medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Identification of novel pathogenic genes of childhood epileptic encephalopathies

Yi-Wu Shi[1*], Jian-Guo Zhang[2,3*], Na He[1*], Zi-Long Ye[1], Wen-Bin Li[1], Han-Kui Liu[2,3], Zhi-Gang Liu[4], Nan-Xiang Shen[1], Xiao-Chong Qu[1], Cui-Xia Fan[1], Jie Wang[1], Sheng Luo[1], Yu-Dan Lv[5], Li Gao[6], Jing Chen[7], Shao-Ping Huang[8], Xin-Guo Lu[9], Jing Yu[10], Jie Zhang[11], Xiu-Xia Wang[12], Qu-Wen Gao[13], Li Jiang[14], Yan-Hui Chen[15], Hui Qi[6], Jing-Da Qiao[1], Li-Zhi Chen[1], Yuan-Jin Zeng[1], Xiao-Xue Yang[16], Hong-Jun Yan[17], Cheng-Yan Li[18], Tao Zeng[19], Fu-Li Min[19], Bing Qin[20], Hai-Qing Xu[21], Lin Xu[22], Bing-Mei Li[1], Yong-Hong Yi[1], Zhi-Hong Zhuo[23], Qing-Hui Guo[24], Su-Li He[25], Hong-Wei Zhang[26,27], Li-Ping Guan[2,3], Wei-Yi Deng[1], Xiao-Fan Ren[9], Dong-Fang Zou[9], Wei-Yue Gu[28], Tao Su[1], Xiao-Rong Liu[1], Yi-Bo Qu[29], Xin-Ping Yang[16], Wei-Ping Liao[1#].

1 Department of Neurology, Institute of Neuroscience, Key Laboratory of Neurogenetics and Channelopathies of Guangdong Province and the Ministry of Education of China, The Second Affiliated Hospital, Guangzhou Medical University, Guangzhou, China, 510260

2 BGI Genomics, Shenzhen 518083, China

3 Clin Lab, BGI Genomics, Shijiazhuang 050011, China

4 Affiliated Foshan Maternity and Child Healthcare Hospital, Southern Medical University, Foshan, China

5 The First Hospital of Jilin University, Jilin, China

6 Department of Pediatrics, Henan Province People's Hospital, People's Hospital of

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Zhengzhou University, Zhengzhou, China

7 Department of Neurology, Children's Hospital of Nanjing Medical University, Nanjing, China

8 The Second Affiliated Hospital of Xi'an Jiaotong University, Xian, China

9 Shenzhen Children's Hospital, Shenzhen, China

10 Neurology Department, Children's Hospital of Xinjiang Uygur Autonomous Region, Urumchi, China

11 Hunan Provincial Children's Hospital, Changsha, China

12 The Second Hospital of Hebei Medical University, Shijiazhuang, China

13 Department of Neurology, Foresea Life Insurance Shaoguan Hospital, Shaoguan, China

14 Children's Hospital of Chongqing Medical University, Chongqing, China

15 Department of Pediatrics, Fujian Medical University Union Hospital, Fuzhou, China

16 Center for Genetics and Developmental Systems Biology, Nanfang Hospital, Southern Medical University, Guangzhou, China

17 Guangdong 999 Brain Hospital, Guangzhou, China

18 Affiliated Hospital of Guangdong Medical University, Zhanjiang, China

19 Guangzhou First People's Hospital, Guangzhou, China

20 The First Affiliated Hospital of Jinan University, Guangzhou, China

21 Xuzhou Center Hospital, Xuzhou, China

22 Qingdao Women and Children's Hospital, Qingdao, China

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

23 The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China

24 The Second Hospital of Shandong University, Jinan, China

25 Shantou Chaonan Minsheng Hospital, Shantou, China

26 Children's Hospital Affiliated to Shandong University, Jinan, China

27 Jinan Children's Hospital, Jinan, China

28 Chigene (Beijing) Translational Medical Research Center Co., Beijing, China

29 Key Laboratory of CNS Regeneration (Ministry of Education), Guangdong-Hong Kong-Macau Institute of CNS Regeneration, Jinan University, Guangzhou, China

* These authors contributed equally to the manuscript.

Correspondence to Dr. Liao, Department of Neurology, Institute of Neuroscience, Key Laboratory of Neurogenetics and Channelopathies of Guangdong Province and the Ministry of Education of China, The Second Affiliated Hospital, Guangzhou Medical University, Guangzhou, China, E-mail wpliao@163.net; liao@gzneurosci.com.

Abstract

Background: Epileptic encephalopathy is a devastating epilepsy with etiologies largely elusive, despite whole-gene/exon sequencing of large cohorts. This study targeted the genetic causes of childhood epileptic encephalopathy, typically Lennox-Gastaut syndrome (LGS) featured by age-dependent onset and characteristic clinical manifestations.

Methods: Trio-based whole-exome sequencing was performed in 235 LGS cases with

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

individualized analyses on each trio by explainable inheritance origin with stratified frequency filtration and on each gene in four aspects, and specified statistical analyses including that on compound heterozygous variants with controls of 1942 asymptomatic parents. Animal models were used to validate the roles of novel candidate genes.

Results: We identified three novel causative genes, including *SBF1* with *de novo*, *CELSR2* with recessive, and *TENM1* with X-linked recessive variants. Significantly higher excesses of *de novo SBF1* variants and biallelic *CELSR2* variants, aggregated variant frequencies of *SBF1*, *CELSR2*, and *TENM1*, and frequency of compound heterozygous *CELSR2* variants in the cases were detected. Phenotype severity/outcome was correlated with the genotype of the variants in these genes. In *Drosophila*, knockdown of these genes showed increased seizure-like behavior and increased firing of excitatory neurons. *Sbf1* knockout zebrafish showed seizure-like behavior, premature death, and increased firing of neurons. *Celsr2* knockout mice showed spontaneous seizures with epileptiform discharges. Additional 42 genes were identified as novel candidate pathogenic genes with evidence of the four genetic aspects/statistics.

Conclusions: This study suggests *SBF1*, *CELSR2*, and *TENM1* are pathogenic genes of LGS and highlights the implications of phenotype subclassification and individualized analyses protocol in identifying genetic causes of human diseases.

Keywords

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Lennox-Gastaut syndrome, childhood epileptic encephalopathies, *SBF1*, *CELSR2*, *TENM1*, trio-based whole-exome sequencing

Introduction

Epileptic encephalopathies, also referred to as developmental and/or epileptic encephalopathies (DEEs)[1], represent a clinically and genetically heterogeneous group of devastating epilepsies characterized by refractory seizures, severe electroencephalography (EEG) abnormalities, and neurodevelopmental delay or decline. Epileptic encephalopathy is a common clinical entity and accounts for approximately 40% of newly diagnosed epilepsies before the age of three years[2]. According to age of onset, epileptic encephalopathies can generally be divided into two major groups[1]: early infantile epileptic encephalopathy (EIEE) with onset in neonates or infants, such as West syndrome and Dravet syndrome, and childhood-onset epileptic encephalopathy, typically Lennox-Gastaut syndrome (LGS). LGS is estimated to account for 1%-2% of all patients with epilepsy and up to 10% of patients with childhood epilepsy[3-5]. It usually begins between 1 and 8 years old (before age 18 years) and is characterized by multiple seizure types that must include tonic seizures and/or consistent paroxysmal fast activity/polyspike on EEG[5]. The etiologies of epileptic encephalopathies remain largely elusive. Recent studies have demonstrated an increased number of genes associated with epileptic encephalopathies. To date, more than 100 genes have been reported to be associated with EIEE/DEEs (https://omim.org), explaining 22% to 40% of cases with EIEE[6,7]. In

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

contrast, only a few genes, such as *CHD2*, *DNM1*, *GABRB3*, and *SYNGAP1*, have been identified in patients with LGS, accounting for a small proportion of cases[8-11]. Furthermore, recent studies with large cohorts have encountered challenges in discovering novel EIEE/DEE genes[12-14]. However, the age-dependent feature, specific clinical characteristics, and distinct EEG pattern highly suggest an underlying genetic mechanism in the epileptogenesis of LGS, which remains to be determined.

