








Somatic variants in diverse genes leads to a spectrum of focal cortical malformations

Dulcie Lai,^{1,†} Meethila Gade,^{1,†}  Edward Yang,² Hyun Yong Koh,^{3,4} Jinfeng Lu,¹ Nicole M. Walley,⁵ Anne F. Buckley,⁶ Tristan T. Sands,^{7,8} Cigdem I. Akman,⁸  Mohamad A. Mikati,^{9,10}  Guy M. McKhann,¹¹ James E. Goldman,¹² Peter Canoll,¹² Allyson L. Alexander,¹³ Kristen L. Park,¹⁴ Gretchen K. Von Allmen,^{15,16} Olga Rodziyevska,¹⁶ Meenakshi B. Bhattacharjee,¹⁷ Hart G. W. Lidov,¹⁸ Hannes Vogel,¹⁹  Gerald A. Grant,²⁰ Brenda E. Porter,²¹ Annapurna H. Poduri,^{3,4,†} Peter B. Crino^{22,†} and  Erin L. Heinzen^{1,23,†}

[†]These authors contributed equally to this work.

Post-zygotically acquired genetic variants, or somatic variants, that arise during cortical development have emerged as important causes of focal epilepsies, particularly those due to malformations of cortical development. Pathogenic somatic variants have been identified in many genes within the PI3K-AKT-mTOR-signalling pathway in individuals with hemimegalencephaly and focal cortical dysplasia (type II), and more recently in *SLC35A2* in individuals with focal cortical dysplasia (type I) or non-dysplastic epileptic cortex. Given the expanding role of somatic variants across different brain malformations, we sought to delineate the landscape of somatic variants in a large cohort of patients who underwent epilepsy surgery with hemimegalencephaly or focal cortical dysplasia. We evaluated samples from 123 children with hemimegalencephaly ($n = 16$), focal cortical dysplasia type I and related phenotypes ($n = 48$), focal cortical dysplasia type II ($n = 44$), or focal cortical dysplasia type III ($n = 15$). We performed high-depth exome sequencing in brain tissue-derived DNA from each case and identified somatic single nucleotide, indel and large copy number variants. In 75% of individuals with hemimegalencephaly and 29% with focal cortical dysplasia type II, we identified pathogenic variants in PI3K-AKT-mTOR pathway genes. Four of 48 cases with focal cortical dysplasia type I (8%) had a likely pathogenic variant in *SLC35A2*. While no other gene had multiple disease-causing somatic variants across the focal cortical dysplasia type I cohort, four individuals in this group had a single pathogenic or likely pathogenic somatic variant in *CASK*, *KRAS*, *NF1* and *NIPBL*, genes previously associated with neurodevelopmental disorders. No rare pathogenic or likely pathogenic somatic variants in any neurological disease genes like those identified in the focal cortical dysplasia type I cohort were found in 63 neurologically normal controls ($P = 0.017$), suggesting a role for these novel variants. We also identified a somatic loss-of-function variant in the known epilepsy gene, *PCDH19*, present in a small number of alleles in the dysplastic tissue from a female patient with focal cortical dysplasia IIIa with hippocampal sclerosis. In contrast to focal cortical dysplasia type II, neither focal cortical dysplasia type I nor III had somatic variants in genes that converge on a unifying biological pathway, suggesting greater genetic heterogeneity compared to type II. Importantly, we demonstrate that focal cortical dysplasia types I, II and III are associated with somatic gene variants across a broad range of genes, many associated with epilepsy in clinical syndromes caused by germline variants, as well as including some not previously associated with radiographically evident cortical brain malformations.

- 1 Division of Pharmacology and Experimental Therapeutics, Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
- 2 Department of Radiology, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115, USA
- 3 Division of Epilepsy and Clinical Neurophysiology, Department of Neurology, Boston Children's Hospital, Boston, MA 02115, USA

Received December 16, 2021. Revised February 19, 2022. Accepted March 13, 2022. Advance access publication April 20, 2022

© The Author(s) 2022. Published by Oxford University Press on behalf of the Guarantors of Brain. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

- 4 Epilepsy Genetics Program, Department of Neurology, Boston Children's Hospital and Harvard Medical School, Boston, MA 02115, USA
- 5 Division of Medical Genetics, Department of Pediatrics, Duke University School of Medicine, Durham, NC 27710, USA
- 6 Department of Pathology, Duke University Medical Center, Durham, NC 27710, USA
- 7 Institute for Genomic Medicine, Columbia University Medical Center, New York, NY 10032, USA
- 8 Department of Neurology, Columbia University Medical Center, New York, NY 10032, USA
- 9 Department of Neurobiology, Duke University, Durham, NC 27708, USA
- 10 Division of Pediatric Neurology, Duke University Medical Center, Durham, NC 27710, USA
- 11 Department of Neurosurgery, Columbia University, New York Presbyterian Hospital, New York, NY 10032, USA
- 12 Department of Pathology and Cell Biology, Columbia University, New York, NY 10032, USA
- 13 Department of Neurosurgery, University of Colorado School of Medicine, Aurora, CO 80045, USA
- 14 Department of Pediatrics and Neurology, University of Colorado School of Medicine, Aurora, CO 80045, USA
- 15 Department of Neurology, McGovern Medical School, Houston, TX 77030, USA
- 16 Division of Child Neurology, Department of Pediatrics, McGovern Medical School, Houston, TX 77030, USA
- 17 Department of Pathology and Laboratory Medicine, McGovern Medical School, Houston, TX 77030, USA
- 18 Department of Pathology, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115, USA
- 19 Department of Pathology, Stanford University, School of Medicine, Stanford, CA 94305, USA
- 20 Department of Neurosurgery, Lucile Packard Children's Hospital at Stanford, School of Medicine, Stanford, CA 94305, USA
- 21 Department of Neurology and Neurological Sciences, Stanford University, School of Medicine, Stanford, CA 94305, USA
- 22 Department of Neurology, University of Maryland School of Medicine, Baltimore, MD 21201, USA
- 23 Department of Genetics, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

Correspondence to: Erin L. Heinzen
University of North Carolina at Chapel Hill
Campus Box 7361, 120 Mason Farm Road
1043 Genetic Medicine Building
Chapel Hill, NC 27599–7360, USA
E-mail: eheinzen@unc.edu

Keywords: somatic variants; epilepsy; exome sequencing

Abbreviations: CCDS=consensus coding sequence; CNV=copy number variant; FCD=focal cortical dysplasia; FCDNOS=FCD not otherwise specified; HMEG=hemimegalencephaly; ID=intellectual disability; MCD=malformations of cortical development; MOGHE=mild MCD with oligodendroglial hyperplasia in epilepsy; NLFE=non-lesional focal epilepsy; SNV=single nucleotide variant; VAF=variant allele frequency

Introduction

The cerebral cortex is assembled during embryonic development through a series of tightly regulated processes that include neuronal and glial progenitor cell proliferation, cellular differentiation and fate specification, neuronal migration and ultimately cortical organization. Disruption of any of these processes may result in a range of cortical malformations, spanning those that affect all or most of the cortex to small focal cortical lesions that may or may not be detectable with brain imaging.^{1,2} Regardless of the size of the lesion, malformations of cortical development (MCD) are commonly associated with refractory epilepsy,³ as well as intellectual disability (ID) and autism spectrum disorder.⁴

Most MCD are thought to arise from variants in genes that encode proteins essential for neurodevelopment. Pathogenic genetic variants that result in MCD can be inherited, newly acquired in the germ cell (gamete) of a parent and transmitted to the affected child and thus appearing as *de novo* variants not detected in the parents' blood- or buccal-derived DNA, or post-zygotically acquired during embryonic development, giving rise to a mosaic pattern of variant-positive and variant-negative cells and appearing as somatic variants with variant allelic frequencies (VAFs) less than 50%. Typically, *de novo* germline variants or very early (pre-gastrulation) post-zygotically acquired somatic variants lead to more diffuse

cortical malformations,^{5–7} whereas somatic variants have been commonly identified in some forms of focal MCD.^{8–10}

Focal cortical dysplasia (FCD) is a type of MCD characterized by focal disruption of cortical cytoarchitecture that is highly associated with medication-resistant epilepsy. The International League Against Epilepsy (ILAE) has provided a classification system to define FCD lesions according to pathological findings in resected brain tissue specimens. Specifically, abnormalities in radial or tangential cortical lamination are deemed Type I (FCDI), the presence of dysmorphic neurons and/or balloon cells indicate Type II (FCDII) and cortical lamination abnormalities in combination with another brain lesion indicate Type III (FCDIII).^{11,12} Recently, a new pathological entity has been described as a mild malformation of cortical development (mMCD) with oligodendroglial hyperplasia in epilepsy (MOGHE), characterized by excess Olig2 and PDGFR α -immunoreactive oligodendroglia and heterotopic white matter neurons. The histological features of this newly described pathology in resected tissue specimens may be subtle and difficult to distinguish from FCDI, mMCD or gliosis.^{13,14} It should be noted that clinically, preoperative MRI is sufficient to make the diagnosis of hemimegalencephaly (HMEG) and frequently to make a presumptive diagnosis of FCD, with FCDII frequently associated with a 'transmantle' sign consisting of abnormal signal extending from the ventricular region to the cortex. Postoperatively, a more definitive MCD diagnosis can be made that incorporates available

pathological findings, which may thus result in reclassification from the initial diagnosis. This is particularly relevant for ‘non-lesional’ (MRI-negative) cases that are then found to have pathological evidence of FCD.

Somatic variants in several genes in the PI3K-AKT-mTOR signalling pathway have been identified in individuals with FCDII and HMEG,^{8,10,15–21} and have been reported in one study to account for approximately 90% of HMEG and 60% of FCDII cases.²² Recently somatic variants in SLC35A2 have been identified in resected brain tissue found to have pathology consistent with FCDI, mMCD and MOGHE, but also in individuals with histologically normal tissue.^{9,23,24} The reasons for the variable pathologic phenotypes across studies may reflect differences in the extent of mosaicism in tissue or within specific cell types,^{9,23,25,26} as well as challenges with accurately assigning histopathological diagnoses across the subtle and highly localized findings characteristic of FCDI, mMCD and the recently suggested MOGHE, pathologies.^{14,27} To date, mosaic loss-of-function variants in SLC35A2 explain between 17% and 29% of cases of refractory neocortical epilepsy with and without FCDI pathology.^{9,23,24} SLC35A2 encodes what is believed to be the sole transporter of UDP-galactose from the cytosol to the Golgi apparatus, where it can be acted on by galactosyltransferases to generate galactose-containing glycoproteins, proteoglycans and glycolipids. Absence of this transporter is expected to have wide-reaching effects on protein and lipid glycosylation. In fact, reduced N-glycan formation has been observed in the brain tissue of individuals with pathogenic somatic SLC35A2 variants.²⁴ Before the discovery of somatic SLC35A2 variants in focal neocortical epilepsy, germline SLC35A2 variants had been associated with severe epileptic encephalopathies and intractable seizures.^{28,29}

Here we report the results of a genome-wide analysis for somatic single nucleotide variants (SNV), small insertion or deletion variants (indels) in protein-coding regions and large copy number variants (CNVs) in resected brain tissue samples from 123 individuals with epilepsy who underwent focal surgical resection for medication-resistant epilepsy and genetic sequencing of the resected tissue. Our study demonstrates that at least a portion of FCD are due to somatic, brain-tissue-specific genetic variants in genes previously implicated in epilepsy and other neurodevelopmental disorders, with the associated focal epilepsy phenotypes likely reflecting the tissue and cellular localization of the pathogenic genetic variants identified. Notably, the types of gene variants involved in FCDI and FCDIII were distinct from those characteristic of FCDII providing a new molecular genetic understanding of FCD pathogenesis.

Materials and methods

Study participants

MCD cohort

Individuals who underwent a surgical resection of cortical tissue due to intractable focal epilepsy were enrolled from Boston Children’s Hospital, Duke University Medical Center, Lucile Packard Children’s Hospital at Stanford, The University of Texas Health Science Center at Houston, Columbia University Irving Medical Center, or Children’s Hospital Colorado. Clinical assessment of the candidacy for focal resection was determined by standard clinical practice at each site, incorporating seizure semiology, EMG/EEG data, structural MRI, functional imaging (PET, SPECT), and consensus at epilepsy surgery conferences.

