CrossMark

Identification of De Novo *DNMT3A* Mutations That Cause West Syndrome by Using Whole-Exome Sequencing

Zhenwei Liu[1] · Zhongshan Li[1] · Xiao Zhi[1,2] · Yaoqiang Du[1,3] · Zhongdong Lin[4] ·
Jinyu Wu[1]

Abstract Epileptic encephalopathies (EEs) are a group of severe neurodevelopmental disorders with extreme genetic heterogeneity. Recent trio-based whole-exome sequencing (WES) studies have demonstrated that de novo mutations (DNMs) play prominent roles in severe EE. In this study, we searched for potential causal DNMs by using high-coverage WES of four unrelated Chinese parent-offspring trios affected by West syndrome. Through extensive bioinformatic analysis, we identified three novel DNMs in *DNMT3A*, *CDKL5*, and *MAMDC2* in three trios and two compound heterozygous mutations in *KMT2A* in one trio. The DNMs in *CDKL5* and *DNMT3A* were considered to be deleterious on the basis of the consensus of several genetic damage prediction tools. In addition, spatiotemporal expression patterns revealed a high level of *DNMT3A* expression during the early embryonic stage in nearly all brain regions. We also observed that certain high-confidence genes for epilepsy were shared among the co-expression and genetic interaction networks of *DNMT3A*, *CDKL5*, and *KMT2A*. Furthermore, all the candidate epilepsy genes in the co-expression network of *DNMT3A* were significantly enriched in the early developmental stages of the brain according to a rank-based enrichment test. In particular, we found that the DNMs of *DNMT3A* were shared among EE, autism spectrum disorder (ASD), and intellectual disability (ID) and mainly occurred in the functional domain of *DNMT3A*. Together, our findings support an association between *DNMT3A* mutations and EE susceptibility and suggest a shared molecular pathophysiology among EE and other neuropsychiatric disorders.

Keywords Epileptic encephalopathy · De novo · *DNMT3A* · *CDKL5* · Whole-exome sequencing

✉ Jinyu Wu
  iamwujy@gmail.com

1 Institute of Genomic Medicine, Wenzhou Medical University, Wenzhou 325027, China

2 School of Ophthalmology and Optometry, Wenzhou Medical University, Wenzhou 325035, China

3 School of Laboratory Medicine and Life Science, Wenzhou Medical University, Wenzhou 325035, China

4 Department of Pediatric Neurology, The Second Affiliated Hospital and Yuying Children's Hospital, Wenzhou Medical University, Wenzhou 325027, China

Introduction

Epileptic encephalopathies (EEs) represent a large group of severe early-onset conditions that are characterized by intractable seizures, cognitive impairment and neurological deficits associated with frequent epileptiform activity on an electroencephalography (EEG) and that generally have poor prognoses and often lead to developmental delay or regression [1–4]. Among the various forms of EE, West syndrome, or infantile spasms (IS), is one of the most commonly and closely observed epilepsy syndromes and is followed in prevalence by epileptic spasms, psychomotor regression, and a specific EEG pattern called hypsarrhythmia [5–7]. West syndrome usually occurs during the first year of life and is estimated to affect 2–3.5 per 10,000 children per year [4, 8]. The etiology of West syndrome is variable and comprises hereditary and environmental conditions. To a certain extent, this condition also causes some difficulties in reaching a precise diagnosis [4, 7]. Therefore, the investigation of etiological factors,

especially genetic causes, can provide important insights into the mechanisms underlying West syndrome [7, 9].

A genetic etiology is believed to play a prominent role in most patients with West syndrome, and a set of genes linked to West syndrome have recently been reported to harbor mutations in several patients [2, 5, 10, 11]. First, mutations of two genes, ARX and CDKL5, have been identified in patients with X-linked familial West syndrome on the basis of separate family-based linkage and microsatellite-exclusion mapping analyses [12, 13]. Subsequent large studies have also confirmed that mutations in these two genes are the most common known causes of West syndrome [14–17]. In addition, genetic variations in several other genes, including GRIN1, SPTAN1, SLC25A22, and STXBP1, have been found in patients by using a candidate gene sequencing approach and have been strongly implicated in West syndrome [18–21]. The application of array comparative genomic hybridization (CGH) and SNP microarray technologies has identified a few strong but rare candidate genes and loci contributing to West syndrome risk, such as FOXG1, MAGI2, and MEF2C [22–24].

Furthermore, genetic studies in recent years have established that de novo dominant mutations are often responsible for many cases of sporadic West syndrome. De novo mutations (DNMs) in MEF2C, SCN2A, and CDKL5 have recently been identified in five patients with West syndrome, by using high-throughput targeted sequencing of candidate genes located in EE-related CNVs [25]. Furthermore, the application of a whole-exome sequencing (WES) approach based on large cohorts of parent-offspring trios and quartets has offered a cost-effective method for the discovery of DNMs associated with EE across all gene-coding regions in the genome, thus leading to a rapid increase in the identification of determinants underlying West syndrome [26, 27]. In a recent study, a total of 329 DNMs in 305 genes have been identified in patients with EEs, by using WES in 264 parent-child trios [10]. In addition to known EE genes harboring recurrent mutations, including STXBP1 ($n = 4$) and CDKL5 ($n = 2$), a statistically significant enrichment of mutations in two novel genes, GABRB3 and ALG13, has indicated that these genes may be implicated in West syndrome and Lennox-Gastaut syndrome (LGS). Other recent studies using exome sequencing have identified DNMs in several genes associated with West syndrome, such as GRIN2B [28], GNAO1 [29], KCNT1 [30], and SPTAN1 [31]. Owing to the marked genetic heterogeneity underlying West syndrome, each genetic variant probably accounts for a small proportion of cases [2]. Therefore, many more genes associated with West syndrome are likely to be discovered as a result of advances in DNA sequencing and increases in cohort size.

To gain insight into the characterization of DNMs in West syndrome, we performed a WES study with four unrelated Chinese proband-parent trios of subjects with West syndrome. Consequently, we identified two deleterious DNMs in DNMT3A and CDKL5 and two compound heterozygous mutations in KMT2A. In addition, integrated analysis of spatiotemporal expression patterns, co-expression networks, gene expression enrichment, and genetic interaction networks suggests an etiological role of DNMT3A in EE. DNMs in DNMT3A are shared among EE, autism spectrum disorder (ASD), and intellectual disability (ID), thus further indicating that DNMT3A may be involved in the onset of sporadic neuropsychiatric disorders.

Methods

Patient Recruitment

Four unrelated trios of Han Chinese subjects with West syndrome were recruited from the Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University. This research protocol was approved by the ethics committee of Wenzhou Medical University and was carried out in accordance with the approved guidelines. Written informed consent for clinical and genetic analyses was obtained from all the guardians of the subjects participating in the study. Clinical evaluations of all the patients consisted of an assessment of the presentation of clinical seizures and an EEG recorded by a pediatrician experienced in EEs.