In the present study, we performed trio-based whole-exome sequencing (WES) in a cohort of patients with LGS without acquired etiologies. The potential pathogenic variants were screened by an individualized analyses protocol, including individualized analyses on each trio by explainable inheritance origin with stratified frequency filtration and on each gene from four aspects that include gene expression in the brain, previously reported phenotypes, probability of being intolerant to heterozygous/homozygous variants of loss-of-function (pLI/pRec), and phenotypes produced by knockout/knockdown. Specific statistical analyses were used for variants of different inheritance, including analysis of compound heterozygous variants with the establishment of a control database. *SBF1*, *CELSR2*, and *TENM1* were identified as pathogenic genes of LGS, which were validated by experiments. Additional 42 genes were identified as novel candidate pathogenic genes with evidence of statistics or the four aspects of gene profile. This study highlights the implications of phenotype subclassification and individualized analyses protocol, which are suggestive for identifying novel pathogenic genes in future studies.

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Methods

Subjects

A total of 235 unrelated cases (234 trios and 1 singleton of adopted) of LGS without acquired causes were enrolled from 2005 to 2021 (see the Supplementary Appendix and Table S1). All patients were diagnosed with LGS with frequent seizures characterized by two or more of the following features: (1) multiple seizure types including tonic seizures, (2) generalized polyspikes or fast rhythms during sleep (especially required when daily tonic seizures are obscure), and (3) generalized slow (< 2.5 Hz) spike-wave complex on EEG. The patients were followed up for at least one year. All subjects were Han Chinese with four Han Chinese grandparents and were born to nonconsanguineous Chinese parents.

This study adhered to the guidelines of the International Committee of Medical Journal Editors regarding patient consent for research or participation and received approval from the ethics committee of the Second Affiliated Hospital of Guangzhou Medical University. Written informed consent was obtained from the patients or their legal guardians.

WES and variant evaluation

The details of WES are presented in the Supplementary Appendix. To identify candidate causative variants in each trio, we adopted an individualized analytical framework in the following five steps (Fig. S1).

1. General filtration. The potentially pathogenic variants, including nonsense,

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

frameshift, canonical splice site, initiation codon, in-frame insert/deletion, missense, and synonymous variants predicted to impact splicing, were retained and then filtered with minor allele frequency (MAF) < 0.005 in the Genome Aggregation Database (gnomAD).

2. Variant filtration by explainable inheritance origin in each trio. The origin of variants presents the genetic difference between the affected child and the parents and thus explains the occurrence of phenotype in a given family (trio), i.e., *de novo*, recessive from the two asymptomatic parents each, or hemizygous. Variants with an explainable inheritance origin in each trio were selected.

3. The variants were then filtered with stratified MAF criteria. The MAF of *de novo* variants, hemizygotes, and homozygotes is set as absent in the control populations in gnomAD, and for compound heterozygous variants, the product of multiplying the frequencies of two alleles in gnomAD is $< 1\times10^{-6}$, which is 7 times less than the probability of one individual in the current population of gnomAD ($1/141456 = 7 \times 10^{-6}$).

4. Variant filtration with criteria on the gene profile. The genes with qualified variants were classified into epilepsy-associated genes (977 genes by Wang *et al.*[15] and updated from OMIM) and genes with undefined gene-disease/epilepsy/seizures associations. For variants in epilepsy-associated genes, their pathogenicities were evaluated by the American College of Medical Genetics and Genomics Standards and Guidelines (ACMG)[16]. For those with undefined gene-disease/epilepsy/seizure associations, the following four aspects were assessed: (1) Tissue-specific expression:

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

a candidate pathogenic gene of epilepsy should be expressed in the brain (inclusion criterion), with careful consideration of the possibility of other explainable pathogenic mechanisms, such as ectopic expression and remote toxic effects of abnormal metabolic products. (2) Excluding the disease-causing genes with previously defined gene-disease associations (genotype-phenotype correlation) against the possibility of epilepsy as a phenotype (exclusion criterion). (3) The probability of being intolerant to heterozygous/homozygous variants of loss-of-function (pLI/pRec), genes of pLI $\geq$ 0.9 with *de novo* variants, and genes of pLI $\geq$ 0.9/pRec $\geq$ 0.9/pNull $\leq$ 0.1 with recessive variants were considered. (4) Whether the genetic knockout/knockdown produces phenotypes of the brain.

5. Selection of candidate genes. The genes with variants of statistically significant repetitiveness, including excesses of *de novo*[8] or recessive[17] variants, aggregate frequency of the variants[18], and the frequency of compound heterozygous variants compared with that in an asymptomatic parent control (see the Supplementary Appendix), or genes meeting the four criteria on the gene profile were selected as novel candidate genes. To analyze the significance of the compound heterozygous variants, we established a control cohort that included 1942 asymptomatic parents from trios, in whom the compound heterozygous variants were identified by detecting one of the paired variants in the child, given that one of the paired variants in a parent would transmit to the child.

All candidate variants were validated by Sanger sequencing. Qualified novel candidate genes were subjected to further bioinformatics and experimental studies to

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

define the gene-disease associations.

Functional studies on *SBF1*, *CELSR2*, and *TENM1*

The use of animals and all experimental procedures were approved by the Committee for Animal Care of the Second Affiliated Hospital of Guangzhou Medical University. (Detailed in the Supplementary Appendix)

Statistical analyses

The two-tailed Fisher's exact test was used to compare the allele frequencies between the case and control groups. Student's t test and one-way analyses of variance (ANOVA) with Tukey's post hoc test were used to compare differences between two or more groups. The Mann□Whitney test and Kruskal□Wallis test were used for nonnormally distributed data to compare differences among two and multiple groups, respectively. Kaplan□Meier survival curves were compared using log-rank statistics. All testing was conducted analyzed by using GraphPad Prism 8.2 (La Jolla, CA, USA). *P* values less than 0.05 were considered statistically significant. Bonferroni correction for *P* values was applied for analysis of the excesses of variants. Given the general number of encoding genes in the human genome (19,711), a corrected *P* value (*Pc*) less than 2.54×10^{-6} (0.05/19,711) indicated a significant genome-wide difference.

Results

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Variants from WES

The 235 recruited patients included 164 boys and 71 girls, with an average age of onset of 4.6 years old (4.6 ± 3.8, mean ± SD). General information on WES is presented in the Supplementary Appendix. After variant filtration by inheritance origin and stratified MAF (steps 2 and 3 in Fig. S1), we obtained 1112 qualified variants, including 177 single *de novo* (0.8 per case), 77 homozygous (0.3 per case), 725 compound heterozygous (3.1 per case), and 133 hemizygous variants (0.8 per male) (Table S4 and Fig. S4).