To be included in this study, patients were required to have no known genetic syndrome at the time of enrolment and have a radiographic phenotype that was not inconsistent with MCD, which includes cases without evident focal lesions. Exclusion criteria included encephalomalacia, Rasmussen’s encephalitis, isolated schizencephaly, neoplasm, isolated mesial temporal lobe sclerosis, grey matter heterotopia, polymicrogyria and clinically diagnosed tuberous sclerosis complex. Imaging and neuropathological data or reports were reviewed by a neuroradiologist (E.Y.), neuropathologist (H.G.W.L.) and epileptologist/neurogeneticist (A.P.) to determine eligibility for this study and classify each eligible case into one of four phenotypic categories: first, FCDI+ represents cases with focal epilepsy with MRI evidence of FCD with or without available neuropathological features of FCDI. Included in the FCDI+ group were cases with MRI lesions suggestive of FCDI and neuropathological evidence of FCDI and related subtypes, including FCDIa (FCD with abnormal radial cortical lamination, including vertical microcolumns), FCDIb (FCD with abnormal layering), FCDIc (FCD with vertical and horizontal abnormalities) and mMCD (with excessive heterotopic neurons).^{11,12} Additionally, we included a classification of FCD not otherwise specified (FCDNOS), referring to cases with MRI lesions suggestive of FCDI but with non-specific histopathology (e.g. gliosis) or no pathological abnormalities, and cases with non-lesional focal epilepsy (NLFE), defined as normal brain MRI, with FCDI or related pathology or without FCDI or related pathology (maintaining the assumption that the pathology was likely present but not observed due to incomplete sampling of the specimen). The second category was FCDII and included cases with FCDIIa (FCD with dysmorphic neurons) and FCDIIb (FCD with dysmorphic neurons and balloon cells), with or without FCD detected on MRI. The third category was FCDIII and included cases with FCDIIIa (cortical dyslamination associated with hippocampal sclerosis), FCDIIIc (cortical dyslamination adjacent to vascular malformation) and FCDIIId (cortical dyslamination adjacent to lesion acquired during early life).^{11,14} The fourth was HMEG, including cases with classical HMEG as well as cases with dysplastic megalencephaly (DMEG) in the form of large malformations that do not fully encompass an entire hemisphere or that also involve both cerebral hemispheres. The clinical diagnoses of all individuals are provided in [Supplementary Table 1](#).

All patients were consented to participate in research approved by the respective institutional review boards.

Twenty-five of the 123 cases were included in previous studies ([Supplementary Table 1](#)). Seven of the 25 were previously genetically diagnosed in those studies and four had new genetic diagnoses reported in this study ([Supplementary Table 1](#)).

Control cohort

We obtained sequencing data from brain tissue specimens obtained from autopsies from controls without neurological disorders. Three-hundred and forty-two exome-sequenced individuals were obtained from the North American Brain Expression Consortium (NABEC—dbGaP Study Accession: phs001300.v1.p1), and 15 whole genome-sequenced controls were downloaded from the Brain Somatic Mosaicism Network (National Institute of Mental Health Data Archive; [Supplementary Table 2](#)).

Specimen collection and DNA extraction

Brain tissue specimens from an epileptogenic region were collected from each of the individuals in the MCD cohort. Blood samples were available for DNA extraction for exome sequencing in 74 of the 123

individuals (60%; [Supplementary Table 3](#)). DNA was extracted from the specimens using GenFind V3 (Beckman Coulter) or DNeasy Blood & Tissue Kits (Qiagen) per the manufacturer's protocols.

Sequencing and variant calling

Exome sequencing

Brain-tissue-derived DNA samples from the 123 individuals underwent paired-end high-depth exome sequencing on an Illumina HiSeq2000, HiSeq2500 or NovaSeq at Duke University Medical Center (DUMC), Columbia University Irving Medical Center (CUIMC), Genewiz or the University of North Carolina at Chapel Hill (UNC) high-throughput sequencing facilities. The protein-coding sequences were enriched in the libraries using either Nimblegen SeqCap EZ V3.0, Agilent V6 SureSelect, IDT xGen Exome Research Panel v1, or IDT xGen Exome Research Panel v2 exome enrichment kits. We also performed exome sequencing on DNA extracted from whole blood samples for 74 individuals ([Supplementary Table 3](#)).

Alignment

Next-generation sequence data from cases and controls were aligned to the hg38 reference genome (GRCh38.d1.vd1) using Burrows-Wheeler Aligner (BWA, version 0.7.15).³⁰ Biobambam2 (version 2.0.168) and samtools (1.11)³¹ were used for marking duplicates, sorting and indexing alignment files.

Variant calling

Somatic single-nucleotide and small insertion-deletion variant calling

Somatic SNV and indel variants were called in the brain-derived DNA samples using Mutect2³² following the NCI Genomics Data Commons Somatic Variant Calling Workflow. The pipeline included an assessment of a panel of normal and germline variants compiled from the gnomAD database (gnomad.hg38.vcf.gz)³³ to reduce artefactual calls. While matched blood was also exome-sequenced in some cases, the exome-sequencing data from brain-derived DNA was used to call somatic variants to ensure uniform analysis of all study subjects. When exome sequence data from a matched leucocyte-derived DNA sample was available, brain tissue-specific somatic variants were also called in the pair and compared to the results from the brain-only calling to ensure that candidate variants were not overlooked.

High-quality somatic SNVs and indels were defined as the subset of somatic variant calls meeting the following criteria: (i) annotated to be located in a protein-coding exon or associated splice sites; (ii) flagged as 'PASS' by the Mutect2 software; (iii) called at a site with at least 20 reads covering the site; (iv) at least five reads supporting the variant allele; (v) reads supporting the variant were found on both read 1 and read 2 of the paired-end sequencing of a DNA fragment; (vi) VAF < 35% for calls on autosomes and the X chromosome of females and VAF < 70% for chromosome X variants in a male; and (vii) VAF > 2%. Variants on the Y chromosome and those where multiple different variant alleles were called in the same individual were also excluded.

Somatic SNVs and indels that did not meet the strict high-quality criteria ('permissive' variant calls) were also evaluated for candidate genetic diagnoses. These permissive calls included all somatic variants called by the Mutect2 algorithm in a protein-coding exon or splice site, excluding those suggestive of a

sequencing or alignment error (i.e. variants failing the base_qual, clustered_events, contamination, map_qual, or strand_bias filters, or variants called as multiallelic). In order to capture variants that may have been overlooked with filtering, no minimum coverage filter was applied for this permissive analysis.

Somatic SNV and indel variant calls were inspected for accurate calling using Integrative Genomics Viewer (IGV, version 2.8.7).

Somatic copy number variant calling

Somatic CNVs were called in the brain tissue-derived DNA using CNV Radar.³² This algorithm uses a panel of normal samples, read depth, and VAF patterns to detect CNVs in the absence of matched controls. Called CNVs were annotated using ANNOT-SV.³⁴ The panel of normals included blood-derived DNA samples from individuals without MCD that were sequenced using the same library and exome enrichment kits, allowing us to correct for technical biases due to bait sizes, location and hybridization conditions. This approach, rather than comparing to the study participants' blood, would allow detection of mosaic CNVs that might also be present in the blood. Given the challenges of accurately detecting somatic CNVs from exome-sequencing data, we limited these analyses only to large CNVs > 1 Mb.

Germline single nucleotide and small insertion-deletion variant calling

Germline indels and SNVs were called using GATK Best Practices recommendations.^{35,36} The process included base quality scores recalibration to generate analysis ready reads, individual level variant calling using HaplotypeCaller, indel realignment, duplicate removal and joint genotyping across the full cohort. The resultant SNV and indel calls were then hard-filtered using quality normalized by depth (QD) < 2.0 || FisherStrand (FS) > 60.0 || root mean square mapping quality (MQ) < 40.0 || mapping quality rank sum test (MQRankSum) < -12.5 || ReadPosRankSum < -8.0 and QD < 2.0 || FS > 200.0 || ReadPosRankSum < -20.0, respectively.

Quality control of exome sequence data

Somalier (version 0.2.12) was used to infer the sex and ancestry of each sequenced individual and to test for unexpected relatedness among cases and controls.³⁷ Ancestry was predicted for each subject using the data set from the 1000 Genomes Project-hg38 as a reference.³⁸ We also assessed the rate of G-T substitutions which are a well-recognized Illumina sequencing error and can also arise from oxidative DNA damage in low-quality DNA samples.^{39,40}

Average coverage and percent of bases covered at least 50-fold across the consensus coding sequence (CCDS) protein-coding regions plus two base pairs flanking exons were calculated for each sample using the GATKDepthOfCoverage tool included in the GATK analysis toolkit.

Annotation

Annotation of the somatic and germline variants was performed using ANNOVAR (v20200609).⁴¹ Annotation databases used were either downloaded directly from ANNOVAR or created using the original database ([Supplementary Table 4](#)).

Protein-coding variants analysed in this study were limited to only those located in the CCDS (CCDS version 22, hg38)⁴² and two base pair splice site regions flanking the CCDS exons.

Variant confirmation

Somatic variants were confirmed using digital PCR (dPCR), Sanger sequencing, or amplicon sequencing. Digital PCR was performed using custom-designed Taqman assays targeting the variant and reference alleles for an SNV or targeting a gene within the duplicated or deleted region for a CNV. A PCR QuantStudio 3D Digital PCR System (Thermo Fisher Scientific) and associated QuantStudio 3D AnalysisSuite Cloud Software was used to quantitatively assess allele fraction and copy number per the manufacturer's protocol.

In some cases where the candidate somatic variant had a high VAF estimated from exome sequencing or when a Taqman assay could not be designed, we confirmed variants with either: (i) PCR followed by Sanger sequencing (Genewiz); (ii) allele-specific PCR, whereby two forward primers were designed to each selectively amplify the variant or the reference sequence in the presence of the same reverse primer placed downstream of the variant; or (iii) amplicon sequencing (CDG Genomics or Genewiz), which entailed PCR amplification of a >100 bp fragment followed by targeted next-generation sequencing using an Illumina MiSeq to ~1000-fold coverage of the variant site and flanking genomic regions. All PCR reactions used primers designed using Primer3 and purchased from Integrated DNA Technologies. MyTaq HS Mix (Meridian Bioscience) was used in all PCR reactions according to the manufacturer's instructions. In one individual (neuro1410F49br), the somatic *MTOR* variant identified in the brain tissue sample was confirmed in a clinically certified laboratory with amplicon sequencing in an independent sample.

Diagnostic analyses

Somatic variant diagnoses

We evaluated the following sets of variant calls to comprehensively search for candidate somatic genetic diagnoses:

- (i) High-quality somatic single-nucleotide and indel variants (defined above) that were rare [gnomAD exome minor allele frequency (MAF) $<1 \times 10^{-5}$], putatively functional [splice-site, nonsense, missense excluded Polyphen (HumVar) benign, frameshift and non-frameshift indel] and located in an Online Mendelian Inheritance in Man (OMIM) dominant neurological disease gene, defined as any gene associated with disorders for which a neurologic phenotype entry was included in the clinical synopsis and excluding genes associated with autosomal recessive disorders (Supplementary Table 5).
- (ii) Permissive somatic single nucleotide substitution and indel variants (defined above) that were classified as pathogenic or likely pathogenic in ClinVar located in an OMIM neurological disease gene and all loss-of-function variants (splicing, nonsense, or frameshift indel) in genes classified as haploinsufficient in the ClinGen Dosage Sensitivity Map (<https://dosage.clinicalgenome.org>; Supplementary Table 5).
- (iii) Permissive somatic single nucleotide substitution and indel variants that are rare (gnomAD exome MAF $<1 \times 10^{-5}$) and putatively functional [splice-site, nonsense, missense excluded Polyphen (HumVar) benign, frameshift and nonframeshift indel] in a list of genes previously reported to harbour somatic variants in brain tissue from cases with FCD (AKT3, *DEPDC5*, *MTOR*, *PIK3CA*, *PTEN*, *RHEB*, *RPS6*, *SLC35A2*, *TSC1*, and *TSC2*) and novel genes implicated in our analysis of high-quality variants described above (*CASK*, *KRAS*, *NF1*, *NIPBL*, *PCDH19*).
- (iv) Somatic CNV calls >1 Mb.

We assessed the likelihood that each variant meeting the above criteria contributed to the corresponding patient's clinical phenotype. The genotype of those deemed as possible genetic diagnoses was then independently confirmed as a true somatic variant using an orthogonal genotyping approach (Sanger sequencing, amplicon sequencing, etc.; see above). Variants were classified according to

recently defined ClinVar variant curation rules that govern the pathogenicity assessment of somatic variants specifically related to developmental brain malformations (ClinGen Brain Malformation Expert Panel).

Germline variant diagnoses

While the primary focus of this study was to identify disease-causing somatic variants in MCD, we also evaluated the sequence data for definitive germline diagnoses as there are known germline genetic causes of some MCD. Among the high-quality germline variant calls, we searched for pathogenic or likely pathogenic variants that met American College of Medical Genetics (ACMG) guidelines,⁴³ including all variants classified as pathogenic or likely pathogenic in ClinVar and all loss-of-function variants (splicing, nonsense, or frameshift indel) in genes classified as haploinsufficient in the ClinGen Dosage Sensitivity Map. Each variant was then assessed for the likelihood that those variants contribute to the patient's phenotype using ACMG guidelines.