Whole-Exome Sequencing

Genomic DNA was isolated from peripheral blood leukocytes through standard phenol/chloroform extraction protocols. DNA quality and quantity were assessed by agarose gel electrophoresis and a NanoDrop 2000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA). A total of 2 μg of genomic DNA from each sample was used for library construction with an Agilent SureSelect Library Prep Kit according to the manufacturer's protocol. Exome libraries were captured using an Agilent SureSelect Human All Exon v5 Kit (Agilent Technologies, Santa Clara, CA, USA) and were sequenced with an Illumina HiSeq 4000 through a 150-bp paired-end run (Illumina, San Diego, CA, USA).

Data Processing and Variant Calling

The raw data generated by WES were processed using the Trim Galore program to filter out adapters and low-quality reads. Then, the remaining reads were aligned/mapped to the human reference genome (GRCH37/hg19) by using the BWA program (version 0.7.12) [32]. After alignment, duplicated reads were removed using Picard tools, and only uniquely mapped reads were used for the detection of variation. After this process, three GATK tools, including RealignerTargetCreator, IndelRealigner, and BaseRecalibrator, were incorporated for

realigning and calling the single-nucleotide variants (SNVs) and InDels, as well as for recalibrating the quality scores for each variant [33]. Finally, the sequencing depth and coverage were calculated on the basis of the unique aligned reads.

Detection and Annotation of De Novo SNVs and InDels

The ForestDNM and mirTrios programs were used to detect de novo SNVs [34]. We also utilized mirTrios software to identify putative de novo InDels and rare inherited variants. All the variants were annotated with ANNOVAR and an in-house bioinformatics tool with UCSC annotation (http://www.ncbi.nlm.nih.gov/refseq/).

To further assess the effects of missense variants, SIFT, LRT, SiPhy, and VEST3 were applied to obtain functional predictions. We considered a missense variant deleterious if the variant was consistently predicted to be deleterious or damaging in all four genetic damage prediction tools. For each variant, we obtained the minor allele frequency (MAF) according to normal population variant databases, including dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/), the 1000 Genomes Project (http://www.1000genomes.org/), ESP6500 (http://evs.gs.washington.edu/EVS/), ExAC(http://exac.broadinstitute.org/), CG69 (http://www.completegenomics.com/public-data/69-genomes/), and GWAS Catalog (https://www.ebi.ac.uk/gwas/). The variants with a MAF >0.1% were removed.

Primers and Sanger Sequencing Validation

Validation of each potential DNM was confirmed through conventional PCR assays and Sanger sequencing. All the primers used in these assays are listed in Table S1.

Construction of Gene Co-expression and Genetic Interaction Networks

To build the gene co-expression network, we utilized spatially and temporally rich transcriptome data generated from the BrainSpan project (http://www.brainspan.org/). The Pearson correlation coefficients (r) for the gene co-expression levels were calculated for each pairwise combination of genes. Strong co-expression (absolute correlation $r \geq 0.7$) was selected between seed genes and correlated epilepsy candidate genes determined on the basis of the EpilepsyGene database [35]. To create the genetic interaction network, genetic interaction data were downloaded from GeneMANIA by utilizing the Cytoscape program plugin [36]. Afterward, we used GeneMANIA to calculate the numbers of nodes and edges between the seed genes and epilepsy candidate genes determined on the basis of the EpilepsyGene database. Both the gene co-expression and genetic interaction networks were visualized using the Cytoscape program [37].

Gene Co-expression Enrichment in Specific Brain Regions

All the genes in the co-expression network were collected to analyze the spatiotemporal expression enrichment in the 12 developmental stages, which contained expression data for all 16 brain regions. We used a previously published method [38] based on a modified z-score ≥ 2 to compute a gene expression signature for the gene lists in all 192 combinations of brain regions and developmental stages. The significance of enrichment was determined with a rank-based enrichment test of the spatiotemporal signatures with 100,000 permutations. Multiple tests were used to correct for all 192 region-stage combinations to obtain Q values by using the Benjamini-Hochberg correction (FDR) [39].

Results

Detection of DNMs

Four trios, each comprising an affected child and two unaffected parents, were included in this WES study. After the removal of adapters and low-quality bases, each sample had 6.96–11.26 Gb of clean data, with 49.01 Mb of bases covering the target regions (Table S2). In general, each sample had at least 99.15% of reads aligned to the reference genome, and on average, approximately 56.00% of effective bases were located in the target regions after removing PCR duplications. A minimum of 95.20% of the target regions were covered at least 4-fold, 92.20% of the target regions were covered at least 10-fold, and 84.00% of the target regions were covered at least 20-fold. Above all, these data fully reflect the reliability of our sequencing, thus providing a reliable basis for the follow-up analysis.

De novo SNVs and InDels in coding regions were identified in each sample by using ForestDNM and mirTrios software. To minimize false-negative results, the DNMs identified by both of these software approaches were considered to be putative DNMs and subjected to further annotation and Sanger sequencing validation. After PCR and Sanger sequencing validation, three de novo SNVs (Table 1) in coding regions of the four trios were confirmed. In addition, we also detected two compound heterozygous mutations in one trio by utilizing mirTrios software in this study.

Identification of EE-Associated Genes with DNMs

Recently, WES of parent-offspring trios has established that DNMs are responsible for many EE phenotypes, although the size of this contribution is not yet known [2]. For all the DNMs in various genes detected in this study, a statistical assessment for potential causative genes was not performed

Table 1 Summary of DNMs detected by trios-based WES of EE

Trio	Chr	Gene	Position (hg19)	Mutation	Inheritance	Protein change	SIFT	VEST3	LRT	SiPhy
E1	Chr9	*MAMDC2*	72758660	Missense	De novo	p.I443M	Tolerable	Tolerable	Deleterious	Nonconserved
E2	Chr2	*DNMT3A*	25467120	Missense	De novo	p.M585I	Damaging	Damaging	Deleterious	Conserved
E3	ChrX	*CDKL5*	18602447	Stop-gain	De novo	p.W176X	–	–	–	–
E4	Chr11	*KMT2A*	118354977	Missense	Paternal	p.S1389C	Damaging	Damaging	Neutral	Conserved
			118375998	Missense	Maternal	p.G3128S	Tolerable	Damaging	Deleterious	Conserved

because of the small sample size. Nevertheless, we assessed pathogenic de novo variants on the basis of the concordance of four functional prediction tools (SIFT, VEST3, LRT, and SiPhy). According to the mutation type and the effect of each mutation on protein function, two de novo mutations in two genes (*CDKL5* and *DNMT3A*) were clearly predicted to be functionally deleterious by all four of these prediction tools (Table 1).

In the proband E3P, the validated de novo variant was a G>A nonsense transition (c.528G>A) in the coding sequence of *CDKL5* (MIM 300203), which changed a tryptophan (Trp) at position 176 into a premature termination codon (p.Trp176X) (Fig. S1; Table 1). As previously reported, *CDKL5* is expressed primarily in the brain, muscles, and thymus [40] and is known to cause encephalopathy, an X-linked dominantly inherited disorder characterized by early infantile epileptic encephalopathy or atypical Rett syndrome [41].