Variants in epilepsy-associated genes were identified in 53 cases with 31 genes involved (Table 1 and Table S5). Among these variants, 47 single *de novo* variants in 26 genes were evaluated as pathogenic or likely pathogenic by using ACMG standards, explaining 20.0% of the cases (47/235) in this cohort (Table 1). Variants in *CHD2* and *SETD1B* showed significant excesses of *de novo* variants[8] (Table S6), explaining 5.9% of the cases (14/235).

For variants in genes that have not been previously reported to be associated with epilepsy/seizures, further filtration of the gene profile was performed (step 4 in Fig. S1). By passing the criteria of inclusion and exclusion, 727 genes were included; among which 45 genes were selected as candidate genes (step 5 in Fig. S1, Table 1, and Table S7), including 14 genes with variants of statistically significant repetitiveness (Table 1, Table S6, Table S8, and Fig S5) and 31 additional genes that met the four criteria on the gene profile (Table 1 and Table S7).

SBF1 presented a significant excess of *de novo* variants[8], which was significant

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

after Bonferroni corrections (*Pc* = 0.012, Table 1 and Table S6). The aggregate frequencies of the variants in the cases were significantly higher than those in the controls (Table S8).

CELSR2 biallelic variants were significantly more than expected by chance after Bonferroni correction[17] (1% MAF cutoff for biallelic damaging missense genotypes in *CELSR2*, 8 versus 0.0029, *Pc* = 0.011). The frequency of biallelic variants in the cases was significantly higher than that in the asymptomatic parent controls (Fig. S5 and Table S9). Additionally, the aggregate frequencies of the variants in the cases were significantly higher than those in the controls (Table S8).

TENM1 hemizygous variants were identified in 6 unrelated cases (5 variants, Table 1), which was significantly more frequent than the male controls (Table S8).

SBF1, *CELSR2*, and *TENM1* variants appeared in four or more cases, involved 18 cases in total, and were subjected to further functional validations.

Genetic-clinical features of the patients with variants in *SBF1*, *CELSR2*, and *TENM1*

The DNA sequencing chromatograms and molecular consequences of the variants are shown in the Supplementary Appendix and Fig. S6-8. The clinical features of the patients are summarized in Table S10.

SBF1

Four missense variants in *SBF1* were identified in four unrelated cases, including three *de novo* variants (c.337C>A/p.Gln113Lys, c.4459G>A/p.Gly1487Ser, and

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

c.5424G>C/p.Trp1808Cys) and one with unknown origin (adopted, c.2272A>G/p.Ile758Val) (Fig. 1A and Fig. S6). The three variants located in the functional domains were identified in patients (LG90, LG57, and LG241) with intractable and frequent seizures even under a combination of three or four antiepileptic drugs (AEDs). They presented both generalized and multifocal discharges in EEGs. The patient (LG206) with the variant in the region between the DENN and GRAM domains (p.Ile758Val) presented a relatively late age of onset (10-15 years) with only generalized discharges in EEG. He achieved seizure-free with monotherapy of lamotrigine, despite previous daily myoclonic seizures.

CELSR2

Biallelic variants in *CELSR2* were identified in eight unrelated cases, including a homozygous missense variant (c.7227C>A/p.His2409Gln) in two cases (Fig. 1B and Fig. S7). The variant p.Ser1214Cys was also recurrently identified in two cases. The two cases (LG38 and LG231) with homozygous variant p.His2409Gln exhibited intractable seizures. In contrast, the two cases (LG242 and LG40) with compound heterozygous variants that had both variants located at nonfunctional regions (p.Ser1214Cys/p.Phe2036Leu and p.Thr2105Ser/p.Arg2875Gln) achieved seizure-free with monotherapy of valproate/lamotrigine. Among the other four cases with one of the paired variants located in the functional domains, two cases (LG150 and LG238, who had variants p.Thr1110Ile and p.Ser2329Asn in functional domains, respectively) presented refractory seizures; one case (LG230, with p.Pro1471Thr in the laminin G-like domain) achieved seizure-free but had intellectual disability; and

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

the other case (LG54) achieved seizure-free without intellectual disability and had compound heterozygous variants with two variants located furthest apart (p.Pro588Arg and p.Glu2849Lys, a distance of 2261 amino acids), which have been suggested to be associated with phenotype severity[19,20].

TENM1

Five hemizygous missense variants in *TENM1* were identified in six unrelated cases (Fig. 1C and Fig. S8). All *TENM1* hemizygous variants were inherited from their asymptomatic mothers, consistent with an XLR inheritance pattern. Among the three cases with variants in the N-terminal intracellular teneurin domain, two cases (LG239 and LG60) suffered from refractory seizures even under three AEDs. In contrast, the three cases (LG28, LG189, and LG270) with variants located in the nonfunctional regions achieved seizure-free under the combination therapy of valproate and lamotrigine (Table S10).

Seizures and electrophysiological activities in *Sbf*, *Flamingo*, and *Ten-m* knockdown *Drosophila*

The *Drosophila* homologs of the human *SBF1*, *CELSR2*, and *TENM1* genes are *Sbf*, *Flamingo*, and *Ten-m*, respectively. We established gene knockdown models in *Drosophila* by RNA interference (RNAi) (see the Supplementary Appendix and Fig. S2) and examined bang-sensitivity (BS) seizure-like behavior. Three phases of typical seizure-like behavior in the BS test, including seizure, paralysis, and recovery, were observed in the knockdown flies (Fig. 2A). Seizure-like behaviors were detected in

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

40.2% (*Sbf*), 24.8% (*Flamingo*), and 33.0% (*Ten-m*) of the knockdown flies, which was 8, 3, and 4 times higher than the control flies, respectively (Fig. 2B). The gene knockdown flies also showed a longer duration of seizure-like behavior than the controls. The knockdown flies recovered from seizures within 30 (*Sbf*), 7 (*Flamingo*), and 20 (*Ten-m*) seconds, while the control flies recovered within 3, 5, and 5 seconds, respectively (Fig. 2C).

The *Sbf*, *Flamingo*, and *Ten-m* knockdown flies showed regular bursts of firing in the projection neurons (Fig. 2D). The frequencies of spontaneous excitatory postsynaptic potentials (sEPSPs) in knockdown flies were significantly higher than those in WT flies (Fig. 2E). There was no significant difference in the amplitude of sEPSPs between the knockdown flies and WT flies (Fig. 2F). These results suggested that knockdown of *Sbf*, *Flamingo*, and *Ten-m* resulted in significantly increased seizure susceptibility and epileptic electrophysiological activity in *Drosophila*.

Seizures and electrophysiological activities in *sbf1* knockout zebrafish

We generated a *sbf1* knockout model by using CRISPR/Cas9 technology (see the Supplementary Appendix and Fig. S3). At 5 days postfertilization (d.p.f.), morphological abnormalities, including lack of a swim bladder, severe body curvature, and small body size, were observed in 1.1% of $sbf1^{-/-}$ and 0.9% of $sbf1^{+/-}$ larvae (Fig. 3A, B). At 17 d.p.f., all $sbf1^{-/-}$ zebrafish were dead, while $sbf1^{+/-}$ and $sbf1^{+/+}$ siblings remained alive. Approximately 62.5% of $sbf1^{+/-}$ and 83.3% of $sbf1^{+/+}$ zebrafish survived to 30 d.p.f., suggesting an essential role of *sbf1* in survival (Fig. 3C).