Gene set enrichment analyses

Gene sets analysed

The burden of variants in all cases compared to controls was assessed across multiple gene sets (Supplementary Table 5), including: (i) OMIM⁴⁴ dominant neurological disease-associated genes as defined above; (ii) OMIM non-neurological disease genes, including all OMIM morbid genes and excluding any that have a neurological phenotype entry included in the clinical synopsis; (iii) loss-of-function intolerant genes [pLI >0.9, gnomAD database (v2.0.1)]; (iv) constrained genes [missense Z score >3.09, gnomAD database (v2.0.1)]; and (v) Kyoto Encyclopedia of Genes and Genomes (KEGG) neurological pathways.⁴⁵

Coverage normalization across cases and controls

Before performing burden testing in the case and control cohorts, sample- and locus-level coverage harmonization was performed to minimize the systematic bias due to differences in sequencing depths between the case and control cohorts. First, subjects were excluded if <80% of the protein-coding regions were sequenced at least 50-fold. This exclusion left us with 63 controls and 113 MCD cases [FCDI+ ($n=46$), FCDII ($n=39$), FCDIII ($n=13$) and HMEG ($n=15$)]. To further address the imbalance in coverage between cases and controls, individual site-level coverage levels were generated using the GATKDepthOfCoverage tool at each site in the exonic and 2-bp splice site regions included in the CCDS. Within the case and control cohorts separately, we identified the sites in the autosomes with at least 50 reads in 90% of the samples. To incorporate the differences in depth of sequencing on the X chromosome in males and females, sites were excluded if they did not have at least 25 reads in males and 50 reads in females in 90% of the cohort. After the case and control cohorts were separately normalized, we then selected only sites that were covered per the defined criteria across 90% of the combined case and control cohort. Finally, the coverage normalized sites were then collapsed by CCDS transcripts and genes, and only genes with at least 50% normalized coverage were included in the analysis ($n=15\,622$; Supplementary Table 6).

Gene set enrichment testing

Gene set enrichment analyses were performed using DNENRICH.⁴⁶ The software models exome-wide dispersion of high-quality

somatic variants under a null hypothesis in a way that accounts for the sizes of the genes included in the gene set, the trinucleotide context of single nucleotide substitutions which has been shown to correlate with mutability⁴⁷ and the functional effect of mutations. In this way, a *P*-value can be estimated under a binomial model to evaluate if there are more than the expected number of variants observed in a particular gene set.

In brief, somatic protein-coding SNVs identified in each individual case and control are randomly distributed among the coverage normalized gene set based on the trinucleotide sequence and the functional effect of a mutation (i.e. a single base substitution ATC to AGC that results in a missense variant could be placed at any ATC site located in any of the coverage normalized genes that when the T is substituted for a G would result in a missense variant). Somatic protein-coding indels identified were also randomly placed within the coverage normalized gene set; however, there was no matching for mutability as there are no mutation rate estimates for this class of variants. The somatic variants identified in each individual were randomly placed in the exome in this way 10 000 times to generate the expected rate of somatic variants across a gene set for the cohort.

Once the simulated dataset is generated, DNENRICH then performs a comparative analysis evaluating the observed proportion of functional variants in the specified gene set in cases to controls to the expected proportions generated in the simulated data. A one-sided statistical test was performed to compare the observed case-control proportion to the simulated proportions across the cohorts. This analysis was performed separately comparing the FCDI+, FCDII, FCDIII and HMEG cohorts to the control cohort. Within each cohort, we separately tested for enrichment of somatic variants within 36 sets of genes making up neurologically relevant pathways defined by KEGG (Supplementary Table 5).

Data availability

Exome sequence data from cases evaluated in this study have been deposited into dbGAP (phs002128.v1).

Results

Case ascertainment and phenotypic description of the cohort

One hundred and forty-seven patients with refractory focal epilepsy who had undergone resective epilepsy surgery were evaluated for inclusion in this study. An initial assessment of brain tissue sample quality led to five cases being excluded due to failing quality control checks after sequencing, including one gender mismatch and four due to excessive rates of G–T substitutions indicative of poor-quality DNA.

Each of the 142 remaining cases underwent a centralized review of pre-surgical MRI and neuropathological findings. Nineteen cases were excluded because they did not meet inclusion criteria (e.g. MRI showed evidence of stroke or hypoxic–ischaemic injury). The remaining cases comprised our cohort of 123 individuals. Cases were reviewed and classified by consensus as FCDI+ (*n* = 48), FCDII (*n* = 44), FCDIII (*n* = 15) or HMEG (*n* = 16; Fig. 1A and Supplementary Table 1). A summary of the clinical phenotypes for each of these 123 individuals is provided in Supplementary Table 1. Representative MRIs are shown in Fig. 2.

A subset of FCD specimens (*n* = 12) included tissue resected from individuals with no discernible lesion on preoperative MRI who

were thus preoperatively delineated as NLFE. Pathological review of these NLFE cases revealed findings consistent with FCDI or mMCD (*n* = 5), FCDII (*n* = 2) or FCDIIIc (*n* = 1), while four individuals had no clear neuropathological abnormalities. The five FCDI/mMCD cases and the pathologically normal subset (*n* = 4) were included in the FCDI+ category. The three cases with pathology consistent with FCDII and FCDIII were binned in the FCDII and FCDIII cohorts, respectively, despite being radiographically non-lesional (Supplementary Table 1).

There were approximately equal numbers of males (*n* = 63) and females (*n* = 60) across the whole cohort, but we observed more males in the FCDI+ category (66%, *P* = 0.011) and there was a trend towards more females in the FCDII, FCDIII and HMEG cohorts (Fig. 1B).

Exome sequencing technical summary

The brain tissue samples from the above 123 individuals were sequenced to an average depth of 367-fold (Fig. 1C); 113 (92%) had 80% of protein-coding regions covered at least 50-fold (Fig. 1D). Each sample had on average 9.5 high-quality somatic SNVs or indel variants in the protein-coding regions of the genome (Fig. 1E), seven of which on average were predicted to alter the activity or level of the encoded protein. Among the high-quality somatic variants that were selected for confirmation based on their diagnostic likelihood, we validated 24/26 (92%) using an orthogonal genotyping approach (Supplementary Table 7).

Targeted somatic diagnostic analysis

Given that we expected a fraction of the MCD cases to harbour a known pathogenic variant in either *SLC35A2* or a *PI3K-AKT-mTOR* signalling pathway gene, we first performed a targeted ‘diagnostic’ analysis to identify pathogenic or likely pathogenic variants based on ACMG guidelines.⁴³

Hemimegalencephaly

Somatic gene variants were detected in 12 of 16 (75%) HMEG cases (Table 1). Six individuals each had a somatic *PIK3CA* variant, including five resulting in the previously reported recurrent amino acid substitutions, p.E545K (*n* = 3) and p.E542K (*n* = 2), and one with the previously reported recurrent p.H1047R substitution.^{10,48} Three individuals each had a recurrent somatic *MTOR* variant, two resulting in p.C1483R and one in p.C1483Y at the protein level.¹⁰ Two cases each had a somatic variant in *AKT3*, one resulting in the previously reported p.E17K variant^{8,10} and one with a novel variant resulting in p.T288I. All except the *AKT3* p.T288I variant were identified as high-quality variants in the brain-only call set. The *AKT3* p.T288I variant (VAF = 44%) was only called with the simultaneous analysis of the brain–blood pair as it was miscalled as a germline variant in the brain-only analysis. The *AKT3* p.T288I variant was absent in the leucocyte-derived DNA exome sequencing data, with 300-fold coverage at the variant site, and in data obtained by dPCR. Despite the lack of detectable variant in the blood, the patient presented clinically with neurological and non-neurological features of a segmental overgrowth syndrome including a capillary haemangioma and toe syndactyly (Supplementary Table 8), suggesting that the variant may have been present outside of the central nervous system.

One additional HMEG specimen had a copy number neutral somatic 16p uniparental disomy that has been previously reported.⁴⁹ All somatic variants identified in HMEG cases had VAFs

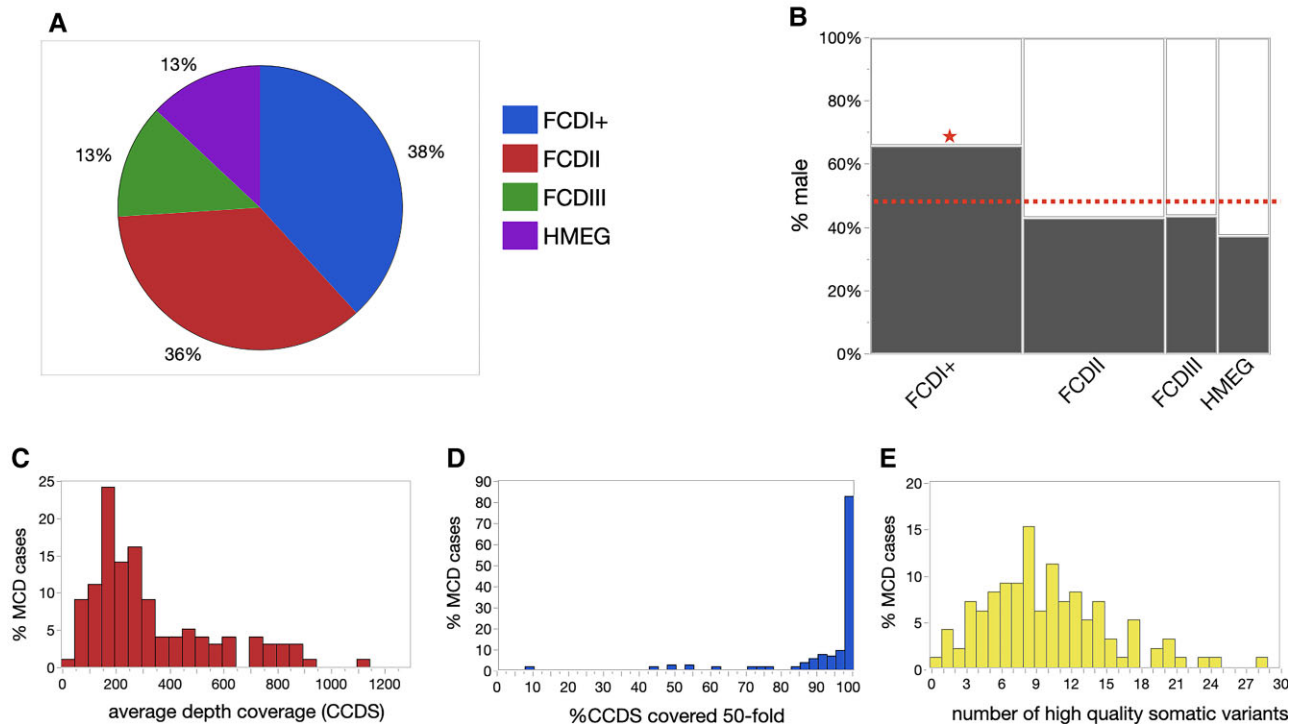


Figure 1 Summary of MCD cohort. (A) Percentages of patients binned in broadly defined MCD phenotypes. (B) Fraction of males within each MCD phenotype. Star indicates a statistically significant deviation from the expected 50% (binomial probability test, $P = 0.011$). (C) Average depth of coverage. (D) Per cent of protein-coding regions sequenced at least 50-fold. (E) Number of high-quality somatic variants called in each subject.

greater than 6.7% (average 20%), reflecting a relatively early mutation (Fig. 3). MRIs for all HMEG cases with a pathogenic somatic variant are provided in Supplementary Fig. 1.

Focal cortical dysplasia type II

Among the 44 individuals diagnosed with FCDII, 13 (30%) had explanatory findings. Eight had a pathogenic or likely pathogenic variant detected in the high-quality somatic variant calls in *MTOR* ($n = 7$) or *PTEN* ($n = 1$; Table 1). The seven individuals with *MTOR* variants harboured either the p.S2215F ($n = 3$), p.S2215Y ($n = 2$), or p.T1977K ($n = 2$) (NM_004958) variants previously reported in individuals with HMEG or isolated FCDII.⁷ The *PTEN* variant (NM_000314:c.G389A:p.R130Q) was identified in one case with a VAF of 9.2% (Table 1). Interestingly, the individual with FCDIIa harbouring the pathogenic *PTEN* variant also had a 1q amplification identified in the CNV analysis (Supplementary Fig. 2), both of which likely contributed to the individual's phenotype. We note that because we cannot determine the specific number of copies of the chromosome arm in our data and because tetrasomy has been reported in some brain cells of an individual with HMEG,⁵⁰ we refer to cases with increased copy number observed in the 1q region as mosaic 1q amplifications.