By Sanger sequencing, we also confirmed that a novel missense mutation consisting of a G to T substitution (c.1755G>T) in *DNMT3A* (MIM 602769) was present in the proband E2P. This single-nucleotide change produced a nonsynonymous substitution (p.M585I), and similarly, the mutation was also suggested to be a damaging missense variant according to the prediction software tools (Table 1). One other DNM was detected in *MAMDC2* in the proband E1P consisting of a nonsynonymous mutation (c.1329C>G) resulting in an isoleucine (Ile) to methionine (Met) substitution at amino acid 443 (p.I443M) in the MAM3 conserved domain. Notably, this mutation was thought to be tolerable or nonconserved on the basis of the predictions of the SIFT, VEST3, and SiPhy software tools (Fig. S1; Table 1).

In addition to the above DNMs, we identified compound heterozygous mutations of *KMT2A* in the proband E4P (Fig. S1; Table 1). A single-nucleotide substitution (c.4166C>G) inherited from the father was found in *KMT2A* (MIM 159555), thus resulting in an amino acid alteration (p.S1389C). A second nonsynonymous mutation (c.9382G>A, p.G3128S) was also identified, which was inherited from the mother. The functional prediction revealed that the two variants were individually suggested to be damaging or deleterious by three out of four prediction tools (Table 1).

Expression Profile of *DNMT3A* in the Human Brain

One of the underlying phenotypes of EE is abnormal brain development. To further investigate the potential contribution to EE of the damaging DNM in *DNMT3A*, we studied the spatiotemporal expression patterns of this gene in the human brain, on the basis of two specialized human brain gene expression databases: BrainSpan and the Human Brain Transcriptome (HBT, http://hbatlas.org/). The expression levels of *DNMT3A* for 16 different brain regions (cerebellar cortex, mediodorsal nucleus of the thalamus, striatum, amygdala, hippocampus, and 11 areas of the neocortex) and 12–15 different stages from embryonic development to adulthood were obtained from the BrainSpan and HBT databases (Fig. 1). First, it was found that *DNMT3A* was widely expressed across all the human brain regions as well as all the different developmental stages analyzed. Interestingly, continuously decreasing *DNMT3A* expression levels were observed from early embryonic stage to late adulthood in all the brain regions analyzed. Notably, a drastic decrease in *DNMT3A* expression occurred near birth and the expression level was maintained at a stable and low level thereafter. In addition, the spatiotemporal expression profile of *DNMT3A* in the human brain described above was highly consistent between the two different databases used, thus indicating the reliability of this finding. Overall, the widespread expression of *DNMT3A* in different human brain regions, especially its markedly higher expression level at early embryonic stages, indicates an important role of *DNMT3A* in early brain development. Therefore, disruption of this gene may cause diseases related to brain dysfunction, such as EE.

Co-expression and Genetic Interaction Network Analyses of *DNMT3A*

Epilepsy-related genes are often involved in brain development. To further interrogate the relationships of *DNMT3A*, *CDKL5*, and *KMT2A* with other genes associated with epilepsy in the context of human brain development, a co-expression network analysis was performed by utilizing spatially and temporally rich transcriptome data from the BrainSpan project. We observed a clear co-expression relationship between each of the three genes and candidate epilepsy genes obtained from the EpilepsyGene database
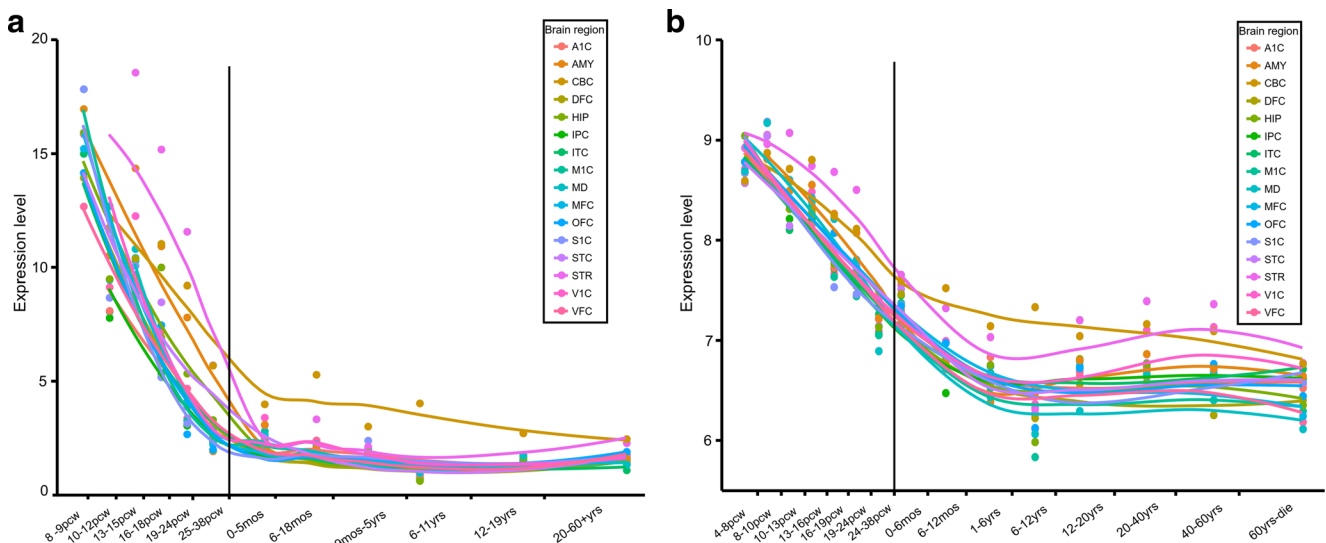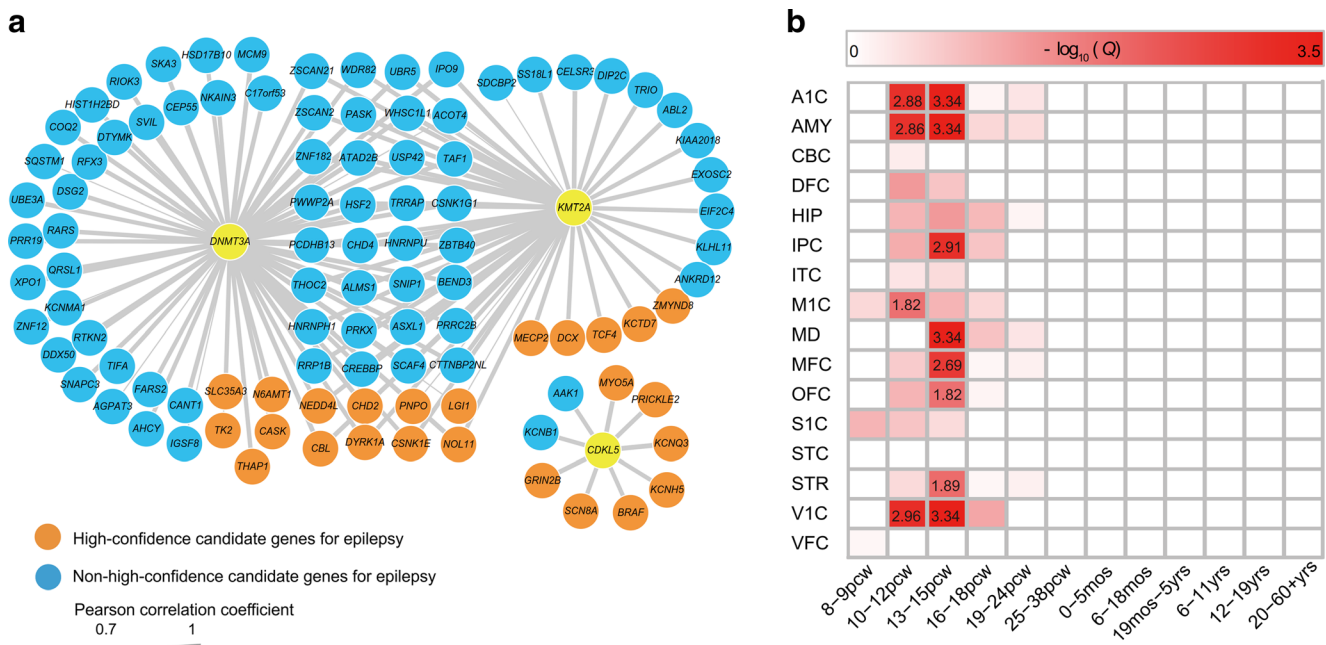
Fig. 1 Expression analysis of *DNMT3A* in 16 human brain regions. Expression profiles of *DNMT3A* in BrainSpan (**a**) and HBT (**b**). The expression levels of *DNMT3A* are shown for the developmental stages from embryonic development to late adulthood. A *solid line* separates the periods into prenatal periods and postnatal periods. *CBC* cerebellar cortex, *MD* mediodorsal nucleus of the thalamus, *STR* striatum, *AMY*

