell. 175, 1507–1519.e16 (2018).
3. Guna, A., Volkmar, N., Christianson, J. C. & Hegde, R. S. The ER membrane protein complex is a transmembrane domain inserter. Science. 359, 470–473 (2018).
4. Harel, T. et al. Monocallelic and biallelic variants in EMC1 identified in individuals with global developmental delay, hypotonia, scoliosis, and cerebellar atrophy. Am. J. Hum. Genet. 98, 562–570 (2016).
5. Abu-Saifeh, L. et al. Autozoyme-guided exome sequencing in retinal dystrophy patients reveals pathogenic mutations and novel candidate disease genes. Genome Res. 23, 236–247 (2013).
6. Philippakis, A. A. et al. The Matchmaker Exchange: a platform for rare disease gene discovery. Hum. Mutat. 36, 915–921 (2015).
7. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 536, 285–291 (2016).
8. The GenomeAsia 100K. Project enables genetic discoveries across Asia. Nature. 576, 106–111 (2019).
9. Greater Middle East Variome. http://igm.ucsd.edu/gme/index.php (2020).
10. Cortese, A. et al. Biallelic mutations in SORD cause a common and potentially treatable hereditary neuropathy with implications for diabetes. Nat. Genet. 52, 473–481 (2020).
11. Ghosh, S. G. et al. Recurrent homozygous damaging mutation in TMX2, encoding a protein disulfide isomerase, in four families with microlissencephaly. J. Med. Genet. 57 (2020).
12. Dickinson, M. E. et al. High-throughput discovery of novel developmental phenotypes. Nature. 537, 508–514 (2016).
13. Diamantopoulou, A. et al. Loss-of-function mutation in Mirt22/Emc1 rescues specific schizophrenia-related phenotypes in a mouse model of the 22q11.2 deletion. Proc. Natl. Acad. Sci. U. S. A. 114, E6127–E6136 (2017).
14. Zhou, Y. et al. EMC10 governs male fertility via maintaining sperm ion balance. J. Mol. Cell. Biol. 10, 503–514 (2018).
15. Reboll, M. R. et al. EMC10 (Endoplasmic Reticulum Membrane Protein Complex Subunit 10) is a bone marrow–derived angiogenic growth factor promoting tissue repair after myocardial infarction. Circulation. 136, 1809–1823 (2017).
16. Tian, S. et al. Proteomic analysis identifies membrane proteins dependent on the ER membrane protein complex. Cell. Rep. 28, 2517–2526.e5 (2019).
17. Shurtleff, M. J. et al. The ER membrane protein complex interacts cotranslationally to enable biogenesis of transmembrane protein complexes. Elife. 7, e37018 (2018).
18. Human Protein Atlas. http://www.proteinatlas.org (2020).

AUTHOR CONTRIBUTIONS

Conceptualization: D.D.S., R.S., H.A., C.B., L.A., W.E., C.A.W. Investigation: R.S., H.A., A.K., A.A.T., M.A., A.A., L.B., A.Z.A., S.S.J., J.L., L.G., L.A., W.E. Data curation: D.D.S., A.K., J.E.N., A. J.B., and A.J.M. Validation and methodology: D.D.S., R.S.H., N.A.A., S.T., T.M.S., R.S.S., and L.M.R. Formal analysis: A.J.M., R.S.H., N.A., I.H., G.E., J.A.B., P.B., S.S., A.J.M., F.H., G.B., and C.B. Supervision: F.H., G.B., M.D., C.B., L.A., W.E., C.A.W. Visualization: E.Y. and A.J.B., D.D. S., R.S., N.A., S.T., A.J.M., and J.E.N. Writing—original draft: D.D.S. Writing—reviewing and editing: all authors.

FUNDING

C.A.W. is supported by National Institute of Neurological Disorders and Stroke (NINDS) (RO1 35129) and is an Investigator of the Howard Hughes Medical Institute. This research was supported by the Allen Discovery Center program, a Paul G. Allen Frontiers Group advised program of the Paul G. Allen family Foundation. D.D.S. is supported by NINDS Neurology Resident Research Education Program (R25 NS070682). R.S.S. is supported by National Institute of Health (NIH) K99NS121604. A.J.M. acknowledges support from NIH Training Grant (T32DK-007726), the 2017 Postdoctoral Fellowship Grant from the Harvard Stem Cell Institute, and the American Society of Nephrology Lippis Research Program 2018 Polycystic Kidney Disease Foundation Jared J. Grantham Research Fellowship. M.D. acknowledges support from NIH (R01NS080833 and R21NS106159), the NIH-funded Harvard Digestive Disease Center (P30DK034854) and BCH Intellectual and Developmental Disabilities Research Center (P30HD18655). M. D. holds the Investigator in the Pathogenesis of Infectious Disease Award from the Burroughs Wellcome Fund. F.H. is the William E. Harmon Professor of Pediatrics and this work was supported by grants from the NIH DK-068306.

ETHICS DECLARATION

Individuals presented herein were identified and evaluated in a clinical setting, and biological samples were collected after obtaining written informed clinical and/or research consent according to protocols approved by their respective institutional review boards (IRBs) (BCH, United Arab Emirates University, Schneider Children’s Medical Center, and Prince Sultan Military Medical City). Postmortem human brain tissue was obtained from the University of Maryland Brain and Tissue Bank of the NIH NeuroBioBank obtained according to IRB for the University of Maryland and NIH Neurobiobank policies. Consent was obtained for publication of photographs with exclusion of eyes for de-identification.

COMPETING INTERESTS

N.A., I.H., P.B., and C.B. are current or former employees of Centogene AG, Rostock, Germany. The other authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41436-021-01097-x

Correspondence and requests for materials should be addressed to C.A.W.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.