Among the permissive variant call set, we identified another p.S2215F *MTOR* variant (VAF = 5.8%), one somatic *TSC1* variant (NM_000368:c.C298T:p.Q100X, VAF = 4.4%) and one low-level somatic variant in *TSC2* (NM_001114382:c.1934_1935insCTGCGAC:p.Y648Lfs*11, VAF = 1.7%). All were initially not detected because of low coverage at the variant site. Consistent with other reports of individuals with FCDII and somatic *TSC1* and *TSC2* variants,^{7,15} neither of these patients met clinical criteria for tuberous sclerosis complex, even when rephenotyped in light of these findings.

Two additional cases each with a somatic *MTOR* variant (p.S2215Y and p.T1977R) were identified, one with clinical exome sequencing of the resected brain tissue specimen and one with targeted sequencing of *MTOR* that was part of another study.⁹ While both had at least one read supporting the variant in our exome sequence data and VAFs in the range we would expect to detect (1.9% and 6.1% for p.S2215Y and p.T1977R, respectively), neither reached the quality threshold required for Mutect2 to call the variant.

All somatic variants identified in FCDII cases had VAFs less than 10% (average 5%; Fig. 3).

Among the 10 individuals with pathogenic somatic *MTOR* variants, five had FCDIIa and five had FCDIIb (Supplementary Table 1). Both specimens with a somatic *TSC1* or *TSC2* variant had FCDIIb (Supplementary Table 1).

FCDI and related phenotypes

We identified one pathogenic variant in *SLC35A2* (NM_005660:c.C435A:p.Y145X, VAF = 27%) in an individual with mMCD and decreased myelination, a phenotype consistent with that previously reported in focal neocortical epilepsy caused by somatic *SLC35A2* variants.²³ In addition, we identified and confirmed disease-causing somatic variants in four additional genes previously implicated in neurodevelopmental disorders. The first of these four cases had a somatic splice site variant in *CASK* (NM_003688:c.2302 + 1G > A, VAF = 6.5%) that, when detected as a germline variant, is associated with an intellectual disorder with microcephaly and pontine and cerebellar hypoplasia often accompanied with seizures.^{51,52} This individual exhibited none of the non-neurological phenotypes associated with germline *de novo* *CASK* variants, consistent with the undetectable levels of the variant allele in the blood (Table 2). Furthermore, this subject showed no evidence of

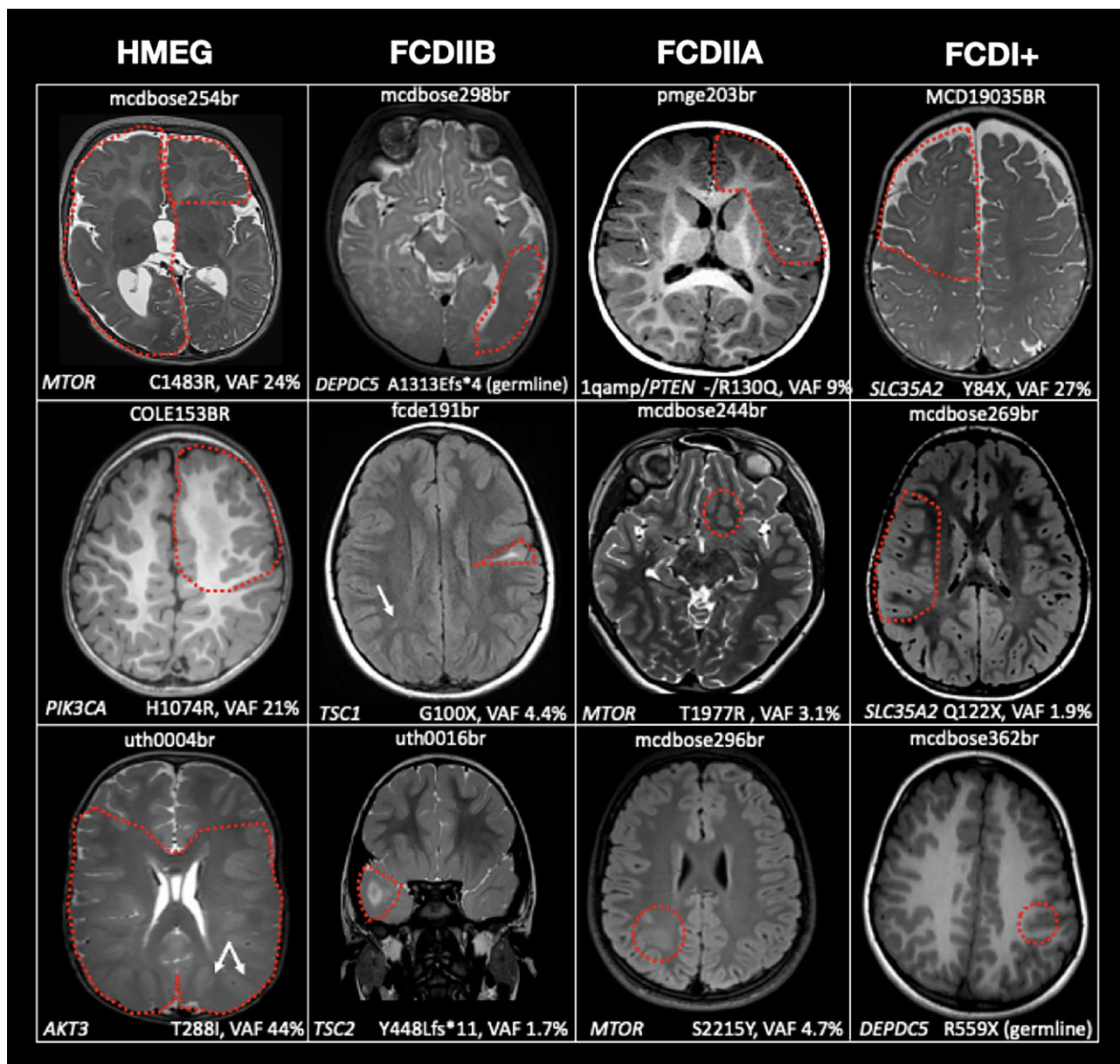


Figure 2 Representative brain MRIs from subjects in the cohort, organized in columns by malformation type (HMEG, FCDIIB, FCDIIA, FCDI+). For each subject (top), the gene, the associated variant and variant allele frequency are provided (bottom). All images are displayed in the axial plane with the exception of uth0016br, which is displayed in coronal, using a mixture of T₁-, T₂- and FLAIR-weighted sequences. The cortical malformation is outlined by a dashed line. For fcd191br and uth0004br, arrows indicate a retrospectively identified transmantle sign in the contralateral hemisphere suspicious for cortical dysplasia and the resected portion of the malformation, respectively.

microcephaly or cerebellar atrophy, features commonly observed in individuals with this variant, at the time of surgery or 18 months post-surgery⁵² (Fig. 4 and Supplementary Table 8), suggesting that the variant may be localized or enriched in the cortical tissue.

The second disease-associated variant detected in our FCDI+ cohort was a somatic variant in *KRAS* (NM_033360:c.G35T;p.G12V, VAF = 20%), a variant previously implicated in arteriovenous brain malformations,⁵³ a range of tumour types^{54,55} and a recently reported FCDIa case.⁵⁶ Other germline variants in *KRAS* have also been associated with Noonan syndrome and cardiofaciocutaneous syndrome,^{57,58} both of which can include developmental delay. The patient with the *KRAS* variant had no radiographic evidence of an arteriovenous malformation or tumour, and head circumference

was normal (49% at 17 months of age; Fig. 4). Due to the lack of a blood sample the tissue localization of the post-zygotically acquired *KRAS* variant was not able to be determined, but the patient exhibited none of the non-neurological phenotypes associated with Noonan syndrome or cardiofaciocutaneous syndrome.

The third likely pathogenic somatic variant identified in FCDI+ was a novel somatic splice site variant in *NIPBL* (NM_133433:c.7685+1G>A, VAF=6.7%), a gene that gives rise to Cornelia de Lange syndrome I, a disorder associated with highly variable phenotypes including dysmorphism, microcephaly, ID and seizures.⁵⁹ The *NIPBL* variant was not found in leucocyte-derived DNA, and the patient had no clinical features of Cornelia de Lange syndrome I (Fig. 4 and Supplementary Table 8).

Table 1 List of pathogenic and candidate disease-causing somatic variants identified in individuals with HMEG and FCDII

Sample	Sex	Phenotype category ^a	Gene—Variant	ACMG classification	Somatic/germline	Brain DNA VAF (exome) [95% CI]	Brain DNA VAF (dPCR/amplicon seq) [95% CI]	Blood DNA VAF (exome), [95% CI]
dukeepi2841br	F	HMEG	16pUPD	LP	Somatic	–	–	–
uth0004br	M	HMEG	AKT3—NM_181690:c.C863T; p.T288I	P	Somatic	44% [38–50%]	–	0% [0–1.2%]
mcdgg17015br	F	HMEG	AKT3—NM_181690:c.G49A; p.E17K	P	Somatic	13% [9.8–19%]	–	0% [0–1.9%]
mcdbose276br	F	HMEG	MTOR—NM_004958:c.G4448A; p.C1483Y	P	Somatic	11% [8.2–15%]	12% [11–13%]	–
mcdbose254br	M	HMEG	MTOR—NM_004958:c.T4447C; p.C1483R	P	Somatic	24% [20–28%]	29% [27–31%]	2% [0–4.7%]
dukeepi2488br	F	HMEG	MTOR—NM_004958:c.T4447C; p.C1483R	P	Somatic	5.7% [1.9–13%]	–	0% [0–6%]
pmge174br	F	HMEG	PIK3CA—NM_006218:c.G1624A; p.E542K	P	Somatic	14% [11–18%]	19% [17–21%]	0% [0–0.6%]
mcdbose274br	F	HMEG	PIK3CA—NM_006218:c.G1624A; p.E542K	P	Somatic	25% [17–36%]	10% [9.2–11%]	–
mcdbose241br	F	HMEG	PIK3CA—NM_006218:c.G1633A; p.E545K	P	Somatic	28% [24–34%]	23% [21–26%]	0% [0–1.1%]
dukeepi5208br	M	HMEG	PIK3CA—NM_006218:c.G1633A; p.E545K	P	Somatic	20% [15–27%]	22% [20–24%]	0% [0–2.5%]
mcdbose348br	M	HMEG	PIK3CA—NM_006218:c.G1633A; p.E545K	P	Somatic	14% [6.8–25%]	27% [26–28%]	–
COLE153BR	M	HMEG	PIK3CA—NM_006218:c.A3140G; p.H1047R	P	Somatic	21% [17–26%]	14% [13–16%]	–
pmge203br	F	FCDII	1q amplification/ PTEN—NM_000314:c.G389A;p.R130Q	P	Somatic	–9% [5.6–14%]	–4.8% [4.4–5.2%]	1.4% [0–3.3%]
mcdbose298br	F	FCDII	DEPDC5—NM_001242896:c.4022delC;p.A1341Efs*4	P	Germline	50% [48–53%]	–	49% [46–52%]
mcdbose286br	M	FCDII	MTOR—NM_004958:c.C6644T; p.S2215F	P	Somatic	5.3% [2.9–8.9%]	5.7% [4.6–6.9%]	–
mcdbose271br	M	FCDII	MTOR—NM_004958:c.C6644T; p.S2215F	P	Somatic	4.0% [2.3–6.6%]	3.9% [2.8–5.4%]	0% [0–1.5%]
COLE141BR	F	FCDII	MTOR—NM_004958:c.C6644T; p.S2215F	P	Somatic	3.9% [2.7–5.4%]	2.2% [1.8–2.7%]	–
fcde183br	F	FCDII	MTOR—NM_004958:c.C6644T; p.S2215F	P	Somatic	5.8% [1.2–16%]	3.5% [3.2–3.8%]	0% [0–2.6%]
mcdbose296br	F	FCDII	MTOR—NM_004958:c.C6644A; p.S2215Y	P	Somatic	4.7% [2.8–7.4%]	7.4% [3.7–7.4%]	0% [0–1%]
neuro1410F49br	M	FCDII	MTOR—NM_004958:c.C6644A; p.S2215Y	P	Somatic	7.2% [4.7–10%]	–	0% [0–2.4%]
mcdbose316br2	M	FCDII	MTOR—NM_004958:c.C6644A; p.S2215Y	LP	Somatic	1.9% [0–5.4%] ^b	5.2% [4.4–6.0%]	0% [0–1.4%]
mcdbose244br	F	FCDII	MTOR—NM_004958:c.C5930G; p.T1977R	LP	Somatic	6.1% [1.2–16.8%] ^b	3.1% [2.7–2.6%]	0% [0–7.2%]
MCDBOSE364BR	F	FCDII	MTOR—NM_004958:c.C5930A; p.T1977K	P	Somatic	4.6% [3.3–6.4%]	4.7% [4.1–5.3%]	–
mcdbose252br	M	FCDII	MTOR—NM_004958:c.C5930A; p.T1977K	P	Somatic	3.8 [2.0–6.5%]	2.9% [2.5–3.4%]	0% [0–0.9%]
mcdbose247br	M	FCDII	SCN1A—NM_001165963:c.T662C;p.L221P	P	Germline	46% [38–54%]	–	42% [35–48]
fcde191br	F	FCDII	TSC1—NM_000368:c.C298T; p.Q100X	P	Somatic	4.4% [0.92–12]	4.1% [3.6–4.6%]	1% [0–4.4%]
uth0016br	F	FCDII	TSC2—NM_001114382:c.1934_1935insCTGCGAC; p.Y648Lfs*11	P	Somatic	1.7% [1.0–2.6%]	–	0% [0–0.6%]

LP = likely pathogenic; P = pathogenic.