amygdala, *HIP* hippocampus, *OFC* orbital prefrontal cortex, *DFC* dorsolateral prefrontal cortex, *VFC* ventrolateral prefrontal cortex, *MFC* medial prefrontal cortex, *M1C* primary motor cortex, *S1C* primary somatosensory cortex, *IPC* posterior inferior parietal cortex, *A1C* primary auditory cortex, *STC* posterior superior temporal cortex, *ITC* inferior temporal cortex, *V1C* primary visual cortex

(Fig. 2a). Among these networks, *DNMT3A* was co-expressed with 75 epilepsy candidate genes, including 13 predicted high-confidence genes for epilepsy, on the basis of gene prioritization conducted by adopting the annotations of 10 functional prediction tools from the EpilepsyGene database

(Fig. 2a). Among the high-confidence genes, *CHD2* was reported as a known causal gene for ASD and ID [25, 42, 43], and another three genes (*NEDD4L*, *CASK*, and *DYRK1A*) were considered to be associated with ASD [44]. In addition, *KMT2A* was co-expressed with 56 epilepsy candidate genes,

Fig. 2 Co-expression analysis and specific spatiotemporal patterns of expression in the brain. **a** Co-expression network analysis of *DNMT3A*, *CDKL5*, and *KMT2A*. Gene co-expression levels were estimated using the Pearson correlation coefficients (*r*) between each pair of genes. Candidate epilepsy genes were extracted from the EpilepsyGene database. **b** Enrichment of spatiotemporal gene expression signatures with

the top-ranking genes spatially and temporally co-expressed with *DNMT3A* in the brain at specific developmental stages. *P* values were calculated according to a rank-based score followed by a permutation test and then converted to *Q* values for multiple hypothesis testing on the basis of FDR

including 13 high-confidence genes for epilepsy (Fig. 2a). Moreover, we found that 40 epilepsy candidate genes, including 8 high-confidence genes for epilepsy, were shared among the co-expression networks of *DNMT3A* and *KMT2A* (Fig. 2a). Notably, the shared high-confidence gene *CHD2* was also predicted to be an ID candidate gene [25, 45]. The known causal gene for EE, *CDKL5*, was co-expressed with nine epilepsy candidate genes, most of which were also high-confidence genes (Fig. 2a).

To determine whether the co-expressed genes are enriched in specific brain regions or developmental stages, we used a permutation test in 100,000 iterations based on gene expression during human brain development to investigate the enrichment of the co-expressed genes across 16 human cortical and subcortical structures during 12 human developmental periods, from early embryonic stages to late adulthood. The *DNMT3A* and the co-expressed genes were significantly enriched ($Q < 0.05$) across two different early developmental stages: the early prenatal (10–12 PCW) and early mid-prenatal (13–15 PCW) stages (Fig. 2b). In contrast, significant enrichment was widely spatially distributed across many brain regions, including the amygdala (AMY), mediodorsal nucleus of the thalamus (MD), striatum (STR), and six areas of the neocortex (the primary auditory cortex (A1C), posterior inferior parietal cortex (IPC), primary motor cortex (M1C), medial prefrontal cortex (MFC), orbital prefrontal cortex (OFC), and primary visual cortex (V1C)) (Fig. 2b). Specifically, the most

significant enrichment was observed in the A1C, AMY, MD, and V1C regions during the early prenatal stage (13–15 PCW, $Q = 4.57 \times 10^{-4}$) (Fig. 2b). These findings also indicated that genetic changes in *DNMT3A* and the co-expressed epilepsy-associated genes may affect early human brain development.