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Seizure-like behavior characterized by high-speed (> 20 mm/sec) "whirlpool-like" circling swimming was observed in the *sbf1*$^{+/-}$ and *sbf1*$^{-/-}$ larvae at 7 d.p.f. (Fig. 3D). The *sbf1*$^{+/-}$ larvae exhibited more hyperactive/seizure-like behavior with longer distance and duration of high-speed movement than the *sbf1*$^{+/+}$ larvae (Fig. 3E and 3F). It is noted that the *sbf1*$^{-/-}$ larvae exhibited more hyperactive/seizure-like behavior than *sbf1*$^{+/+}$ larvae, but not as apparent as the *sbf1*$^{+/-}$ larvae (only a significantly longer distance of high-speed movement at the 60 min observation; Fig. 3E).

We further performed electrophysiological recordings in the forebrain of larvae at 7-10 d.p.f. as reported previously[21]. Spontaneous epileptiform discharges were recorded in *sbf1*$^{-/-}$ and *sbf1*$^{+/-}$ larvae but not in *sbf1*$^{+/+}$ larvae (Fig. 3G). Both the *sbf1*$^{-/-}$ and *sbf1*$^{+/-}$ larvae exhibited a greater number and duration of discharges than the *sbf1*$^{+/+}$ larvae, which was also more apparent in the *sbf1*$^{+/-}$ larvae (Fig. 3H, I).

Seizures and electrophysiological activities in *Celsr2* knockout mice

Spontaneous seizures with typical EEG alterations were observed in homozygous *Celsr2* knockout mice (4/15, 26.7%, Fig. 4A-C, and Supplementary Video). Among the four *Celsr2*$^{-/-}$ mice with spontaneous seizures, two mice exhibited seizures to stage 3. The other two mice displayed seizures up to stage 5, followed by death (Fig. 4D). Analysis of the power spectra of the background EEGs showed that the *Celsr2*$^{-/-}$ mice had higher spectral power, characterized by high-amplitude and low-frequency oscillations (Fig. 4E). The average duration of seizure activity in the *Celsr2*$^{-/-}$ mice was 194.9 ± 168.0 seconds (mean ± SD, Fig. 4F).

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Discussion

Despite the wide application of high-throughput sequencing, the gene-disease associations in more than three-fourths of the genes in the human genome remain undetermined (https://omim.org). Recent studies in large cohorts of epilepsy patients have encountered challenges in discovering novel epilepsy genes[12-14]. Using trio-based WES with individualized analyses, the present study identified three novel pathogenic genes of LGS, including *SBF1* with *de novo* origin, *CELSR2* with biallelic recessive inheritance, and *TENM1* with X-linked recessive inheritance. These genes presented significant repetitiveness of variants in LGS, including significantly higher excesses of *de novo* variants in *SBF1* and biallelic variants in *CELSR2*, and aggregated frequencies of variants in *SBF1*, *CELSR2*, and *TENM1* compared with that in the controls of the gnomAD populations. The frequency of compound heterozygous/homozygous *CELSR2* variants in the cases was significantly higher than that in the asymptomatic parent controls. Clinically, the phenotype severity and outcome were correlated with genotype. Animal models with knockdown or knockout of these genes exhibited increased seizure-like behavior and increased firing of excitatory neurons. Additional 42 genes were identified as potential candidate genes with evidence of statistics or the four aspects of gene profile. The identified potential causative genes and the disclosed genotype-phenotype correlations imply a significance in the diagnosis and management of the patients. This study highlights the implications of phenotype subclassification and individualized analyses protocol,

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

which are potentially helpful in identifying novel pathogenic genes among the three-fourths of the human genome, providing the basis for future individualized medicine.

Currently, statistical analyses are commonly used to identify pathogenic genes in studies with large cohorts, which is limited in identifying novel pathogenic genes and some classical pathogenic genes, such as *KCNT1*[12,13] and *TSC2*[12-14], would be missed in such studies. In fact, only *CHD2* and *SETD1B* presented significant excess variants in this study, among the 31 defined epilepsy-associated genes. This study focused on LGS and employed an individualized analysis procedure in combination with specific statistical analyses, which potentially improves the specificity and sensitivity in identifying novel pathogenic genes.

First, epilepsy syndromes are characterized by age-dependent clinical manifestations, especially the age of onset[1], which are potentially associated with distinct pathogenic genetic causes[15]. The present study focused on epileptic encephalopathy of childhood onset with characteristic clinical manifestations, differing from most of the DEEs that are commonly of early onset, and identified three novel potential pathogenic genes, thus reinforcing the implication of phenotype subclassification by fining clinical features.

Second, individualized analyses on each trio included criteria of explainable inheritance origin and subsequently stratified MAF. The explainable inheritance origin is essential for the selection of variants, which potentially improves the specificity. The variants were then filtered with MAF according to the inheritance

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

pattern, instead of simply "absence" in the control populations, thus avoiding omission of a single low-MAF variant in biallelic variants and potentially improving the sensitivity. In fact, a single heterozygous variant of "damaging" in recessive disorders is not "pathogenic" and thus may be prevalent with low MAF in the general populations, typically p. Arg208Ter (MAF of 0.00025) and c.509-1G>C (MAF of 0.00039) in *TPP1*, which were identified in 89% of patients with ceroid lipofuscinosis[22], a severe neurodegenerative disease with epilepsy. Pathogenic biallelic variants are potentially common. This could be basically understood by the quantitative dependent feature of genes[23,24]. Since the human genome is diploid, and usually, a portion of functioning from one copy of a gene is sufficient for biophysiological function. Subsequently, biallelic variants are required for pathogenicity in most of the genes. This study used specific statistical analyses for variants of different inheritance, including the Poisson cumulative distribution function for *de novo* variant[8] and the cumulative binomial probability for recessive variants[17]. The establishment of a control cohort of compound heterozygous variants in this study enabled the statistical analysis of the significance of biallelic variants.

The individualized analyses on each gene were from four aspects, including tissue-specific gene expression, previously reported phenotypes, pLI/pRec, and phenotypes produced by knockout/knockdown, which potentially improves specificity and helps targeting and defining novel pathogenic genes. The *SBF1*, *CELSR2*, and *TENM1* genes are highly expressed in the brain. The pLIs of the three genes are 1.0, indicating that loss of function of the three genes are potentially pathogenic.

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Knockout of *SBF1*, *CELSR2*, and *TENM1* exhibited postnatal lethality or brain abnormalities in previous studies (https://www.informatics.jax.org/), and knockdown or knockout models in the present study exhibited increased seizure-like behavior and increased firing of excitatory neurons.

Regarding previously reported phenotypes, *CELSR2* and *TENM1* have not been defined to be associated with human diseases (https://omim.org). Our previous studies have shown that two genes of the CELSR subfamily, *CELSR1* and *CELSR3*, are associated with epilepsy[25,26]. *SBF1* encodes a member of the protein-tyrosine phosphatase family, which plays a vital role in cell growth and differentiation[27]. Previously, *SBF1* biallelic recessive variants (Fig. S9) were reported to be associated with Charcot-Marie-Tooth (CMT) type 4B3, which is characterized by progressive limb muscle weakness and distal sensory impairment[28]. In contrast, *SBF1* variants identified in this study are *de novo* heterozygous variants. *SBF1* is expressed in the brain and muscle at similar levels. It is noted that the heterozygous *Sbf1* knockout zebrafish presented more hyperactive/seizure-like behavior than the homozygous knockout larvae. Additionally, the pLI of *SBF1* is 1.00, indicating that *SBF1* is highly intolerant to heterozygous loss-of-function variants. These evidences suggest a potential association between epilepsy and heterozygous *SBF1* variants. The mechanism underlying the pathogenesis of *SBF1* biallelic variants in CMT4B3 is unclear[29]. Genes associated with both CMT and epilepsy have been previously reported, such as *AARS1*[30,31] and *DYNC1H1*[32,33], for which further studies are needed to determine the underlying mechanisms.