^aMore detailed information about radiographic and pathologic phenotypes are provided in [Supplementary Table 1](#).^bSomatic variant not called in our somatic variant calling pipeline.

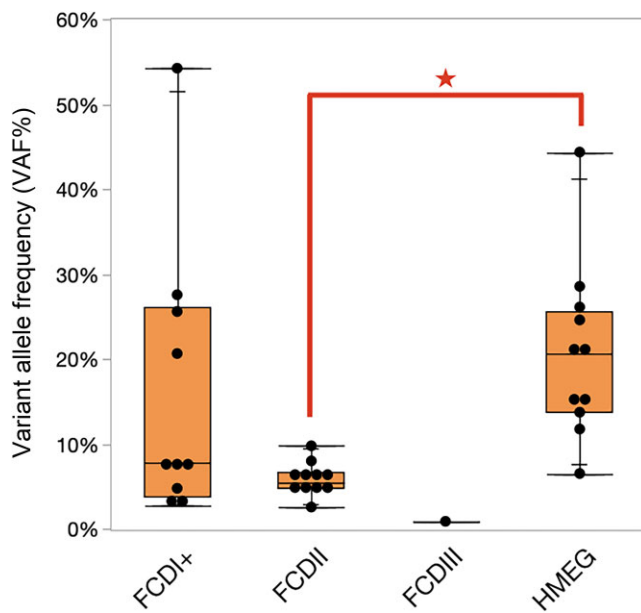


Figure 3 Quantile box plot showing VAF as a function of phenotype. Individual points (black) representing single cases are superimposed. Statistically significant difference (Student's t-test) indicated by bold red bar and star.

Finally, the fourth variant identified was a mosaic frameshift deletion in *NF1* (NM_000267—c.1017_1018del:p.S340Cfs*12, VAF = 3%). Loss-of-function *NF1* variants cause neurofibromatosis 1, a disorder associated with macrocephaly, ID and seizures.^{60,61} While leucocyte-derived DNA from this individual with the somatic *NF1* variant was not available for sequencing in this research study, exome sequencing performed clinically on leucocyte-derived DNA did not detect the variant allele despite >500-fold coverage (Table 2). This individual had no clinical features suggestive of neurofibromatosis 1 (Fig. 4 and Supplementary Table 8). However, in addition to the *NF1* somatic variant identified in this study, an *NF1* germline variant of unknown significance (VOUS; NM_000267: c.G4122T:p.Q1374H) was also identified during clinical exome sequencing testing in the individual. 'Second hits' in *NF1* have been previously identified in benign neurofibromas,⁶² although not specifically to our knowledge within an epileptic zone of an individual with focal epilepsy.

One low-level VOUS was identified and confirmed in *ATP10A* (NM_024490:c.C647A:p.S216X, VAF = 2.9%; Table 2). This variant appeared in two separate exome sequencing runs of two specimens from the same individual. No other sample was sequenced multiple times and no other somatic variant identified in this case was found in both sequencing runs. *ATP10A* encodes a phospholipid-transporting ATPase and is not a known disease gene. At a population level, loss-of-function variants in this gene were identified slightly less frequently than expected in the gnomAD database (loss-of-function observed/expected upper bound fraction score of 0.43), which may indicate that this class of variants in *ATP10A* may be under negative selection.³³ The specimen from this individual had pathology suggesting FCDIa with disrupted cytoarchitecture and neuronal heterotopia (Supplementary Table 8).

Because several of the previously reported and confirmed *SLC35A2* variants identified from the same sequencing data were not called in high-quality variant calls, we also evaluated a more

permissive set of calls (see 'Materials and methods' section) whereby we relaxed the read depth requirements and VAF and excluded only variant calls that are likely due to sequencing or alignment errors. Given the likely high number of false positives we expect in the permissive call set, we limited our analyses only to known pathogenic variants or variants in known MCD genes. Using this approach, we identified three additional somatic *SLC35A2* variants in three different individuals in the FCDI+ cohort, including one novel low-level somatic variant in *SLC35A2* (NM_005660:c.C547T; p.Q183X, VAF = 1.9%) and two somatic *SLC35A2* variants (NM_005660:c.757_758insGCTCTGGTGGG:p.A253Gfs*100, VAF = 25%; and c.G164T:p.R55L, VAF = 54%) that were previously identified and reported on in another publication.⁹ All had not been initially identified because of low coverage at the variant site, lowering confidence that the somatic variant was truly present. One of the cases with a somatic *SLC35A2* variant had mMCD, whereas the others exhibited very subtle cortical dyslamination although not meeting formal diagnostic criteria for FCDIa or Ib (Supplementary Table 1).

Only one high-quality somatic CNV was identified in the FCDI+ cohort, a chromosome 1q amplification (Supplementary Fig. 2). Chromosome 1q amplifications have been reported in individuals with HMEG⁸ and with unilateral polymicrogyria with dysmorphic neurons.⁶³ This individual was diagnosed with FCDNOS with heterotopic white matter neurons and pan-cortical eosinophilic astrocytic inclusions, a pattern described in detail (see Case 2 in Fischer et al.⁶⁴).

Focal cortical dysplasia type III

In FCDIII, we identified only one somatic variant among the 16 cases evaluated. This individual has FCD plus radiographic evidence of mesial temporal sclerosis (Fig. 4) and had a very low-level mosaic loss-of-function *PCDH19* variant (NM_001105243: c.1958_1959del:p.S653Cfs*66, VAF = 0.37%) identified in the cortical specimen from that patient that was identified in the permissive variant call set and confirmed with deep amplicon sequencing.

Germline variant analysis

We identified a total of four likely germline diagnoses (Tables 1 and 2). Two of the pathogenic germline variants were found in *DEPDC5* (NM_001242896:c.C1759T:p.R587X and c.4022delC:p.A1341Efs*4) diagnosed with probable FCDIa and FCDIIb, respectively. Somatic second hits were not identified in either of the germline *DEPDC5* individuals, although we cannot rule out that they were undetected due to low VAF or technical limitations. The third individual had a known pathogenic germline *SCN1A* variant (NM_001165963: c.T662C:p.L221P) and had been diagnosed radiographically and by pathology with FCDIIb. The *SCN1A* variant was also found to be *de novo* with clinical sequencing after inclusion in this study. While not previously associated with FCDII, the *SCN1A* variant was deemed by a diagnostic laboratory to be likely contributing to the individual's epilepsy. Finally a fourth variant in *RANBP2* (NM_006267:c.C1754T:p.T585M) was identified in an individual with FCDIII, specifically FCD with pathology consistent with hippocampal sclerosis.⁶⁵ This *RANBP2* variant is associated with acute infection-induced encephalopathy⁶⁶ and indeed, on further review of the clinical records, the patient had presented at 4 months of age with coma for 4 weeks due to presumed encephalitis; an infectious aetiology was never identified. At age 9 years, the patient developed intractable neocortical seizures that resulted in surgical resection.

Table 2 List of pathogenic and candidate disease-causing somatic variants identified in individuals with FCDI+ or FCDIII

Sample	Sex	Phenotype category ^a	Gene—Variant	ACMG classification ^b	Somatic/germline	Brain DNA VAF (exome) [95% CI]	Brain DNA VAF (dPCR/amplicon seq) [95% CI]	Blood DNA VAF (exome), [95% CI]
uth0008br	M	FCDI ⁺	1q amplification	LP	Somatic	–	–	–
MCDBOSE212BR	M	FCDI ⁺	ATP10A—NM_024490:c.C647A; p.S216X	VOUS	Somatic	2.9% [1.6–5.0%]	1.5% [1.2–1.9]	–
mcdbose208br	F	FCDI ⁺	CASK—NM_003688:c.2302 + 1G > A	P	Somatic	6.5% [4.5–9.2%]	4.8% [2.6–9.5%]	0% [0–1.6%]
uth0017br	M	FCDI ⁺	CUL1—NM_001370660:c.C187T; p.R63X	VOUS	Somatic	7.2% [3.5–13%]	5.5% [4.8–6.3%]	0% [0–3.6%]
mcdbose362br	M	FCDI ⁺	DEPDC5—NM_001242896: c.C1759T;p.R587X	P	Germline	50% [42–58%]	–	–
mcdbose325br	M	FCDI ⁺	KRAS—NM_033360:c.G35T; p.G12V	P	Somatic	20% [15–25%]	18% [17–19%]	–
COLE160BR	M	FCDI ⁺	NF1—NM_000267: c.1017_1018del;p.S340Cfs*12	P	Somatic	3.0% [1.7–4.7%]	2.1% [1.6–2.6%]	0% [0–0.06%] ^b
uth0005br	F	FCDI ⁺	NIPBL—NM_133433:c.7685 + 1G > A	LP	Somatic	6.7% [2.5–14%]	16% [15–18%]	0% [0–2%]
mcdbose240br	M	FCDI ⁺	SLC35A2—NM_005660: c.757_758insGCTCTGGTGGG; p.A253Gfs*100	P	Somatic	25% [13–40%]	–	0% [0–11%]
mcdbose269br	M	FCDI ⁺	SLC35A2—NM_005660:c.C547T; p.Q183X	P	Somatic	1.9% [0.39–5.5%]	2.7% [2.1–3.3]	0% [0–3.6%]
MCD19035BR	M	FCDI ⁺	SLC35A2—NM_005660:c.C435A: p.Y145X	P	Somatic	27% [21–34%]	3.8% [3.0–4.8%]	–
mcdbose233br	M	FCDI ⁺	SLC35A2—NM_005660:c.G164T: p.R55L	LP	Somatic	54% [39–69%]	–	0% [0–5%]
COLE131BR	F	FCDIII	PCDH19—NM_001105243: c.1958_1959del;p.S653Cfs*66	P	Somatic	0.37% [0.12–0.86%]	0.34% [0.31–0.37%]	–
dukeepi3209br	M	FCDIII	RANBP2—NM_006267:c.C1754T: p.T585M	P	Germline	48% [42–53%]	–	–

LP = likely pathogenic; P = pathogenic; VOUS = variant of unknown significance.

^aMore detailed information about radiographic and pathologic phenotypes are provided in [Supplementary Table 1](#).

^bSomatic variant not called in our somatic variant calling pipeline.