Furthermore, to gain insight into the functional associations of *DNMT3A*, *CDKL5*, and *KMT2A* with other genes related to epilepsy, we constructed a gene-interaction network model based on the interaction dataset collected from the GeneMANIA database. As a result, we observed varying degrees of genetic interactions between *DNMT3A* and 20 candidate epilepsy genes, including 8 high-confidence genes for epilepsy, obtained from the EpilepsyGene database (Fig. 3). Among the high-confidence genes, the *GRIN2B* was found to be shared with ASD and ID [46] and the *CNTNAP2* and *UBE3A* were considered as high likely candidate genes for ASD [44, 46]. Similarly, *KMT2A* displayed genetic interactions with 42 candidate epilepsy genes, including 17 high-confidence genes. The known causal gene *CDKL5* was associated with 40 candidate epilepsy genes, including 19 high-confidence genes. Notably, eight genes had interactions with both *DNMT3A* and *KMT2A*, whereas nine genes had interactions with both *KMT2A* and *CDKL5*, and five genes had interactions with both *DNMT3A* and *CDKL5* (Fig. 3). In particular, three high-confidence genes for epilepsy, *UBE3A*, *IER3IP1*, and *CNTNAP2* were present in the genetic interaction networks of *DNMT3A*, *CDKL5*, and *KMT2A*.
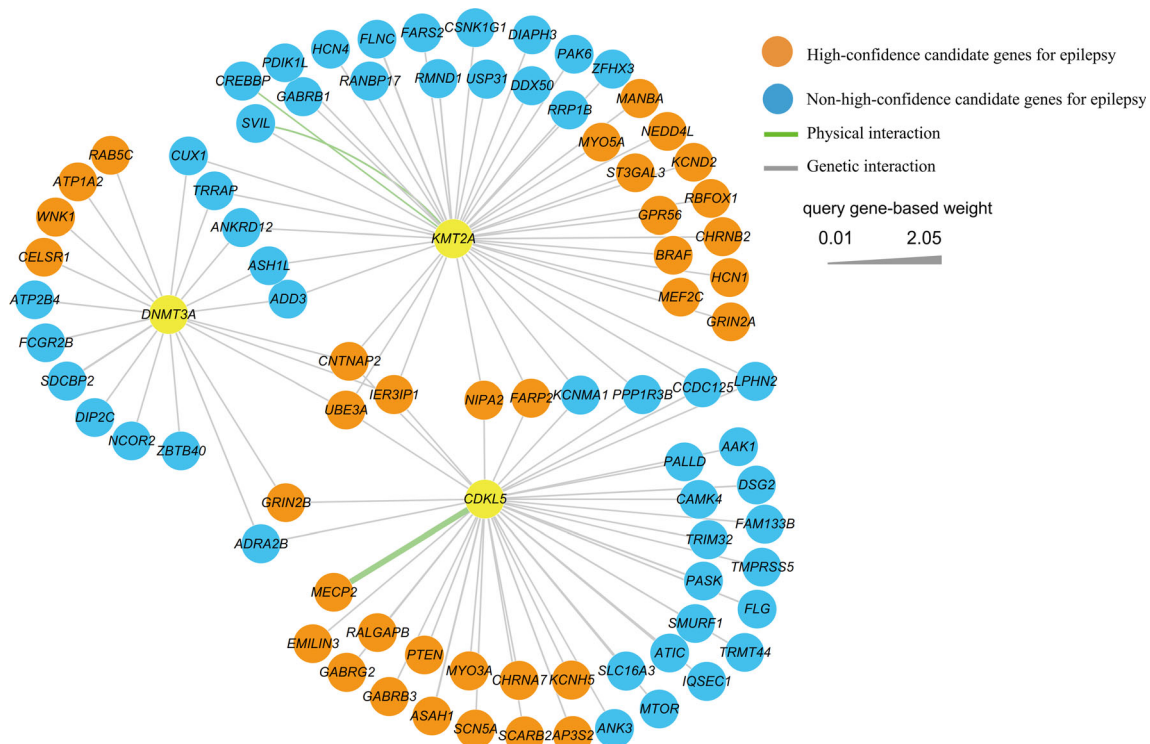
Fig. 3 Genetic interaction network analysis of *DNMT3A, CDKL5,* and *KMT2A*. Nodes denote genes, and edges denote interactions between two genes. The thickness of an edge denotes the degree of genetic interaction. Candidate epilepsy genes were extracted from the EpilepsyGene database

Furthermore, we also found that *KMT2A* and *SVIL* have not only a genetic interaction but also display a physical interaction between both gene products. In addition, *CDKL5* and the high-confidence gene *MECP2* also exhibit a strong physical interaction (Fig. 3). These results suggest that *DNMT3A*, *CDKL5*, and *KMT2A* may have similar functions regarding interaction with several candidate epilepsy genes.
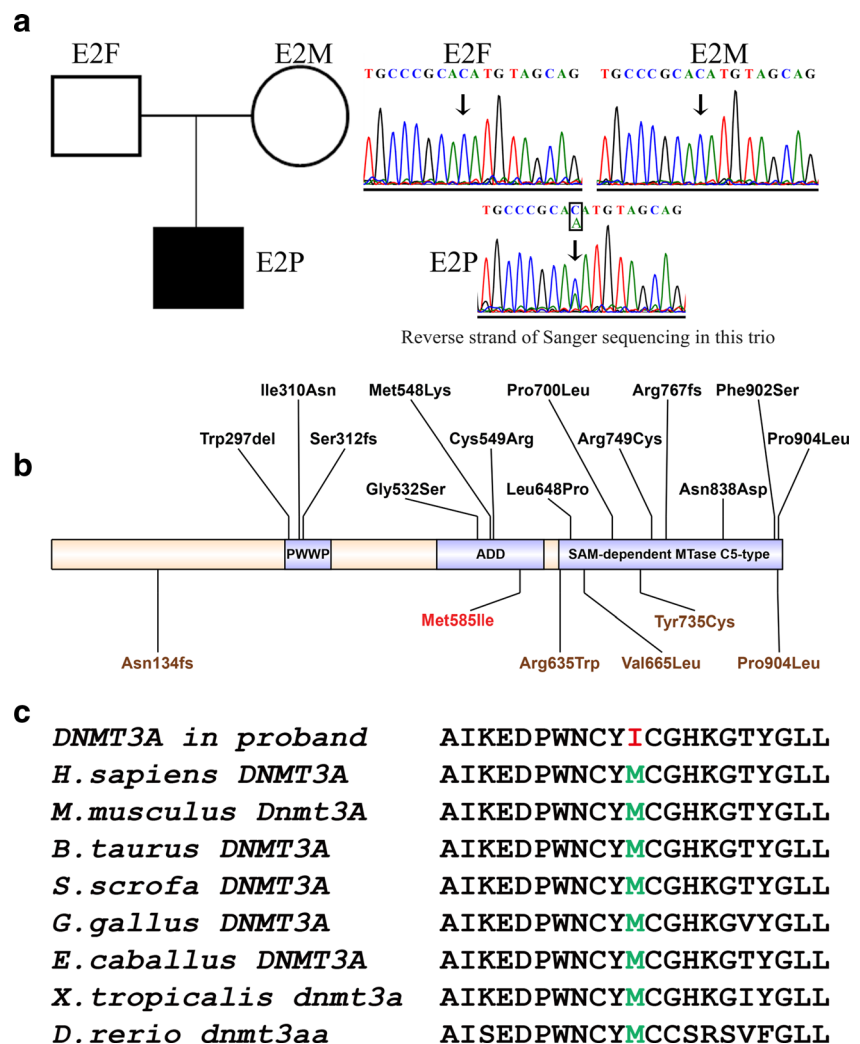
Discussion

Advances in WES in recent years have accelerated understanding of the genetic etiology of sporadic EE, in which DNMs have been shown to be dominant contributors to causality [1]. Certain genes implicated in West syndrome have been identified in larger cohorts of trios with affected offspring [10, 14]. In this work, we used WES to search for DNMs in four parent-offspring trios affected by West syndrome in an effort to identify potential disease-causing DNMs on the basis of a bioinformatics analysis. As a result,

we successfully identified three coding DNMs in three probands and two compound heterozygous mutations in one proband.

With proband E3P, one stop-gain DNM of the *CDKL5* was identified and found to potentially cause protein truncation at position 176, thus suggesting that the function of the mutant protein would be impaired. Notably, the *CDKL5* is a well-documented known causal gene for EE [12] and other severe brain disorders [47]. In one study of ten sporadic EE trios using WES, one de novo stop-gain mutation at position 832 of the protein encoded by the *CDKL5* has been found to be a likely cause in one proband [14]. Notably, the truncation of the CDKL5 protein at the upstream position 176 found in our study is expected to cause more severe defects than the truncation of this protein at the downstream position 832 found in the referenced study. In another large-scale study of 264 EE trios that has used WES, three DNMs have been found on the *CDKL5*, and the significantly high mutation rate observed suggests that *CDKL5* is one of the most common disease-causing genes [10].