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Additional 42 genes (in 57 cases) presented variants with repetitiveness of statistical significance or met the four criteria of gene profile, which are highly suggestive for candidate pathogenic genes of childhood epileptic encephalopathies and warrant further validation.

This study suggested that *SBF1*, *CELSR2*, and *TENM1* were pathogenic genes of LGS and highlights the implication of phenotype subclassification and individualized analyses protocol in identifying novel pathogenic genes in future studies.

References

1. Wirrell, E.C. *et al.* Methodology for classification and definition of epilepsy syndromes with list of syndromes: Report of the ILAE Task Force on Nosology and Definitions. *Epilepsia* **63**, 1333-1348 (2022).
2. Eltze, C.M. *et al.* A population-based study of newly diagnosed epilepsy in infants. *Epilepsia* **54**, 437-45 (2013).
3. Asadi-Pooya, A.A. Lennox-Gastaut syndrome: a comprehensive review. *Neurol Sci* **39**, 403-414 (2018).
4. Jahngir, M.U., Ahmad, M.Q. & Jahangir, M. Lennox-Gastaut Syndrome: In a Nutshell. *Cureus* **10**, e3134 (2018).
5. Specchio, N. *et al.* International League Against Epilepsy classification and definition of epilepsy syndromes with onset in childhood: Position paper by the ILAE Task Force on Nosology and Definitions. *Epilepsia* **63**, 1398-1442 (2022).

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

6. Sheidley, B.R. *et al*. Genetic testing for the epilepsies: A systematic review. *Epilepsia* **63**, 375-387 (2022).

7. Ruggiero, S.M., Xian, J. & Helbig, I. The current landscape of epilepsy genetics: where are we, and where are we going? *Curr Opin Neurol* **36**, 86-94 (2023).

8. Epi, K.C. *et al. De novo* mutations in epileptic encephalopathies. *Nature* **501**, 217-21 (2013).

9. Carvill, G.L. *et al.* Targeted resequencing in epileptic encephalopathies identifies *de novo* mutations in CHD2 and SYNGAP1. *Nature Genetics* **45**, 825-830 (2013).

10. Euro, E.-R.E.S.C., Epilepsy Phenome/Genome, P. & Epi, K.C. *De novo* mutations in synaptic transmission genes including DNM1 cause epileptic encephalopathies. *Am J Hum Genet* **95**, 360-70 (2014).

11. Shi, Y.W. *et al*. Synaptic clustering differences due to different GABRB3 mutations cause variable epilepsy syndromes. *Brain* **142**, 3028-3044 (2019).

12. Epi25 Collaborative. Electronic address, s.b.u.e.a. & Epi, C. Ultra-Rare Genetic Variation in the Epilepsies: A Whole-Exome Sequencing Study of 17,606 Individuals. *Am J Hum Genet* **105**, 267-282 (2019).

13. Epi25, C., Chen, S., Neale, B.M. & Berkovic, S.F. Shared and distinct ultra-rare genetic risk for diverse epilepsies: A whole-exome sequencing study of 54,423 individuals across multiple genetic ancestries. Preprint at *medRxiv* https://doi.org/10.1101/2023.02.22.23286310 (2023).

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

14. Epi25 Collaborative. Electronic address, j.c.c.e. & Epi, C. Sub-genic intolerance, ClinVar, and the epilepsies: A whole-exome sequencing study of 29,165 individuals. *Am J Hum Genet* **108**, 965-982 (2021).

15. Wang, J. *et al*. Epilepsy-associated genes. *Seizure* **44**, 11-20 (2017).

16. Richards, 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* **17**, 405-24 (2015).

17. Martin, H.C. *et al*. Quantifying the contribution of recessive coding variation to developmental disorders. *Science* **362**, 1161-1164 (2018).

18. Strande, N.T. *et al*. Evaluating the Clinical Validity of Gene-Disease Associations: An Evidence-Based Framework Developed by the Clinical Genome Resource. *Am J Hum Genet* **100**, 895-906 (2017).

19. Luo, S. *et al*. Recessive LAMA5 Variants Associated With Partial Epilepsy and Spasms in Infancy. *Front Mol Neurosci* **15**, 825390 (2022).

20. Luo, J. *et al*. MPDZ variants associated with epilepsies and/or febrile seizures and the individualized genotype-phenotype correlation. *Genes & Diseases* https://doi.org/10.1016/j.gendis.2023.06.006 (2023).

21. Baraban, S.C. Forebrain electrophysiological recording in larval zebrafish. *J Vis Exp* https://doi.org/10.3791/50104 (2013).

22. Kousi, M., Lehesjoki, A.E. & Mole, S.E. Update of the mutation spectrum and clinical correlations of over 360 mutations in eight genes that underlie the

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

neuronal ceroid lipofuscinoses. *Hum Mutat* **33**, 42-63 (2012).

23. He, N. *et al.* Common genetic epilepsies, pathogenicity of genes/variants, and genetic dependence. *Seizure* **109**, 38-39 (2023).
24. Li, B. *et al.* Genetic Dependence and Genetic Diseases. Preprint at *bioRxiv* https://doi.org/10.1101/2023.08.02.551736 (2023).
25. Chen, Z. *et al*. CELSR1 variants are associated with partial epilepsy of childhood. *Am J Med Genet B Neuropsychiatr Genet* **189**, 247-256 (2022).
26. Li, J. *et al*. CELSR3 variants are associated with febrile seizures and epilepsy with antecedent febrile seizures. *CNS Neurosci Ther* **28**, 382-389 (2022).
27. Cui, X. *et al.* Association of SET domain and myotubularin-related proteins modulates growth control. *Nat Genet* **18**, 331-7 (1998).
28. Nakhro, K. *et al*. SET binding factor 1 (SBF1) mutation causes Charcot-Marie-Tooth disease type 4B3. *Neurology* **81**, 165-73 (2013).
29. Mammel, A.E. *et al*. Distinct roles for the Charcot-Marie-Tooth disease-causing endosomal regulators Mtmr5 and Mtmr13 in axon radial sorting and Schwann cell myelination. *Hum Mol Genet* **31**, 1216-1229 (2022).
30. Latour, P. *et al*. A major determinant for binding and aminoacylation of tRNA (Ala) in cytoplasmic Alanyl-tRNA synthetase is mutated in dominant axonal Charcot-Marie-Tooth disease. *Am J Hum Genet* **86**, 77-82 (2010).
31. Simons, C. *et al*. Loss-of-function alanyl-tRNA synthetase mutations cause an autosomal-recessive early-onset epileptic encephalopathy with persistent myelination defect. *Am J Hum Genet* **96**, 675-81 (2015).

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

32. Becker, L.L. *et al*. The clinical-phenotype continuum in DYNC1H1-related disorders—genomic profiling and proposal for a novel classification. *Journal of Human Genetics* **65**, 1003-1017 (2020).

33. Wu, W.C. *et al*. DYNC1H1 variants associated with infant-onset epilepsy without neurodevelopmental disorders. *Seizure* https://doi.org/10.1016/j.seizure.2023.10.010 (2023).

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgements

We thank the affected patients and their families for participating in this study. This work was funded by the National Natural Science Foundation of China (grant nos. 82171439, 82271505, and 81971216) and the Guangdong Basic and Applied Basic Research Foundation (grant no. 2021A1515010986). The funders had no role in the study design, data collection, and analysis or in the decision to publish or the preparation of the manuscript.