Novel gene search

After completing a targeted evaluation for variants in known genes, we next sought to evaluate the high-quality call set for candidate somatic variants that may contribute to the aetiological landscape of the MCDs. The majority of pathogenic variants in dominant acting genes involved in rare neurodevelopmental conditions, such as epilepsy and brain malformations, are ultra-rare in the population, putatively functional and are often found in intolerant or constrained genes (i.e. genes that tend to be depleted of functional variants, specifically loss-of-function variants, in the general population).^{67–69} Given this observation, we first evaluated only those variants that are rare (gnomAD exome MAF < 1 × 10^{−5}), functional [splice-site, nonsense, missense excluded Polyphen (HumVar) benign, frameshift and non-frameshift indel], and located in a constrained gene, or a rare loss-of-function variant (nonsense, splicing, or frameshift indel) in a gene with a high score reflecting genic loss-of-function variant intolerance (pLI > 0.9).⁶⁸ Across all 123 cases, there were 49 variants meeting these criteria. There were only two genes that harboured more than one high-quality somatic variant call confirmed with independent genotyping in different individuals, including MTOR (*n* = 10) and PIK3CA (*n* = 5). Each of these MTOR and PIK3CA variants had been identified in the diagnostic analyses. Among the genes with single somatic variants identified, five had been identified in the diagnostic analysis above (AKT3, CASK, NF1, NIPBL and PTEN). Among the 29 qualifying variants remaining, all except one were not felt to be strong

candidates either because they appeared to be an artefact with visual inspection, were found to be more common in other control data sets, did not confirm with secondary genotyping, or were in genes unlikely to contribute to disease. The one additional candidate identified and confirmed in this analysis was a somatic loss-of-function variant in *CUL1* (NM_001370660:c.C187T:p.R63X, VAF = 7.2%). *CUL1* encodes a protein comprising the ubiquitin ligase complex that serves to ubiquitinate proteins involved in cell cycle progression. *CUL1* has not to date been implicated in human disease, but it is embryonic lethal when knocked out in mice.⁷⁰

Aggregate analyses

Enrichment of known pathogenic somatic variants in MCD

Given the lack of knowledge of the somatic genetic landscape in brain tissue-derived DNA from healthy individuals, we next sought to compare the rate of low-level pathogenic somatic variants in MCD cases to that of controls. The control cohort used for this analysis consisted of exome and whole genome sequence data from autopsy-collected brain tissue specimens from 63 individuals without neurological disorders made available for biomedical research through the North American Brain Expression Consortium and the Brain Somatic Mosaicism Network. To ensure a fair comparison between cases and controls, we identified a list of 15 622 of 18 078 CCDS genes that were sequenced to approximately equal depths and extents in cases and controls ([Supplementary Table 6](#), see

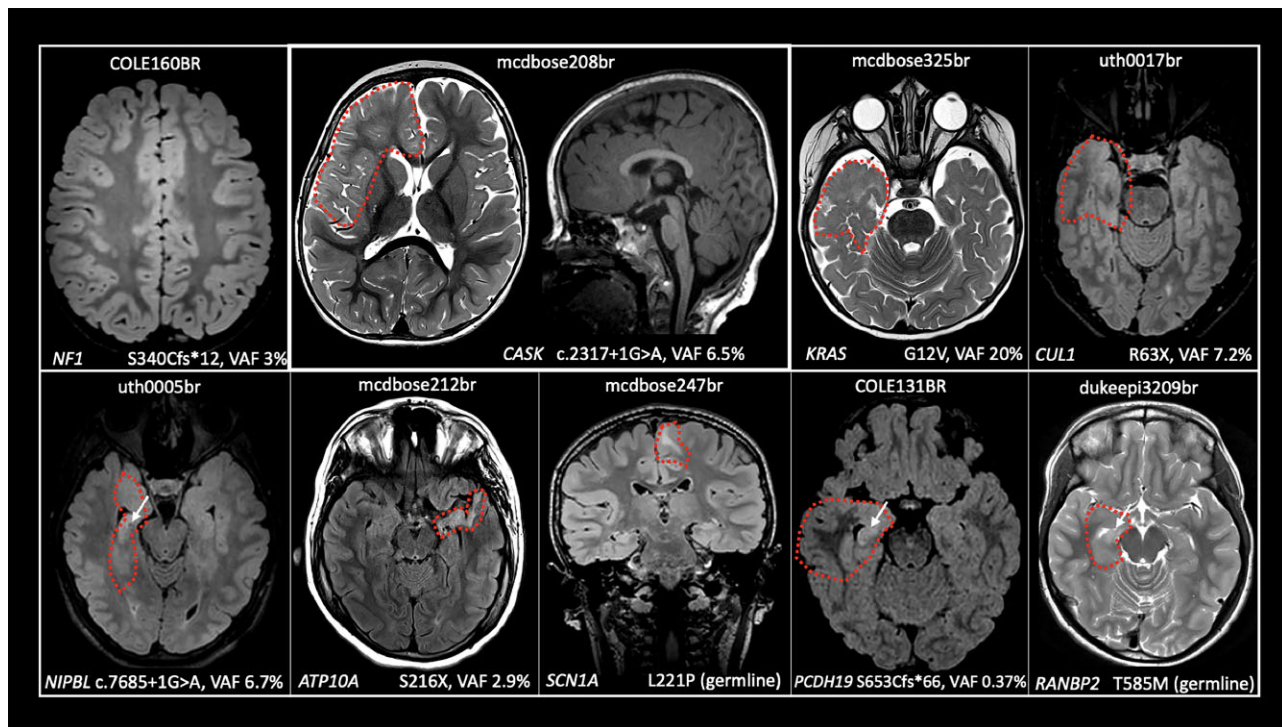


Figure 4 Brain MRIs for subjects with novel somatic variants. Each subject is shown using a mixture of axial T₂- and FLAIR-weighted imaging sequences with the exception of mcdbose208br, for which a sagittal T₁ is also provided, and mcdbose247, which is depicted in the coronal plane (annotations as in Fig. 2). The cortical malformation is outlined by a dashed line except for COLE160BR, which was a non-lesional case. Lesions were classified as FCDIIB for mcdbose247br and FCDIII in three cases (uth0005br, COLE131BR and dukeepi3209br with hippocampal sclerosis indicated by the arrow). The remaining cases were classified as FCDI+ (COLE160BR, mcdbose208br, mcdbose325br, uth0017br, mcdbose212br). Note that mcdbose208br demonstrates normal hindbrain volume, unlike germline variants in CASK. VAFs obtained from exome sequence data are provided for each somatic variant in the bottom right.

'Materials and methods' section). Evaluating only these 15 622 genes we performed the somatic diagnostic analysis (see Materials and methods) initially applied to our MCD cases in the control cohort. Because DNA was not available to us to confirm variant calls in controls we did not consider variant validation status in our analysis and only relied on the bioinformatically defined filtering criteria that could be applied consistently across both cohorts. Within the coverage normalized gene set (15 622 genes), we found no candidate somatic genetic diagnoses in controls, whereas in the FCDI+, FCDII and HMEG cohorts we identified 5/46 (11%, two-sided Fisher's exact test $P=0.017$), 7/39 (18%, $P=8 \times 10^{-4}$), and 10/15 (67%, $P=2.4 \times 10^{-9}$), respectively. Not unexpectedly, the enrichment of somatic genetic diagnoses in the FCDII and HMEG cohorts was driven by somatic variants in *PIK3CA*, *AKT3* and *MTOR*. The enrichment in FCDI+ was driven by the singleton variants discussed in the diagnostic analysis in *CASK*, *KRAS*, *NF1*, *NIPBL* and *SLC35A2*, and suggests that these variants are contributing to the somatic genetic risk of FCDI+. Even after removing the one variant in the only known FCDI+ associated gene, *SLC35A2*, there were still more than expected somatic genetic variants observed in the FCDI+ cohort compared to controls (Fisher's exact test $P=0.028$), also suggesting the presence of novel pathogenic genetic variants present in our dataset.

Pathway enrichment analyses

Finally, we performed a pathway analysis to evaluate if high-quality somatic variants were enriched in biologically informed gene sets in any of the MCD cohorts. Across 36 KEGG-defined neurological pathways, we observed no enrichment of somatic variants

in cases compared to controls in the FCDI+ and FCDIII cohorts. Predictably, within FCDII and HMEG cohorts somatic variants in the PI3K-AKT-mTOR signalling pathway genes were observed more than expected (Fig. 5). The Ras signalling pathway genes had more than expected somatic variants in HMEG cases, although the signal was driven exclusively by variants in *AKT3* and *PIK3CA* (Fig. 5).

Discussion

The role of post-zygotically acquired somatic variants that arise during foetal brain development in epilepsy-associated MCD has been increasingly recognized over the past several years. Prior reports of somatic variants have focused largely on deep sequencing of genes known to be involved in specific types of focal cortical malformations. Here we report the results of the first and largest genome-wide investigation for post-zygotically acquired somatic genetic variants in the resected brain tissue of 123 patients with epilepsy. Based on rigorous phenotypic assessment of their MRI and neuropathological findings, we classified the lesions into four clinically distinct forms of MCD—FCDI+, FCDII, FCDIII and HMEG—which allowed for a novel comparison of the genetic profiles of different MCD types.

Similar to what we and others have previously reported, we identified a high rate of pathogenic or likely pathogenic genetic findings in HMEG (75%) as well as the commonly encountered and radiographically and pathologically well-defined FCDII (30%; Fig. 6). All pathogenic and likely pathogenic variants that we

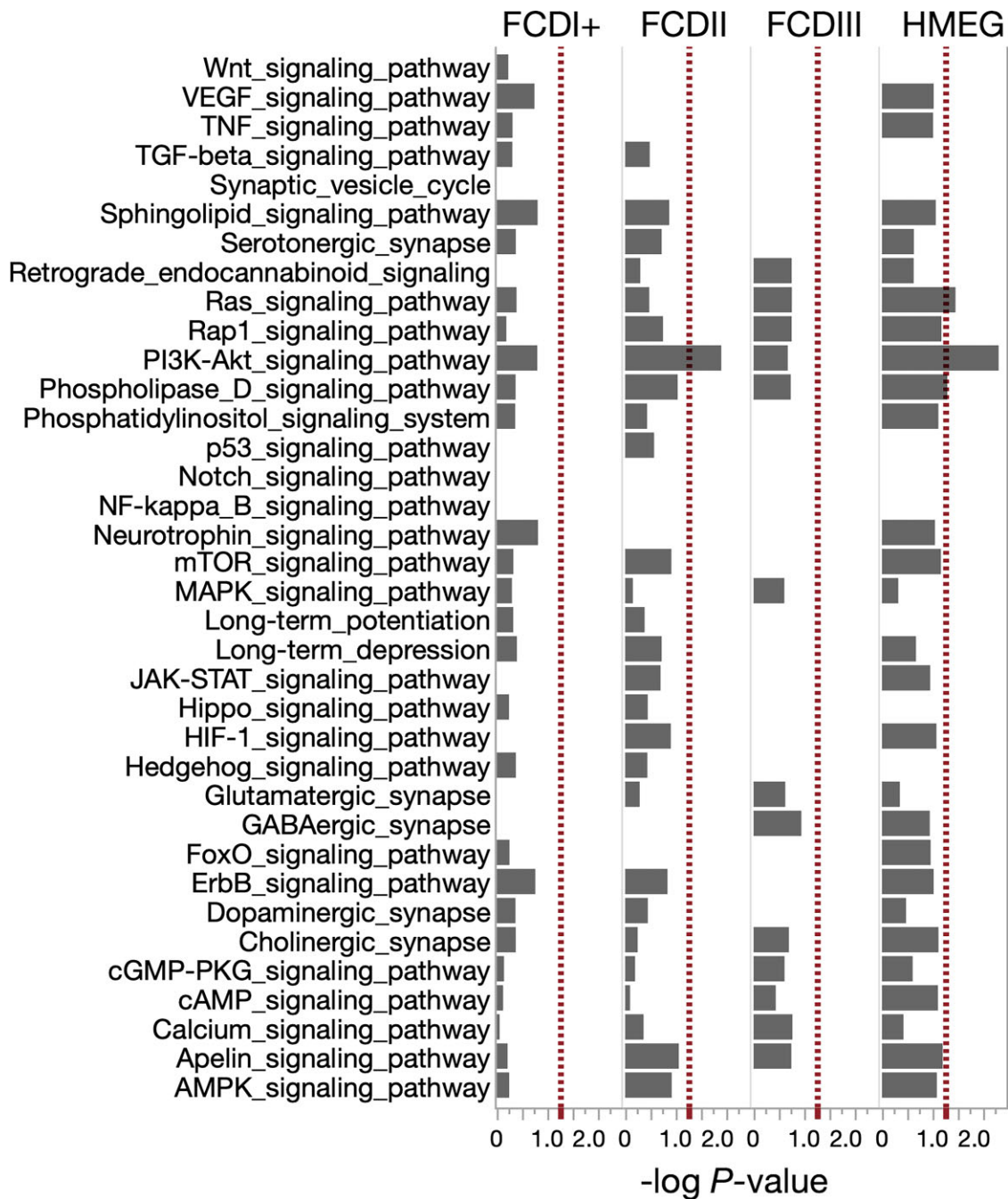


Figure 5 DGENRICH analysis. Transformed P-values ($-\log P$ -value) reflecting the probability that the number, types (indel, SNV) and effects (missense, loss-of-function, etc.) of observed somatic variants in the genes comprising each KEGG pathway that could have occurred by chance in the FCDI+, FCDII, FCDIII and HMEG cohorts. The dashed line is the threshold of significance corresponding to a P-value of <0.05 .

identified in HMEG and FCDII cases were in genes comprising the PI3K-AKT-mTOR signalling pathway, which is not surprising given the shared pathological features between these malformations. The variants in *AKT3*, *DEPDC5*, *MTOR*, *PIK3CA*, *PTEN*, *RHEB*, *RPS6*, *TSC1* and *TSC2* in HMEG and FCDII reinforce the pivotal role of mTOR hyperactivation as a shared mechanism underlying these select MCD subtypes,^{8,10,15,18–22,48} substantiating their characterization as ‘mTORopathies’.⁷¹

While somatic gene discovery in FCDII has progressed rapidly in the past few years, identifying specific causes for FCDI has been

more challenging. As we begin to identify somatic variants as the cause of FCDI, it is clear multiple distinct genes associated with brain development lead to FCDI. This genetic heterogeneity for FCDI is in sharp contrast to FCDII, for which virtually all associated gene variants have been linked to the PI3K-AKT-mTOR signalling pathway. Despite the challenges identifying genes in FCDI and related phenotypes, we note that the rate of genetic diagnosis in FCDI+ approaches that of FCDII when considering the novel findings identified here (Fig. 6).