Fig. 4 DNMs identified in *DNMT3A*. **a** The DNM of *DNMT3A* (c.1755G>T, p.M585I) was confirmed in E2P by using Sanger sequencing. **b** Schematic representation of the DNMT3A protein and the DNMs of DNMT3A identified in EE, ASD, and ID. Amino acid changes are shown in *black* for ID, *brown* for ASD, and *red* for EE. **c** The conservation of the variant c.1755G>T, p.M585I among different species

DNMT3A in proband	AIKEDPWNCYICGHKGTYGLL
H.sapiens DNMT3A	AIKEDPWNCYMCGHKGTYGLL
M.musculus Dnmt3A	AIKEDPWNCYMCGHKGTYGLL
B.taurus DNMT3A	AIKEDPWNCYMCGHKGTYGLL
S.scrofa DNMT3A	AIKEDPWNCYMCGHKGTYGLL
G.gallus DNMT3A	AIKEDPWNCYMCGHKGVYGLL
E.caballus DNMT3A	AIKEDPWNCYMCGHKGTYGLL
X.tropicalis dnmt3a	AIKEDPWNCYMCGHKGIYGLL
D.rerio dnmt3aa	AISEDPWNCYMCCSRSVFGLL

The *DNMT3A* is located on chromosome 14q32 and encodes a DNA methyltransferase that is essential for establishing methylation during embryogenesis and can methylate unmethylated and hemimethylated DNA, thereby playing important roles in genomic imprinting and X chromosome inactivation [48]. In this study, with the aid of a trio-based WES, we found a de novo SNV (c.1755G>T, p.Met585Ile) located in exon 8 of *DNMT3A* in the proband E2P (Fig. 4a, b). We also found that this mutation was not present in any large control population study, including the Exome Aggregation Consortium database, 1000 Genomes Project, dbSNP 147, and the NHLBI Exome Sequencing Project. Similarly, according to a search of the published literature, this mutation has not been previously reported. Additionally, the residue Met585 is evolutionarily conserved among different vertebrate species, and the variant was predicted to be deleterious on the basis of multiple bioinformatic algorithms (Fig. 4c; Table 1). In the analysis of the spatial and temporal expression profiles of *DNMT3A* in the human brain, high expression levels of *DNMT3A* at the embryonic stage (4–8 PCW) across multiple brain regions indicated that the gene may play an important role in early brain development. The co-expression network analysis in BrainSpan suggested that *DNMT3A* is co-expressed with several genes known to be associated with epilepsy in different human brain regions and developmental stages. Furthermore, the genetic interaction network analysis provided additional evidence that the *DNMT3A* can interact with genes known to

be associated with epilepsy. These observations suggest that *DNMT3A* may play an important role in epilepsy.

Notably, the DNMs have been repeatedly reported in WES studies of probands with different sporadic neuropsychiatric disorders, such as EE, ASD, ID, and SCZ [46, 49]. In a recent study of 13 unrelated patients with human overgrowth syndromes, characterized by ID and a distinctive facial appearance, 13 different de novo heterozygous mutations have been identified in *DNMT3A* (Fig. 4b; Table 2) [50]. Notably, seizures have been reported in two unrelated patients, thus further suggesting an important role of *DMNT3A* in epilepsy. Additionally, other four different de novo nonsynonymous mutations and one de novo frameshift mutation in *DNMT3A* associated with ASD have been found in five different trios from three different cohorts (Fig. 4b; Table 2) [42, 43, 51]. These results suggest that DNMs of *DNMT3A* are shared among EE, ASD, and ID. In fact, a shared genetic etiology between EE and other neuropsychiatric disorders, including ASD and ID, has previously been reported [52], and in this study, we report a new gene shared among EE, ASD, and ID. We are not clear about the exact mechanisms for genotype-phenotype correlation with respect to different *DNMT3A* mutations. However, it is proposed that the nature of the different mutations, alternative neurobiological conditions, genetic background, and other factors including environmental influence, mutation timing, and epigenetic factors could affect the genotype to phenotype relationship [2, 25, 50, 53].

Table 2 The shared DNMs of *DNMT3A* among ID, ASD, and EE

Chr	Mutation	Protein change	Mutation type	Disorder	Method	Reference
Chr2	c.889_891delTGG	p.Trp297del	Deletion	ID	WES	Tatton-Brown et al. [50]
Chr2	c.929T>A	p.Ile310Asn	SNV	ID	WES	Tatton-Brown et al. [50]
Chr2	c.934_937dupTCTT	p.Ser312fs	Insertion	ID	WES	Tatton-Brown et al. [50]
Chr2	c.1594G>A	p.Gly532Ser	SNV	ID	WES	Tatton-Brown et al. [50]
Chr2	c.1643T>A	p.Met548Lys	SNV	ID	WES	Tatton-Brown et al. [50]
Chr2	c.1645T>C	p.Cys549Arg	SNV	ID	WES	Tatton-Brown et al. [50]
Chr2	c.1943T>C	p.Leu648Pro	SNV	ID	WES	Tatton-Brown et al. [50]
Chr2	c.2099C>T	p.Pro700Leu	SNV	ID	WES	Tatton-Brown et al. [50]
Chr2	c.2245C>T	p.Arg749Cys	SNV	ID	WES	Tatton-Brown et al. [50]
Chr2	c.2297dupA	p.Arg767fs	Insertion	ID	WES	Tatton-Brown et al. [50]
Chr2	c.2512A>G	p.Asn838Asp	SNV	ID	WES	Tatton-Brown et al. [50]
Chr2	c.2705T>C	p.Phe902Ser	SNV	ID	WES	Tatton-Brown et al. [50]
Chr2	c.2711C>T	p.Pro904Leu	SNV	ID	WES	Tatton-Brown et al. [50]
Chr2	c.1903C>T	p.Arg635Trp	SNV	ASD	WGS	Jiang et al. [51]
Chr2	c.2204A>G	p.Tyr735Cys	SNV	ASD	WES	De Rubeis et al. [43]
Chr2	c.2711C>T	p.Pro904Leu	SNV	ASD	WES	Iossifov et al. [42]
Chr2	c.1993G>T	p.Val665Leu	SNV	ASD	WES	Iossifov et al. [42]
Chr2	c.401dupA	p.Asn134fs	Insertion	ASD	WES	Iossifov et al. [42]
Chr2	c.1755G>T	p.Met585Ile	SNV	EE	WES	In this study

On the basis of the predictions from the UniProt database, DNMT3A contains three functional domains: a PWWP domain, an ADD domain and a C-terminal SAM-dependent MTase C5-type domain (Fig. 4b). Notably, most of the DNMs in DNMT3A were identified in these functional domains in ASD, ID, and EE, with the exception of a frameshifting insertion (c.401dupA, p.Asn134fs) in ASD (Fig. 4b; Table 2). It has been shown that these domains play an important role in DNMT3A function. In this study, the DNM resulting in Met585Ile occurred in the ADD domain, which inhibits the enzymatic activity of the catalytic domain by blocking the binding of this domain to DNA, according to previous structural and biochemical analyses [54]. The disruption of the ADD domain may affect the recognition of unmethylated lysine 4 of histone H3 and, therefore, the regulation of DNA methylation, thus indicating that the ADD domain of DNMT3A is essential for establishing methylation [54]. In addition, two paralogs of DNMT3A: DNMT3B and DNMT3L were also examined as for their possible role in EE or other neuropsychiatric disorders based on previous studies. DNMT3B was considered a candidate gene for ID [55]. As for DNMT3L, the interaction of DNMT3L with DNMT3A and DNMT3B was well documented [50], which may indicate its role in similar genetic disorders.

In the proband E2P, we successfully identified two compound heterozygous mutations (p.S1389C and p.G3128S) in KMT2A (MLL1), which were individually predicted to be pathogenic by three bioinformatic algorithms. KMT2A encodes a histone methyltransferase that methylates histone H3 lysine 4 (H3K4) and mediates chromatin modifications, playing an essential role in the regulation of gene expression during early development and hematopoiesis [56]. It has been reported that MLL1-deficient mouse neural stem cells located in the subventricular zone could efficiently differentiate into glial lineage cells; hence, KMT2A might be implicated in neurogenesis in the mouse postnatal brain [57]. To date, this gene has not been clearly determined to be associated with any neuropsychiatric disorders. Nevertheless, in one recent study of 2270 trios with ASD, two different DNMs in KMT2A have been described in two unrelated ASD probands [43]. Another DNM in KMT2A has also been identified in a proband with ASD, by using a WES approach [42]. Moreover, one DNM (c.334A>G; p.M112V) in KMT2A has recently been described in a gene discovery study that conducted WES in 264 trios [10]. Additionally, analogously to the spatiotemporal expression and enrichment patterns for DNMT3A, KMT2A exhibited high expression levels and was co-expressed with epilepsy-associated genes that were also significantly enriched in early brain developmental periods across widespread brain regions (Fig. S2). These observations indicated that KMT2A may be a potential candidate gene for EE.

Through the results obtained by performing high-coverage WES and an in-depth bioinformatics analysis, we identified three different de novo coding mutations in the four trios with West syndrome. In addition to the clear link of CDKL5 to EE, the heterozygous DNMs of DNMT3A identified in this study suggest an important role of DNMT3A in EE, but elucidation of the molecular mechanism will require further genetic and functional studies. In particular, the application of WES should aid further understanding of the genetic architecture of EE and facilitate the molecular diagnosis of EE.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no competing interests.

References

1. Myers CT, Mefford HC (2016) Genetic investigations of the epileptic encephalopathies: recent advances. Prog Brain Res 226:35–60. doi:10.1016/bs.pbr.2016.04.006

2. McTague A, Howell KB, Cross JH et al (2016) The genetic landscape of the epileptic encephalopathies of infancy and childhood. Lancet Neurol 15(3):304–316. doi:10.1016/S1474-4422(15)00250-1

3. Auvin S, Cilio MR, Vezzani A (2016) Current understanding and neurobiology of epileptic encephalopathies. Neurobiol Dis 92(Pt A):72–89. doi:10.1016/j.nbd.2016.03.007

4. Nieh SE, Sherr EH (2014) Epileptic encephalopathies: new genes and new pathways. Neurotherapeutics 11(4):796–806. doi:10.1007/s13311-014-0301-2

5. Shbarou R, Mikati MA (2016) The expanding clinical spectrum of genetic pediatric epileptic encephalopathies. Semin Pediatr Neurol 23(2):134–142. doi:10.1016/j.spen.2016.06.002

6. Covanis A (2012) Epileptic encephalopathies (including severe epilepsy syndromes). Epilepsia 53(Suppl 4):114–126. doi:10.1111/j.1528-1167.2012.03621.x

7. Pellock JM, Hrachovy R, Shinnar S et al (2010) Infantile spasms: a U.S. consensus report. Epilepsia 51(10):2175–2189. doi:10.1111/j.1528-1167.2010.02657.x

8. Saemundsen E, Ludvigsson P, Rafnsson V (2008) Risk of autism spectrum disorders after infantile spasms: a population-based study nested in a cohort with seizures in the first year of life. Epilepsia 49(11):1865–1870. doi:10.1111/j.1528-1167.2008.01688.x

9. Paciorkowski AR, Thio LL, Dobyns WB (2011) Genetic and biologic classification of infantile spasms. Pediatr Neurol 45(6):355–367. doi:10.1016/j.pediatrneurol.2011.08.010

10. Allen AS, Berkovic SF, Cossette P et al (2013) De novo mutations in epileptic encephalopathies. Nature 501(7466):217–221. doi:10.1038/nature12439

11. Kato M (2006) A new paradigm for West syndrome based on molecular and cell biology. Epilepsy Res 70(Suppl 1):S87–S95. doi:10.1016/j.eplepsyres.2006.02.008

12. Weaving LS, Christodoulou J, Williamson SL et al (2004) Mutations of CDKL5 cause a severe neurodevelopmental disorder with infantile spasms and mental retardation. Am J Hum Genet 75(6):1079–1093. doi:10.1086/426462

13. Stromme P, Mangelsdorf ME, Shaw MA et al (2002) Mutations in the human ortholog of Aristaless cause X-linked mental retardation and epilepsy. Nat Genet 30(4):441–445. doi:10.1038/ng862

14. Veeramah KR, Johnstone L, Karafet TM et al (2013) Exome sequencing reveals new causal mutations in children with epileptic encephalopathies. Epilepsia 54(7):1270–1281. doi:10.1111/epi.12201

15. Kato M, Koyama N, Ohta M et al (2010) Frameshift mutations of the ARX gene in familial Ohtahara syndrome. Epilepsia 51(9):1679–1684. doi:10.1111/j.1528-1167.2010.02559.x

16. Giordano L, Sartori S, Russo S et al (2010) Familial Ohtahara syndrome due to a novel ARX gene mutation. Am J Med Genet A 152A(12):3133–3137. doi:10.1002/ajmg.a.33701

17. Bahi-Buisson N, Nectoux J, Rosas-Vargas H et al (2008) Key clinical features to identify girls with CDKL5 mutations. Brain 131(Pt 10):2647–2661. doi:10.1093/brain/awn197

18. Saitsu H, Tohyama J, Kumada T et al (2010) Dominant-negative mutations in alpha-II spectrin cause West syndrome with severe cerebral hypomyelination, spastic quadriplegia, and developmental delay. Am J Hum Genet 86(6):881–891. doi:10.1016/j.ajhg.2010.04.013

19. Ding YX, Zhang Y, He B et al (2010) A possible association of responsiveness to adrenocorticotropic hormone with specific GRIN1 haplotypes in infantile spasms. Dev Med Child Neurol 52(11):1028–1032. doi:10.1111/j.1469-8749.2010.03746.x

20. Molinari F, Kaminska A, Fiermonte G et al (2009) Mutations in the mitochondrial glutamate carrier SLC25A22 in neonatal epileptic encephalopathy with suppression bursts. Clin Genet 76(2):188–194. doi:10.1111/j.1399-0004.2009.01236.x

21. Saitsu H, Kato M, Mizuguchi T et al (2008) De novo mutations in the gene encoding STXBP1 (MUNC18-1) cause early infantile epileptic encephalopathy. Nat Genet 40(6):782–788. doi:10.1038/ng.150

22. Brunetti-Pierri N, Paciorkowski AR, Ciccone R et al (2011) Duplications of FOXG1 in 14q12 are associated with developmental epilepsy, mental retardation, and severe speech impairment. Eur J Hum Genet 19(1):102–107. doi:10.1038/ejhg.2010.142

23. Nowakowska BA, Obersztyn E, Szymanska K et al (2010) Severe mental retardation, seizures, and hypotonia due to deletions of MEF2C. Am J Med Genet B Neuropsychiatr Genet 153B(5):1042–1051. doi:10.1002/ajmg.b.31071

24. Marshall CR, Young EJ, Pani AM et al (2008) Infantile spasms is associated with deletion of the MAGI2 gene on chromosome 7q11.23-q21.11. Am J Hum Genet 83(1):106–111. doi:10.1016/j.ajhg.2008.06.001

25. Carvill GL, Heavin SB, Yendle SC et al (2013) Targeted resequencing in epileptic encephalopathies identifies de novo mutations in CHD2 and SYNGAP1. Nat Genet 45(7):825–830. doi:10.1038/ng.2646

26. Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet 17(6):333–351. doi:10.1038/nrg.2016.49

27. Bamshad MJ, Ng SB, Bigham AW et al (2011) Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet 12(11):745–755. doi:10.1038/nrg3031

28. Lemke JR, Hendrickx R, Geider K et al (2014) GRIN2B mutations in West syndrome and intellectual disability with focal epilepsy. Ann Neurol 75(1):147–154. doi:10.1002/ana.24073

29. Nakamura K, Kodera H, Akita T et al (2013) De novo mutations in GNAO1, encoding a Galphao subunit of heterotrimeric G proteins, cause epileptic encephalopathy. Am J Hum Genet 93(3):496–505. doi:10.1016/j.ajhg.2013.07.014

30. Ohba C, Kato M, Takahashi N et al (2015) De novo KCNT1 mutations in early-onset epileptic encephalopathy. Epilepsia 56(9):e121–e128. doi:10.1111/epi.13072

31. Writzl K, Primec ZR, Strazisar BG et al (2012) Early onset West syndrome with severe hypomyelination and coloboma-like optic discs in a girl with SPTAN1 mutation. Epilepsia 53(6):e106–e110. doi:10.1111/j.1528-1167.2012.03437.x

32. Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26(5):589–595. doi:10.1093/bioinformatics/btp698

33. McKenna A, Hanna M, Banks E et al (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20(9):1297–1303. doi:10.1101/gr.107524.110

34. Li J, Jiang Y, Wang T et al (2015) mirTrios: an integrated pipeline for detection of de novo and rare inherited mutations from trios-based next-generation sequencing. J Med Genet 52(4):275–281. doi:10.1136/jmedgenet-2014-102656

35. Ran X, Li J, Shao Q et al (2015) EpilepsyGene: a genetic resource for genes and mutations related to epilepsy. Nucleic Acids Res 43(Database issue):D893–D899. doi:10.1093/nar/gku943

36. Warde-Farley D, Donaldson SL, Comes O et al (2010) The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res 38(Web Server issue):W214–W220. doi:10.1093/nar/gkq537

37. Smoot ME, Ono K, Ruscheinski J et al (2011) Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 27(3):431–432. doi:10.1093/bioinformatics/btq675

38. Krishnan A, Zhang R, Yao V et al (2016) Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder. Nat Neurosci 19(11):1454–1462. doi:10.1038/nn.4353

39. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B (Methodological):289–300

40. Guerrini R, Parrini E (2012) Epilepsy in Rett syndrome, and CDKL5- and FOXG1-gene-related encephalopathies. Epilepsia 53(12):2067–2078. doi:10.1111/j.1528-1167.2012.03656.x

41. Kato T, Morisada N, Nagase H et al (2015) Somatic mosaicism of a CDKL5 mutation identified by next-generation sequencing. Brain and Development 37(9):911–915. doi:10.1016/j.braindev.2015.03.002

42. Iossifov I, O'Roak BJ, Sanders SJ et al (2014) The contribution of de novo coding mutations to autism spectrum disorder. Nature 515(7526):216–221. doi:10.1038/nature13908

43. De Rubeis S, He X, Goldberg AP et al (2014) Synaptic, transcriptional and chromatin genes disrupted in autism. Nature 515(7526):209–215. doi:10.1038/nature13772

44. Xu LM, Li JR, Huang Y et al (2012) AutismKB: an evidence-based knowledgebase of autism genetics. Nucleic Acids Res 40(Database issue):D1016–D1022. doi:10.1093/nar/gkr1145

45. Rauch A, Wieczorek D, Graf E et al (2012) Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet 380(9854):1674–1682. doi:10.1016/S0140-6736(12)61480-9

46. Li J, Cai T, Jiang Y et al (2016) Genes with de novo mutations are shared by four neuropsychiatric disorders discovered from NPdenovo database. Mol Psychiatry 21(2):290–297. doi:10.1038/mp.2015.40

47. Bartnik M, Derwinska K, Gos M et al (2011) Early-onset seizures due to mosaic exonic deletions of CDKL5 in a male and two females. Genet Med 13(5):447–452. doi:10.1097/GIM.0b013e31820605f5

48. Yanagisawa Y, Ito E, Yuasa Y et al (2002) The human DNA methyltransferases DNMT3A and DNMT3B have two types of promoters with different CpG contents. Biochim Biophys Acta 1577(3):457–465

49. Lee BH, Smith T, Paciorkowski AR (2015) Autism spectrum disorder and epilepsy: disorders with a shared biology. Epilepsy Behav 47:191–201. doi:10.1016/j.yebeh.2015.03.017

50. Tatton-Brown K, Seal S, Ruark E et al (2014) Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability. Nat Genet 46(4):385–388. doi: 10.1038/ng.2917

51. Jiang YH, Yuen RK, Jin X et al (2013) Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing. Am J Hum Genet 93(2):249–263. doi:10.1016/j.ajhg.2013.06.012

52. Gonzalez-Mantilla AJ, Moreno-De-Luca A, Ledbetter DH et al (2016) A cross-disorder method to identify novel candidate genes for developmental brain disorders. JAMA Psychiatry 73(3):275–283. doi:10.1001/jamapsychiatry.2015.2692

53. Kobow K, El-Osta A, Blumcke I (2013) The methylation hypothesis of pharmacoresistance in epilepsy. Epilepsia 54(Suppl 2):41–47. doi:10.1111/epi.12183

54. Guo X, Wang L, Li J et al (2015) Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. Nature 517(7536):640–644. doi:10.1038/nature13899

55. Kou Y, Betancur C, Xu H et al (2012) Network- and attribute-based classifiers can prioritize genes and pathways for autism spectrum disorders and intellectual disability. Am J Med Genet C Semin Med Genet 160C(2):130–142. doi:10.1002/ajmg.c.31330

56. Krivtsov AV, Armstrong SA (2007) MLL translocations, histone modifications and leukaemia stem-cell development. Nat Rev Cancer 7(11):823–833. doi:10.1038/nrc2253

57. Barabe F, Kennedy JA, Hope KJ et al (2007) Modeling the initiation and progression of human acute leukemia in mice. Science 316(5824):600–604. doi:10.1126/science.1139851