Author Contributions

WPL and YWS conceptualized the study, analyzed and interpreted the data, and drafted and revised the manuscript. WPL had full access to all data in the study and

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

takes responsibility for the integrity of the data and accuracy of the data analysis. JGZ, NH, ZLY, NXS, and HKL performed the clinical data and whole-exome sequencing analysis. WBL, XCQ, and CXF performed seizure activity studies. WPL, YWS, JGZ, NH, ZLY, NXS, HKL, WBL, XCQ, and CXF drafted and revised the manuscript and contributed to the statistical analysis. All authors collected data, revised the manuscript, and contributed to the writing.

Competing interests

All authors claim that there are no conflicts of interest.

Ethics statement

All procedures performed were in accordance with the ethical standards of the institutional committee. Ethical approval was approved by the ethics committee of the Second Affiliated Hospital of Guangzhou Medical University (approval ethics number 2020-hs-49). We declared that all patients' IDs were not identified to anyone outside the research group.

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Table 1. Overview of potential pathogenic genes in this cohort

Known epilepsy-associated genes (31 genes in 53 cases)

Genes with pathogenic or likely pathogenic variants (26 genes in 47 cases):

Epilepsy/DEE genes (10 genes in 24 cases):

CHD2 (10)[*], *SETD1B* (4)[*], *CELF2* (2), *PCDH19* (2), *GABRB3* (1), *GRIN2B* (1), *KCNB1* (1), *NSF* (1), *SCN1A* (1), *STXBP1* (1).

Developmental disorder with seizures genes (16 genes in 23 cases):

ATP1A3 (2), *CACNA1B*[†] (2), *NUS1* (2), *SRCAP* (2), *SYNGAP1* (2), *TSC1* (2), *ZMYM2* (2), *COL4A2* (1), *GRIA1* (1), *KIF1A* (1), *KMT2D* (1), *MACF1* (1), *NEXMIF* (1), *PPP2CA* (1), *SIN3A* (1), *TSC2* (1).

Genes with possible pathogenic variants[‡] (5 genes in 6 cases):

KIF4A (2), *ANK3* (1), *CNPY3* (1), *SLC7A6OS* (1), *TAF1* (1).

Novel candidate genes (45 genes in 75 cases)

Genes with *de novo* variants (7 genes in 10 cases):

SBF1 (3)[*, §], *SRRM2* (2), *AHCTF1* (1), *BSN* (1), *RAB6A* (1), *SBNO1* (1), *UBR2* (1).

Genes with homozygous/compound heterozygous variants (32 genes in 49 cases):

CELSR2 (8)[*, §, ¶], *ADGRA1* (2)[¶], *AKAP13* (2)[¶], *CRMP1* (2)[¶], *CCDC66* (2)[¶], *MICAL2* (2)[¶], *SYNJ2* (2)[¶], *TOGARAM2* (2)[¶], *FAT1* (2), *LAMA1* (2), *LRP2* (2), *LTN1* (1), *KIF26B* (1), *ARNTL* (1), *ATXN1L* (1), *BOC* (1), *CHD6* (1), *CLUH* (1), *DLX6* (1), *FASN* (1), *HEG1* (1), *JMJD1C* (1), *KDM4C* (1),

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

LMTK2 (1), *MYO18A* (1), *NCOA2* (1), *NCOA6* (1), *NCOR2* (1), *PTGFRN* (1), *RFX1* (1), *TENM2* (1), *UBE2J1* (1).

Genes with X-linked recessive variants (6 genes in 16 cases):

TENM1 (6)[§], *AMER1* (2)[§], *NRK* (2)[§], *ZNF674* (3)[§], *MAGEC1* (2)[§], *ZMYM3* (1).

[*]Genes with a significant excess of variants in this cohort compared with controls ($Pc < 0.05$ after Bonferroni correction).

[†]The two variants in the *CACNA1B* gene were *de novo*, while previously reported variants were recessive.

[‡]The variants in these genes were all recessive and evaluated as uncertain significance by ACMG but were considered to be possibly pathogenic by highly consistent clinical phenotype as previously reported, absence or very low MAF in the general population (criteria in step 3 in Fig. S1), and explainable inheritance origin (details in Table S5).

[§]The aggregate frequencies of the variants in the cases were significantly higher than those in the controls ($P < 0.05$).

[¶]The frequencies of genes with biallelic variants in this cohort were significantly higher than those in the asymptomatic parent controls ($P < 0.05$).

The underline indicates the genes that met the four criteria on the gene profile in step 4 for selection of novel candidate genes (Fig. S1).

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

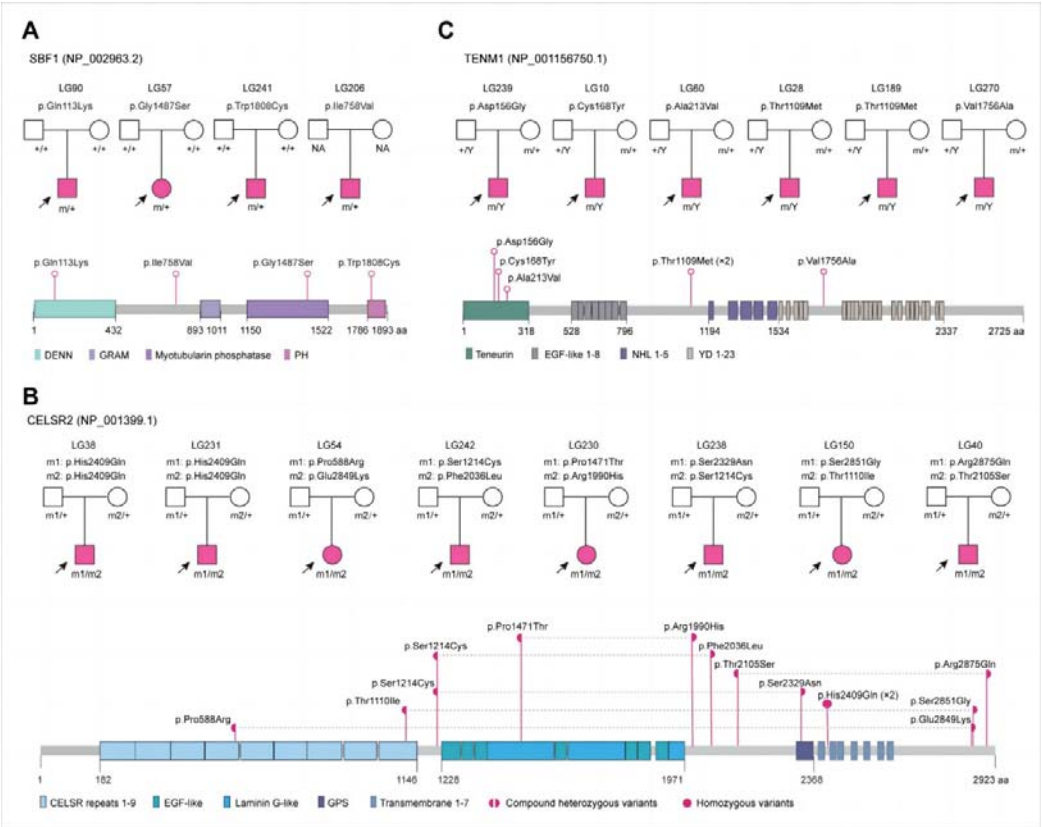

Figure 1. *SBF1*, *CELSR2*, and *TENM1* variants identified in patients with LGS. Pedigrees of the unrelated cases with *SBF1* (Panel A), *CELSR2* (Panel B), and *TENM1* (Panel C) variants (top) and the localization of the variants in the protein domains (bottom). The gray dotted lines in Panel B indicate a pair of compound heterozygous variants of *CELSR2*. Abbreviations: CELSR, cadherin EGF LAG seven-pass G-type receptor; DENN, differentially expressed in normal and neoplastic cells; EGF, epidermal growth factor; GPS, G-protein-coupled receptor proteolytic site; GRAM, glucosyltransferases, Rab-like GTPase activators and myotubularins; NHL, ncl-1, HT2A and lin-41; PH, pleckstrin homology; YD, Tyr and Asp dipeptide.