Despite the contrasting genetics between FCDI and FCDII/HMEG, we did identify a single individual with FCDNOS (FCD1+

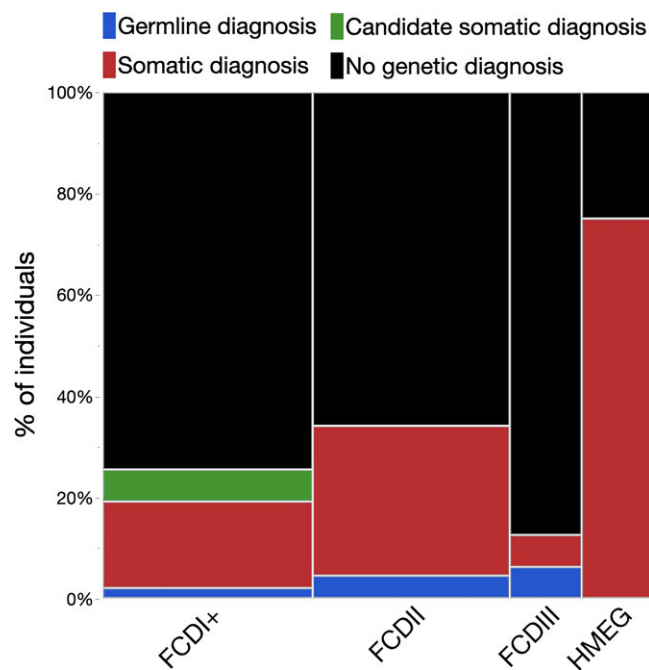


Figure 6 Rate of genetic diagnoses across phenotypic categories. Percent of individuals with a germline genetic diagnosis (blue), a somatic genetic diagnosis (red), a candidate somatic genetic diagnosis (green), or no genetic diagnosis identified (black). The size (width) of the bar corresponds with size of the cohort.

category) with a somatic chromosome 1q amplification, a CNV that has been reported previously in HMEG⁸ and two individuals with a *DEPDC5* variant, one with FCDIa and one with FCDIIb. For 1q amplification, the variable phenotype may be caused by excess expression of *AKT3*, although it may equally likely be due to combined effects of triplosensitivity of a large number of genes residing on the q arm of chromosome 1 and/or variable cell-type specific mutation burdens. Consistent with our observations, variable phenotypes in individuals with *DEPDC5* variants, including some cases with pathology consistent with FCDI and some with FCDII, have been reported previously.^{72–75} The variable phenotype reported in this study for *DEPDC5*, as well as what has been reported previously, may be explained by the presence or absence of a ‘second hit’, as has been shown in a subset of cases, or by variant-specific effects on the protein or variable cell-type specific variant burden. While no second hit was identified in either individual with a germline loss-of-function *DEPDC5* variant in this study, it remains possible that they were present but simply undetected at the sequencing depth achieved in these cases.

Consistent with prior studies of lesional and non-lesional focal neocortical epilepsies, we identified somatic variants in *SLC35A2* in FCDI+ cases in this cohort. Three of the four individuals with *SLC35A2* had pathological features consistent with MOGHE. Pathological assessments of these cases at the time of ascertainment did not specifically reference MOGHE because this was not well recognized as a distinct pathological entity and still is not a formal ILAE classification term. Additional quantitative studies will likely be needed to develop more definitive diagnostic guidelines that take into account region-specific oligodendrocyte development.^{76,77} Interestingly, we also identified somatic variants in *CASK*, *KRAS*, *NF1* and *NIPBL*, genes previously associated with neurodevelopmental disorders but in the presence of *de novo* germline

variants. All of the somatic variants identified were either identical to the previously reported pathogenic germline *de novo* variants (*CASK* and *KRAS*) or were predicted to result in the same functional effect on the protein (*NF1*, *NIPBL* and *SLC35A2*). The phenotypes associated with each of these gene variants in our MCD patients were restricted to brain-associated phenotypes, not surprisingly, because the variants were mosaic and enriched in brain tissue in all cases where a blood sample was available for comparison (Table 2). For example, the patient with the somatic *NIPBL* variant lacked limb abnormalities and facial dysmorphisms, but shared seizures, brain malformation and developmental delay associated with germline *de novo* variants in *NIPBL* in patients with Cornelia de Lange syndrome 1.⁵⁹ These examples provide additional evidence, similar to that seen with *SLC35A2*, that at least some genes with germline variants previously implicated in neurodevelopmental disorders, but not focal brain malformations, can nonetheless cause epilepsy-associated focal MCD when they occur as somatic variants.

NF1, *NIPBL* and *KRAS* play known roles in normal neuronal development. For example, approximately 50% of autopsied brains from neurofibromatosis patients show disordered cortical architecture with random orientation of neurons, focal heterotopic neurons, proliferation of glial cells to form well-defined gliofibrillary nodules and hyperplastic gliosis.⁷⁸ In addition, *NF1* knockout mice have reduced cerebral cortical thickness and increased cell packing density.⁷⁹ Mouse embryonic brains electroporated in utero with *Nipbl*-targeting shRNAs at embryonic Day 14.5 exhibit defects in neuronal migration including significant accumulation of transfected neurons in the intermediate zone and a reduction of cell numbers in the cortical plate.⁸⁰ Interestingly, the somatic *KRAS* variant detected in FCDI has been specifically implicated in arteriovenous malformations when localized to brain endothelial cells,⁵³ and more recently, reported in a single case with FCDIa.⁵⁶ The former raises questions about the impact of cell-type specific variant burden in dictating phenotype and highlights an area where additional research is needed.

FCDIII is pathophysiologically more complex than FCDI and II, because it is characterized by FCDI+ associated with an additional neuropathological finding (i.e. vascular malformation, glioneuronal tumour, or hippocampal sclerosis).¹¹ Our gene discovery in FCDIII is completely novel as no study to date has identified a genetic association with this MCD subtype. In our study, we identified a single very low-level (VAF = 0.37%) somatic *PCDH19* variant in the temporal cortical sample from a female individual classified as FCDIII based on the presence of a mild FCD in the presence of mesial temporal sclerosis. This patient had seizure onset at age 2 years and neurobehavioural difficulties; she had normal overall cognitive ability but relative weakness in verbal conceptual skills, which is milder than what is typically seen in germline *PCDH19*-related epilepsy in females.⁸¹ We also identified a case of *RANBP2*-associated encephalopathy⁶⁶ that was linked to an undefined infectious illness and later developed into focal epilepsy with mesial temporal sclerosis. As mesial temporal sclerosis may represent the secondary effects of repetitive or prolonged seizures, it is possible that it represents a secondary finding that can influence the classification of the FCD from FCDI (what they would have been diagnosed in the absence of mesial temporal sclerosis) to FCDIII. If that is the case, we would expect overlap between the genetics of FCDI and FCDIII to emerge over time.

Our study also suggests that while low-level pathogenic somatic variants can be detected, a considerable fraction of FCDI, FCDII and FCDIII remain genetically unexplained despite high-depth exome sequencing. In our study, the VAF averaged around 20% for

HMEG, compared to less than 5% in FCDII which are much smaller brain malformations (Fig. 3). Even within individuals specifically with pathogenic MTOR variants, larger lesions were observed with higher VAFs (Supplementary Fig. 3). This observation aligns with the previously reported relationship between lesion size and balloon cell burden with VAF in brain malformations associated with mTOR-associated pathologies.^{7,22} In addition, we observed two instances in which pathogenic somatic MTOR variants were initially missed because of a combination of the low level of the variants and insufficient depth of sequencing at the locus to detect the mutation, one case of an extremely low-level pathogenic somatic PCDH19 variant in an individual with FCDIII and multiple somatic variant calls in FCDI⁺ with VAFs less than 5%. These data collectively suggest that even with high-depth exome sequencing, pathogenic somatic variants may be missed that may be captured with even higher-depth sequencing, single-cell sequencing, or more advanced sequencing and variant calling approaches to allow for more accurate calling of low-level somatic variants. Such low-level variants may explain at least a subset of the ~67% of genetically unexplained individuals in our study and should be the focus of future investigations.

There are a few notable limitations to our study. First, we performed exome sequencing on brain tissue specimens resected in the course of epilepsy surgery; while these specimens represented a portion of a larger surgically resected epileptogenic focus, they were adjacent to but not necessarily representative of the same region that was analysed for neuropathological findings. Second, we did not sequence samples from across the resected regions in this study, so we were not able to detect regional variability in pathogenic variants. Future studies could more directly ascertain tissue for sequencing that is in direct proximity to where the neuropathological assessments are made and conduct comprehensive sampling across resected tissue regions.

In conclusion, our study sheds important new light on the somatic genetic bases of MCD, most notably emerging causes for FCDI and related phenotypes in the form of somatic variants in a diverse set of genes involved in neurodevelopmental disease. While FCDII and HMEG have been well-characterized by MRI, pathology and now increasingly genetic causes involving the mTOR pathway, FCDI and FCDIII are more heterogeneous disorders. We are only at the beginning of achieving diagnostic precision for all forms of FCD, which will also require translation to the clinical evaluation of patients with focal epilepsy. Our findings further support the incorporation of genetic findings, in addition to the currently used pathological classification, into stratifying diverse forms of MCD¹⁴ and motivates even larger-scale studies to further define the somatic genetic landscapes of MCD.

Acknowledgements

The authors kindly thank all the study participants and their families and referring physicians for their willingness to contribute samples and clinical data to this work. We thank the Boston Children's Hospital Core for Neurological Diseases for ascertainment, patient consent and sample procurement.

The authors also thank the Bioinformatics and Analytics Research Collaborative, and in particular Dr Hemant Kelkar and Halina Krzystek for their guidance developing and implementing the bioinformatics pipeline used in this study. We also thank Mr Corey Bailey in the Duke Pathology Department for his extensive assistance with locating and organizing archived slides for review of clinical cases.

We also thank the artist who created the thumbnail image associated with this manuscript, Fawn Gracey.

Data and/or research tools used in the preparation of this manuscript were obtained from the National Institute of Mental Health (NIMH) Data Archive (NDA). NDA is a collaborative informatics system created by the National Institutes of Health to provide a national resource to support and accelerate research in mental health. Data set identifier(s): 2962. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or of the Submitters submitting original data to the NDA.

Funding

This work was supported by the NINDS (R01-NS094596, R01-NS114122 and R01-NS089552). E.L.H. was supported by the Irving Institute for Clinical and Translational Research at Columbia University. A.P. was supported by the Boston Children's Hospital Translational Research Program. P.B.C. is supported by a Javits Award (NINDS R37-NS125632).

Competing interests

The authors report no competing interests.

Supplementary material

Supplementary material is available at *Brain* online.

References

1. Barkovich AJ, Kuzniecky RI, Dobyns WB, Jackson GD, Becker LE, Evrard P. A classification scheme for malformations of cortical development. *Neuropediatrics*. 1996;27(2):59–63.
2. Barkovich AJ, Guerrini R, Kuzniecky RI, Jackson GD, Dobyns WB. A developmental and genetic classification for malformations of cortical development: Update 2012. *Brain*. 2012;135(Pt 5):1348–1369.
3. Leventer RJ, Phelan EM, Coleman LT, Kean MJ, Jackson GD, Harvey AS. Clinical and imaging features of cortical malformations in childhood. *Neurology*. 1999;53(4):715–722.
4. Casanova MF, El-Baz AS, Kamat SS, et al. Focal cortical dysplasias in autism spectrum disorders. *Acta Neuropathol Commun*. 2013;1:67.
5. Jamuar SS, Lam A-TN, Kircher M, et al. Somatic mutations in cerebral cortical malformations. *N Engl J Med*. 2014;371(8):733–743.
6. Guerrini R, Dobyns WB. Malformations of cortical development: Clinical features and genetic causes. *Lancet Neurol*. 2014;13(7):710–726.
7. D'Gama AM, Woodworth MB, Hossain AA, et al. Somatic mutations activating the mTOR pathway in dorsal telencephalic progenitors cause a continuum of cortical dysplasias. *Cell Rep*. 2017;21(13):3754–3766.
8. Poduri A, Evrorny GD, Cai X, et al. Somatic activation of AKT3 causes hemispheric developmental brain malformations. *Neuron*. 2012;74(1):41–48.
9. Winawer MR, Griffin NG, Samanamud J, et al. Somatic SLC35A2 variants in the brain are associated with intractable neocortical epilepsy. *Ann Neurol*. 2018;83(6):1133–1146.
10. Lee JH, Huynh M, Silhavy JL, et al. De novo somatic mutations in components of the PI3K-AKT3-mTOR pathway cause hemimegalencephaly. *Nat Genet*. 2012;44(8):941–945.