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

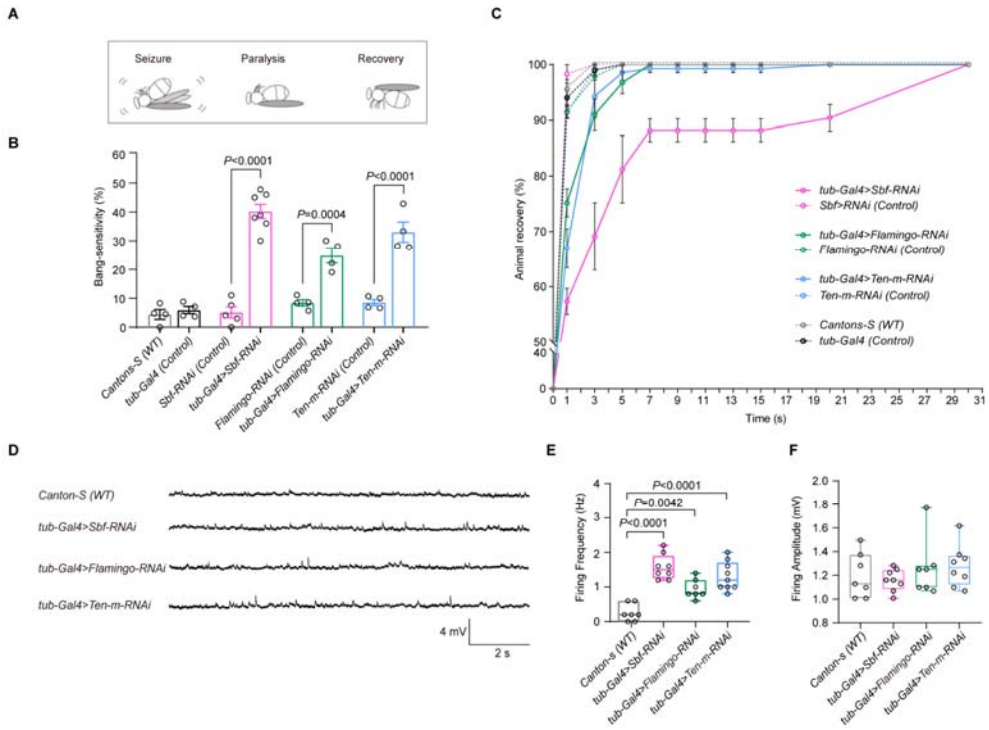

Figure 2. Seizures and electrophysiological activities in *Sbf*, *Flamingo*, and *Ten-m* knockdown *Drosophila*. Panel A shows three stages of seizure-like behaviours in the (bang-sensitivity) BS paralysis test, including seizure (manifesting as vibrating wings), paralysis, and recovery, were observed in *Sbf*, *Flamingo*, and *Ten-m* knockdown flies. Panel B shows that seizure-like behaviors occurred at a higher rate in the knockdown flies (*tub-Gal4>Sbf-/Flamingo-/Ten-m-RNAi*) than in WT flies (*Canton-S*), *tub-Gal4* (*Control*), and the respective RNAi control flies. Panel C shows the recovery time from seizures in the knockdown flies (*tub-Gal4>Sbf-/Flamingo-/Ten-m-RNAi*) and in the control groups. Data were collected from 4-7 groups, 4-6 flies in each group, for BS paralysis and recovery tests. All data are represented as the mean ± s.e.m. One-way ANOVA and Tukey's post hoc test were used for multiple comparisons.

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Panel D shows representative traces of electrical activity in projection neurons (PNs) showing spontaneous excitatory postsynaptic potentials (sEPSPs) of WT and *tub-Gal4>Sbf-/Flamingo-/Ten-m-RNAi* flies. Panel E shows the frequencies of sEPSPs in PNs of WT and knockdown flies. Panel F shows the amplitude of sEPSPs between WT and knockdown flies. The box plots show all data points from minimum to maximum. Boxes represent data from the lower (25th percentile) to the upper (75th percentile) quartiles. The box center corresponds to the 50th percentile. A horizontal line indicates the median. Data are represented as the mean ± s.e.m. (*Canton-S*, n = 7; *tub-Gal4>Sbf-RNAi*, n = 8; *tub-Gal4>Flamingo-RNAi*, n = 7, *tub-Gal4>Ten-m-RNAi*, n = 8), from three independent experiments; one-way ANOVA test.