11. Blümcke I, Thom M, Aronica E, et al. The clinicopathologic spectrum of focal cortical dysplasias: A consensus classification proposed by an ad hoc Task Force of the ILAE Diagnostic Methods Commission. *Epilepsia*. 2011;52(1):158–174.
12. Palmini A, Najm I, Avanzini G, et al. Terminology and classification of the cortical dysplasias. *Neurology*. 2004;62(6 Suppl 3):S2–S8.
13. Schurr J, Coras R, Rossler K, et al. Mild malformation of cortical development with oligodendroglial hyperplasia in frontal lobe epilepsy: A new clinico-pathological entity. *Brain Pathol*. 2017;27(1):26–35.
14. Blumcke I, Cendes F, Miyata H, Thom M, Aronica E, Najm I. Toward a refined genotype–phenotype classification scheme for the international consensus classification of Focal Cortical Dysplasia. *Brain Pathol*. 2021;31(4):e12956.
15. Lim JS, Gopalappa R, Kim SH, et al. Somatic mutations in TSC1 and TSC2 cause focal cortical dysplasia. *Am J Hum Genet*. 2017;100(3):454–472.
16. Lim JS, Kim W-I, Kang H-C, et al. Brain somatic mutations in MTOR cause focal cortical dysplasia type II leading to intractable epilepsy. *Nat Med*. 2015;21(4):395–400.
17. Moller RS, Weckhuysen S, Chipaux M, et al. Germline and somatic mutations in the MTOR gene in focal cortical dysplasia and epilepsy. *Neurol Genet*. 2016;2(6):e118.
18. Nakashima M, Saitou H, Takei N, et al. Somatic mutations in the MTOR gene cause focal cortical dysplasia type IIb. *Ann Neurol*. 2015;78(3):375–386.
19. Pelorosso C, Watrin F, Conti V, et al. Somatic double-hit in MTOR and RPS6 in hemimegalencephaly with intractable epilepsy. *Hum Mol Genet*. 2019;28(22):3755–3765.
20. Salinas V, Vega P, Piccirilli MV, et al. Identification of a somatic mutation in the RHEB gene through high depth and ultra-high depth next generation sequencing in a patient with hemimegalencephaly and drug resistant epilepsy. *Eur J Med Genet*. 2019;62(11):103571.
21. Zhao S, Li Z, Zhang M, et al. A brain somatic RHEB doublet mutation causes focal cortical dysplasia type II. *Exp Mol Med*. 2019;51(7):1–11.
22. Baldassari S, Ribierre T, Marsan E, et al. Dissecting the genetic basis of focal cortical dysplasia: A large cohort study. *Acta Neuropathol*. 2019;138(6):885–900.
23. Bonduelle T, Hartlieb T, Baldassari S, et al. Frequent SLC35A2 brain mosaicism in mild malformation of cortical development with oligodendroglial hyperplasia in epilepsy (MOGHE). *Acta Neuropathol Commun*. 2021;9(1):3.
24. Sim NS, Seo Y, Lim JS, et al. Brain somatic mutations in SLC35A2 cause intractable epilepsy with aberrant N-glycosylation. *Neurol Genet*. 2018;4(6):e294.
25. Heinzen EL. Somatic variants in epilepsy—Advancing gene discovery and disease mechanisms. *Curr Opin Genet Dev*. 2020;65:1–7.
26. Miller KE, Koboldt DC, Schieffer KM, et al. Somatic SLC35A2 mosaicism correlates with clinical findings in epilepsy brain tissue. *Neurol Genet*. 2020;6(4):e460.
27. Chamberlain WA, Cohen ML, Gyure KA, et al. Interobserver and intraobserver reproducibility in focal cortical dysplasia (malformations of cortical development). *Epilepsia*. 2009;50(12):2593–2598.
28. Kodera H, Nakamura K, Osaka H, et al. De novo mutations in SLC35A2 encoding a UDP-galactose transporter cause early-onset epileptic encephalopathy. *Hum Mutat*. 2013;34(12):1708–1714.
29. Ng BG, Buckingham KJ, Raymond K, et al. Mosaicism of the UDP-galactose transporter SLC35A2 causes a congenital disorder of glycosylation. *Am J Hum Genet*. 2013;92(4):632–636.
30. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009;25(14):1754–1760.
31. Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–2079.
32. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*. 2013;31(3):213–219.
33. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434–443.
34. Geoffroy V, Herenger Y, Kress A, et al. AnnotSV: An integrated tool for structural variations annotation. *Bioinformatics*. 2018;34(20):3572–3574.
35. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high-confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics*. 2013;43:11.10.1–11.10.33.
36. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43(5):491–498.
37. Pedersen BS, Bhetariya PJ, Brown J, et al. Somalier: Rapid relatedness estimation for cancer and germline studies using efficient genome sketches. *Genome Med*. 2020;12(1):62.
38. Genomes Project C, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68–74.
39. Costello M, Pugh TJ, Fennell TJ, et al. Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Res*. 2013;41(6):e67.
40. Chen L, Liu P, Evans TC Jr, Ettwiller LM. DNA damage is a pervasive cause of sequencing errors, directly confounding variant identification. *Science*. 2017;355(6326):752–756.
41. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.
42. Pruitt KD, Harrow J, Harte RA, et al. The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. *Genome Res*. 2009;19(7):1316–1323.
43. Richards S, Aziz N, Bale S, et al. 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*. 2015;17(5):405–424.
44. Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD), May, 14, 2020. World Wide Web URL: <https://omim.org/>
45. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. 2000;28(1):27–30.
46. Fromer M, Pocklington AJ, Kavanagh DH, et al. De novo mutations in schizophrenia implicate synaptic networks. *Nature*. 2014;506(7487):179–184.
47. Krawczak M, Ball EV, Cooper DN. Neighboring-nucleotide effects on the rates of germ-line single-base-pair substitution in human genes. *Am J Hum Genet*. 1998;63(2):474–488.
48. D’Gama AM, Geng Y, Couto JA, et al. Mammalian target of rapamycin pathway mutations cause hemimegalencephaly and focal cortical dysplasia. *Ann Neurol*. 2015;77(4):720–725.
49. Griffin NG, Cronin KD, Walley NM, et al. Somatic uniparental disomy of chromosome 16p in hemimegalencephaly. *Cold Spring Harb Mol Case Stud*. 2017;3(5):a001735.
50. Cai X, Evrony GD, Lehmann HS, et al. Single-cell, genome-wide sequencing identifies clonal somatic copy-number variation in the human brain. *Cell Rep*. 2014;8(5):1280–1289.

51. Piluso G, D'Amico F, Saccone V, et al. A missense mutation in CASK causes FG syndrome in an Italian family. *Am J Hum Genet.* 2009;84(2):162–177.
52. Najm J, Horn D, Wimplinger I, et al. Mutations of CASK cause an X-linked brain malformation phenotype with microcephaly and hypoplasia of the brainstem and cerebellum. *Nat Genet.* 2008;40(9):1065–1067.
53. Nikolaev SI, Vetiska S, Bonilla X, et al. Somatic activating KRAS mutations in arteriovenous malformations of the brain. *N Engl J Med.* 2018;378(3):250–261.
54. Kranenburg O. The KRAS oncogene: Past, present, and future. *Biochim Biophys Acta.* 2005;1756(2):81–82.
55. Pulciani S, Santos E, Lauver AV, Long LK, Aaronson SA, Barbacid M. Oncogenes in solid human tumours. *Nature.* 1982;300(5892):539–542.
56. Pepi C, de Palma L, Trivisano M, et al. The role of KRAS mutations in cortical malformation and epilepsy surgery: A novel report of nevus sebaceous syndrome and review of the literature. *Brain Sci.* 2021;11(6):793.
57. Schubert S, Zenker M, Rowe SL, et al. Germline KRAS mutations cause Noonan syndrome. *Nat Genet.* 2006;38(3):331–336.
58. Niihori T, Aoki Y, Narumi Y, et al. Germline KRAS and BRAF mutations in cardio-facio-cutaneous syndrome. *Nat Genet.* 2006;38(3):294–296.
59. Tonkin ET, Wang T-J, Lisgo S, Bamshad MJ, Strachan T. NIPBL, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome. *Nat Genet.* 2004;36(6):636–641.
60. Valero MC, Velasco E, Moreno F, Heméndez-Chico C. Characterization of four mutations in the neurofibromatosis type 1 gene by denaturing gradient gel electrophoresis (DGGE). *Hum Mol Genet.* 1994;3(4):639–641.
61. Wallace MR, Andersen LB, Saulino AM, Gregory PE, Glover TW, Collins FS. A *de novo* Alu insertion results in neurofibromatosis type 1. *Nature.* 1991;353(6347):864–866.
62. Serra E, Puig S, Otero D, et al. Confirmation of a double-hit model for the NF1 gene in benign neurofibromas. *Am J Hum Genet.* 1997;61(3):512–519.
63. Kobow K, Jabari S, Pieper T, et al. Mosaic trisomy of chromosome 1q in human brain tissue associates with unilateral polymicrogyria, very early-onset focal epilepsy, and severe developmental delay. *Acta Neuropathol.* 2020;140(6):881–891.
64. Fischer GM, Vaziri Fard E, Shah MN, et al. Filamin A–negative hyaline astrocytic inclusions in pediatric patients with intractable epilepsy: Report of 2 cases. *J Neurosurg Pediatr.* 2020;26:38–44.
65. Blumcke I, Cross JH, Spreafico R. The international consensus classification for hippocampal sclerosis: An important step towards accurate prognosis. *Lancet Neurol.* 2013;12(9):844–846.
66. Neilson DE, Adams MD, Orr CM, et al. Infection-triggered familial or recurrent cases of acute necrotizing encephalopathy caused by mutations in a component of the nuclear pore, RANBP2. *Am J Hum Genet.* 2009;84(1):44–51.
67. Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet.* 2013;9(8):e1003709.
68. Samocha KE, Robinson EB, Sanders SJ, et al. A framework for the interpretation of *de novo* mutation in human disease. *Nat Genet.* 2014;46(9):944–950.
69. Epi4K Consortium, Epilepsy Phenome/Genome Project. *De novo* mutations in epileptic encephalopathies. *Nature.* 2013;501(7466):217–221.
70. Wang Y, Penfold S, Tang X, et al. Deletion of the *Cul1* gene in mice causes arrest in early embryogenesis and accumulation of cyclin E. *Curr Biol.* 1999;9(20):1191–1194.
71. Crino PB. mTOR signaling in epilepsy: Insights from malformations of cortical development. *Cold Spring Harb Perspect Med.* 2015;5(4):a022442.
72. Baulac S, Ishida S, Marsan E, et al. Familial focal epilepsy with focal cortical dysplasia due to DEPDC5 mutations. *Ann Neurol.* 2015;77(4):675–683.
73. Liu L, Chen ZR, Xu HQ, et al. DEPDC5 variants associated malformations of cortical development and focal epilepsy with febrile seizure plus/febrile seizures: The role of molecular sub-regional effect. *Front Neurosci.* 2020;14:821.
74. Sim NS, Ko A, Kim WK, et al. Precise detection of low-level somatic mutation in resected epilepsy brain tissue. *Acta Neuropathol.* 2019;138(6):901–912.
75. Ribierre T, Deleuze C, Bacq A, et al. Second-hit mosaic mutation in mTORC1 repressor DEPDC5 causes focal cortical dysplasia-associated epilepsy. *J Clin Invest.* 2018;128(6):2452–2458.
76. Chang A, Nishiyama A, Peterson J, Prineas J, Trapp BD. NG2-positive oligodendrocyte progenitor cells in adult human brain and multiple sclerosis lesions. *J Neurosci.* 2000;20(17):6404–6412.
77. Zerlin M, Milosevic A, Goldman JE. Glial progenitors of the neonatal subventricular zone differentiate asynchronously, leading to spatial dispersion of glial clones and to the persistence of immature glia in the adult mammalian CNS. *Dev Biol.* 2004;270(1):200–213.
78. Rosman NP, Pearce J. The brain in multiple neurofibromatosis (von Recklinghausen's disease): A suggested neuropathological basis for the associated mental defect. *Brain.* 1967;90(4):829–838.
79. Zhu Y, Romero MI, Ghosh P, et al. Ablation of NF1 function in neurons induces abnormal development of cerebral cortex and reactive gliosis in the brain. *Genes Dev.* 2001;15(7):859–876.
80. van den Berg DLC, Azzarelli R, Oishi K, et al. Nipbl interacts with Zfp609 and the integrator complex to regulate cortical neuron migration. *Neuron.* 2017;93(2):348–361.
81. Smith L, Singhal N, El Achkar CM, et al. PCDH19-related epilepsy is associated with a broad neurodevelopmental spectrum. *Epilepsia.* 2018;59(3):679–689.