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

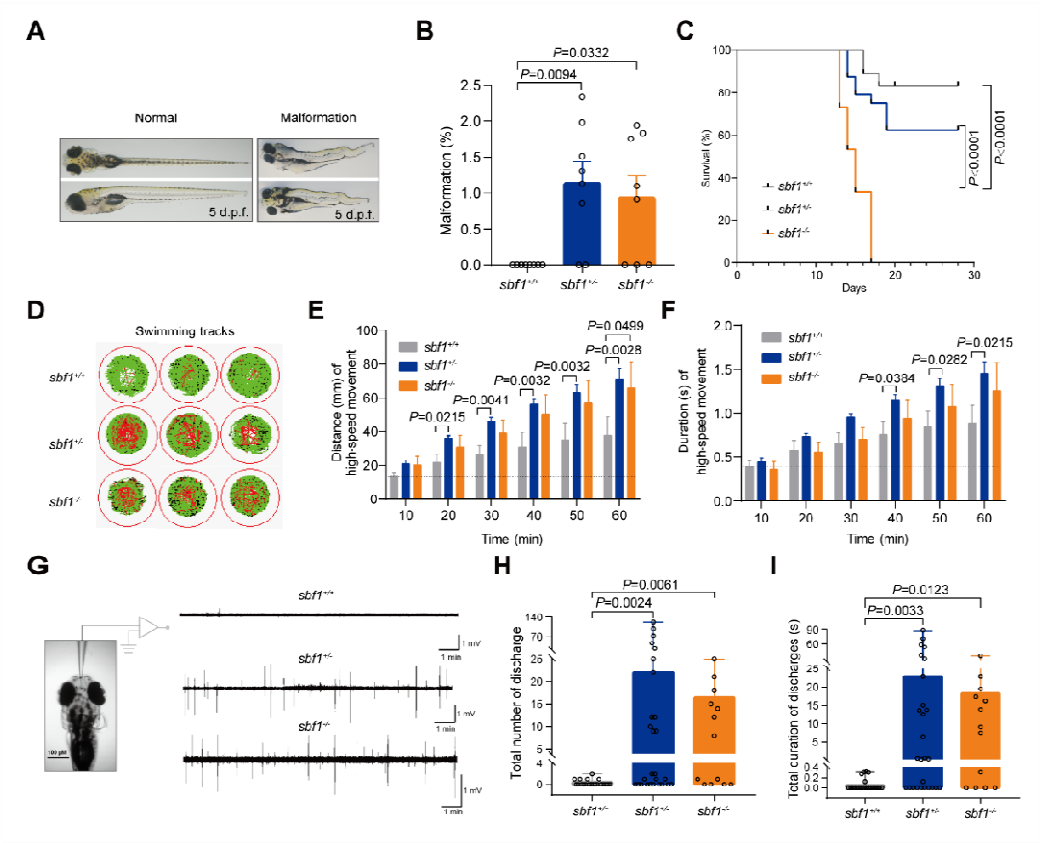

Figure 3. Seizure activity in *sbf1* knockout zebrafish. Panel A shows malformation in 5 days post-fertilization (d.p.f.) larvae of *sbf1*$^{+/-}$ and *sbf1*$^{-/-}$ but not *sbf1*$^{+/+}$ (WT). Panel B shows that malformation rates in the *sbf1*$^{+/-}$ and *sbf1*$^{-/-}$ larvae were significantly higher than those in *sbf1*$^{+/+}$ larvae (n = 8, one-way ANOVA); *sbf1*$^{+/+}$ (grey), *sbf1*$^{+/-}$ (blue), and *sbf1*$^{-/-}$ (orange). Kaplan–Meier survival analysis for *sbf1*$^{+/+}$ (n = 18), *sbf1*$^{+/-}$ (n = 24), and *sbf1*$^{-/-}$ (n = 15) in Panel C. Panel D shows representative swimming tracks of 7 d.p.f. larvae within 60 minutes. Red, green, and black lines represent high- (≥ 20 mm/s), intermediate- (8 - 20 mm/s), and slow-speed movements (< 8 mm/s), respectively. Panels E and F show the distance (Panel E) and duration (Panel F) of high-speed movement in each genotype. Data are represented as the mean ± s.e.m. (*sbf1*$^{+/+}$, n = 10; *sbf1*$^{+/-}$, n = 20; *sbf1*$^{-/-}$, n = 10), from three independent

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

experiments; one-way ANOVA test. Panel G shows extracellular field recordings that were performed in immobilized and agar-embedded zebrafish larvae forebrains. Left panel, schematic showing the insert position of a glass microelectrode. Right panel, representative field recording trace of epileptiform discharge from *sbf1*$^{+/+}$ (top), *sbf1*$^{+/-}$ (middle), and sbf*1*$^{-/-}$ (bottom) larvae at 7 d.p.f. Panels H and I show the total number of electrographic seizure-like bursts (including interictal and ictal discharges) (Panel H) and the duration of the discharges (Panel I). Data were from 15 min of recording in 7-11 d.p.f. larvae of *sbf1*$^{+/+}$ (n = 22), *sbf1*$^{+/-}$ (n = 27), and *sbf1*$^{-/-}$ (n = 13); Kruskal□Wallis test.

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

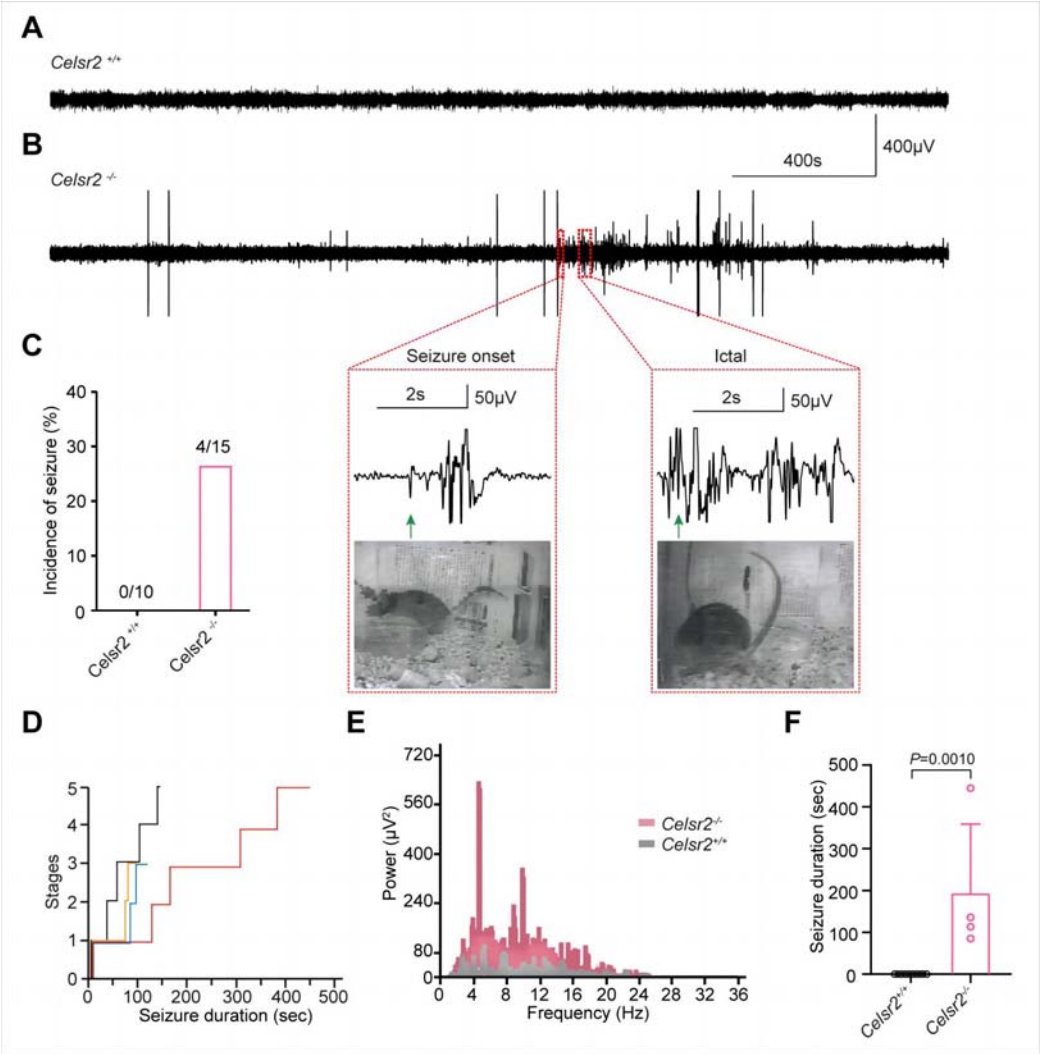

Figure 4. Spontaneous seizures and electroencephalography (EEG) recording in *Celsr2* knockout mice. Panel A shows normal EEGs that were detected from the WT mice. Panel B shows representative traces of ictal epileptiform discharges in *Celsr2*$^{-/-}$ mice. The dotted boxes indicate the zoomed-in traces of the onset and ictal epileptiform discharges. The green arrows indicate the time point of behavior of the mouse from video recording. Panel C shows the proportion of *Celsr2*$^{+/+}$ and *Celsr2*$^{-/-}$ mice exhibiting spontaneous seizures. Panel D shows seizure duration in *Celsr2*$^{+/+}$ (n = 10) and *Celsr2*$^{-/-}$ (n = 4) mice; Mann□Whitney test. Panel E shows seizure

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

behaviors that were rated with a modified Racine scale in *Celsr2*$^{+/+}$ and *Celsr2*$^{-/-}$ mice. Panel F shows the power spectrum analysis of EEG signals in *Celsr2*$^{+/+}$ and *Celsr2*$^{-/-}$ mice.

medRxiv preprint doi: https://doi.org/10.1101/2023.07.25.23293037; this version posted January 5, 2024. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .