Genetic and cellular analysis of genes and variants for juvenile myoclonic epilepsy

A thesis submitted for the degree of

Doctor of Philosophy

by

Shrilaxmi V Joshi

Molecular Biology and Genetics Unit
Jawaharlal Nehru Centre for Advanced Scientific Research
(A Deemed University)
Bengaluru 560 064, India

February 2022

Dedicated to

Appaji, Amma, Anju and Ani

Declaration

I hereby declare that this thesis, entitled **"Genetic and cellular analysis of genes and variants for juvenile myoclonic epilepsy",** is an authentic record of research work carried out by me under the guidance of Prof. Anuranjan Anand in the Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bengaluru, India and that this work has not been submitted elsewhere for the award of any other degree.

In keeping the norm of reporting scientific observations, due acknowledgements have been made wherever the work described here has been based on the findings of other investigators. Any omission, which might have occurred by oversight or misjudgment, is regretted.

Shrilaxmi V Joshi

Place: Bengaluru

Date: 24.02.2022

Certificate

This is to certify that the work described in this thesis ***"Genetic and cellular analysis of genes and variants for juvenile myoclonic epilepsy"*** is the result of the investigations carried out by Ms. Shrilaxmi V Joshi in the Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bengaluru, under my guidance. The results presented in this thesis have not previously formed the basis for the award of any other diploma, degree, or fellowship.

Anuranjan Anand

Place: Bengaluru

Date: 24.02.2022

Acknowledgements

My first words of gratitude are to my thesis supervisor Prof. Anuranjan Anand. I am thankful to him for the opportunity, resources, and encouragement. He has helped me develop lab and resource management skills. Most of all, I am grateful to him for having taught me to be a thorough scientist.

I would like to thank the faculty members of the MBGU and NSU: Prof. M.R.S. Rao, Prof. Namita Surolia, Prof. Hemalatha Balaram, Prof. Maneesha Inamdar, Prof. Tapas Kumar Kundu, Prof. Ranga Uday Kumar, Prof. Kaustav Sanyal, Prof. Sheeba Vasu, Prof. Ravi Manjithaya, Prof. James Chelliah, and Dr. Kushagra Bansal for the discussions and suggestions throughout the course of my Ph.D.

I extend my thanks to our clinical collaborators from the NIMHANS, SCTIMST and other epilepsy clinics for patient recruitment and sample collection.

This study would not have been possible without the kind collaboration of Dr. Shyam. K. Sharan, NIH, USA. I thank him for the mouse strains and inputs. I am grateful to Prof. Shubha Tole, TIFR, Mumbai for the support and the opportunity to learn *in utero* electroporation in her lab.

I am indebted to the mentoring received from Dr. Raghunand Tirumalai and Mr. A. N. Jayachandra. They have been my pillars of support and I sincerely thank them.

I wish to thank all the former and present members of human molecular genetics lab, Dr. Praveen Raju, Dr. Deepti Chugh, Dr. Nishtha Pandey, Dr. Kalpita Karan, Dr. Shalini Roy Choudhury, Dr. Ram Murthy, Dr. Manpreet Kaur, Dr. Shveta, Pooja, Sambhavi, Girija, Sukanya, Sourav, Moumita, Angshumi, Yashwini, Nazia, Mohan, Pallawi, Sri Charani, and Chandrashekhara for making my stay in the lab pleasant and for the academic environment. Special thanks to Sourav and Ram for help in the WES analysis.

The financial support of the JNCASR is gratefully acknowledged. I also

acknowledge support from the confocal, sequencing, and central instrumentation facilities at MBGU and NSU, computer lab, library, security office, purchase, academic and administrative sections. I especially thank the staff of animal house (Dr. Prakash, Sagar, Vasanth and Manjesh) and bio-imaging (Suma, Sunil and Keerthana) for their co-operation and their help in conducting my experiments successfully. My stay at JNCASR has been made comfortable by the hostel facilities that include, mess, book club, and Dhanvantri. I thank the ESHG, EMBO-EMBL, Keystone symposia for the conference fellowships.

Throughout my term in JNCASR, Vikas, Pooja and Aisha have been keepers of my sanity. I thank them for wise, helpful feedback and for all the fun times. I offer thanks to all my friends, too many to name here for cheering me on through this journey. I have had the support of my batchmates and members of the MBGU-NSU department. They have extended to me an arm of support whenever I needed it. My sincere thanks also to the faculty and friends from CCMB, Hyderabad, St. Joseph's college, and St. Francis School, Bengaluru.

I am eternally grateful to Ajji, my parents, Mr. Venkatesh D Joshi, Mrs. Geeta D Patil and my sister Ms. Anupama Joshi for the numerous sacrifices they have made, for their love and endless encouragement. They are my strength, sanity, and source of power. Special thanks to Prasanna; my niece Paavani and nephew Aprameya for being glimmering golden rays of sunshine and joy at every turn.

Aniruddha, my best friend and husband, words do no justice for everything you are to me. Thank you for your faith in me and the process, for the unwavering support and positivity.

Finally, I express profound gratitude to my parents-in-law, Dr. S. Rajagopalan and Mrs. Prabha for their love and constant encouragement.

Abbreviations

°C	Degree Celsius
µg	Microgram
µl	Microliter
ABI	Applied Biosystems
ABS	Absence Seizures
bp	Base pairs
BWA	Burrows-Wheeler Aligner
CADD	Combined Annotation Dependent Depletion
CBZ	Carbamazepine
HEK	Human Embryonic Kidney
CEPH	Centre d'E´ tude du Polymorphisme Human
cm	Centimeters
cM	Centimorgan
DAPI	4, 6-diamidino-2-phenylindole
dbSNP	Single Nucleotide Polymorphism database
DEPC	Diethyl pyrocarbonate
PBS	Phosphate Buffered Saline
DNA	Deoxyribonucleic acid
dNTP	Doxynucleotide triphosphate
EDTA	Ethylenediaminetetraacetic acid
EEG	Electroencephalogram
SD	Standard deviation
EJM	Epilepsy, Juvenile Myoclonic
EtBr	Ethidium bromide
EVS	Exome Variant Server
ExAC	Exome Aggregation Consortium

GABR1	Gamma-aminobutyric acid (GABA) A receptor, alpha-1
GABRD	Gamma-aminobutyric acid (GABA) A receptor, delta
GAIIx	illumina Genome Analyzer IIx platform
gnomAD	Genome Aggregation Database
GRch37	Genome Reference Consortium Human genome build 37
GTCS	Generalized Tonic-Clonic Seizures
HRP	Horseradish peroxidase
ILAE	International League Against Epilepsy
JME	Juvenile Myoclonic Epilepsy
kb	Kilobases
nt	Nucleotide
KCl	Potassium chloride
LOD	Logarithm of Odds
MAF	Minor Allele Frequency
Mb	Megabases
mg	Milligram
$MgCl_2$	Magnesium chloride
ml	Milliliter
mM	Millimolar
mA	Milliampere
MS	Myoclonic Seizures
NaCl	Sodium chloride
NCBI	National Centre for Biotechnology Information
NEB	New England Biolabs
ng	Nanogram
NJ	New Jersey
OMIM	Online Mendelian Inheritance in Man
PB	Phenobarbitone

PCR	Polymerase Chain Reaction
PFA	Paraformaldehyde
pmol	Picomole
PMSF	Phenylmethylsulfonyl fluoride
RT-PCR	Reverse transcriptase polymerase chain reaction
SDS	Sodium dodecyl sulfate
TAE	Tris-Acetate-EDTA
Taq	*Thermus aquaticus*
CDC20B	Cell division cycle 20B
EFHC1	EF hand domain (C-terminal)-containing 1
EFHC2	EF hand domain (C-terminal)-containing 2
CILK1	Ciliogenesis associated kinase 1
Sry	Sex-determining region Y
Tris HCl	Tris Hydrochloride
USA	United States of America
UTR	Untranslated region
V	Volt
VEP	Variant Effect Predictor
VPA	Valproic acid
WES	Whole-Exome Sequencing
WGS	Whole-Genome Sequencing

Table of Contents

Introduction

1.1 Epilepsy

Epilepsy is a common neurological disorder characterized primarily by an enduring predisposition to seizures and the associated cognitive, psychological, and social consequences (Devinsky et al., 2018). An epileptic seizure is "a transient occurrence of signs and/or symptoms due to abnormal excessive or synchronous neuronal activity in the brain" (R. S. Fisher et al., 2017). While operationally practical, conceptually, this definition of epilepsy positions it in a hinterland between a symptom and a disease. Considering that about 10% of individuals experience a seizure during their lifetime, a formal diagnosis involves further testing to identify as having (i) an epilepsy syndrome or (ii) recurrent unprovoked seizures more than 24 hours apart, or (iii) an unprovoked seizure with a high risk (≥60%) of recurrence for the next decade, where risk is determined by assessing relevant clinical characteristics, such as by conducting epileptiform electroencephalography (EEG) examination (R. S. Fisher et al., 2014; Hauser & Beghi, 2008).

An epileptic seizure is a transient behavioral change accompanied by one, or a combination of - loss of awareness, stiffening, jerking, abnormal sensation rising from abdomen to chest, a sensation of odors, and a feeling of déjà vu. These changes are rooted in the excessive synchronous neural activity in the central nervous system. Seizures manifest due to changes at many scales of spatial and temporal detail. In this regard, microscopic entities such as cells, ions, and their channels are as relevant as feedback and crosstalk across larger (1-10 cm) structures in the brain. To enable a systematic study of these

symptoms, the International League Against Epilepsy in 2017 proposed basic seizure classification. Herein, basis the origin of seizures, they can be epileptic seizures classified as focal onset, generalized onset, or unknown onset (Figure 1). Focal seizures originate within networks local to a region limited to a single hemisphere. Generalized seizures involve both the cerebral hemispheres, often beginning in neocortical, thalamocortical, limbic, or brainstem networks (R. S. Fisher et al., 2017).

Classification of seizures is challenging and often requires analysis of seizure type/s, genetic or metabolic markers, brain structure and other factors age at onset, or frequency of seizures and precipitation factors. It also requires careful and continuous monitoring which may be difficult due to the intermittent nature of seizures and the unreliability of self-reporting. EEG recordings and brain imaging are often key indicators that help classify the type of seizures.

1.2 Epilepsy classification

Foundations to the present-day concepts of epilepsy were the seminal contributions of Sir William Gowers, who described predisposing and exciting causes of epilepsy (Gowers, 1901); John Hughlings Jackson's understanding of various seizure types as well as the incorporation of a scientific approach to study epilepsies (Shorvon, 2011) and Sir John Russell Reynold's classification of epilepsy basis etiology (Reynolds, 1862; Shorvon, 2011). Amalgamating these founding ideas with current knowledge of epilepsies and their underlying mechanisms, the International League Against Epilepsy (ILAE) introduced the classification and terminology for different epilepsies and seizures and published it in 1969 (Gastaut, 1969). These were updated in the classifications of epileptic seizures in 1981, 1989, and 2010. The ILAE classification of epilepsies forms a critical tool for clinicians and researchers alike. The most recent framework for classification was

proposed in 2017 (Figure 2;Scheffer et al., 2017). This presents three levels and allows for incorporating findings relating to underlying etiology and co-morbidities. The first step involves the identification of seizure type into focal onset, generalized onset, and unknown onset. A higher-level diagnosis is made upon the availability of information relating to access to EEG, video, and imaging studies.

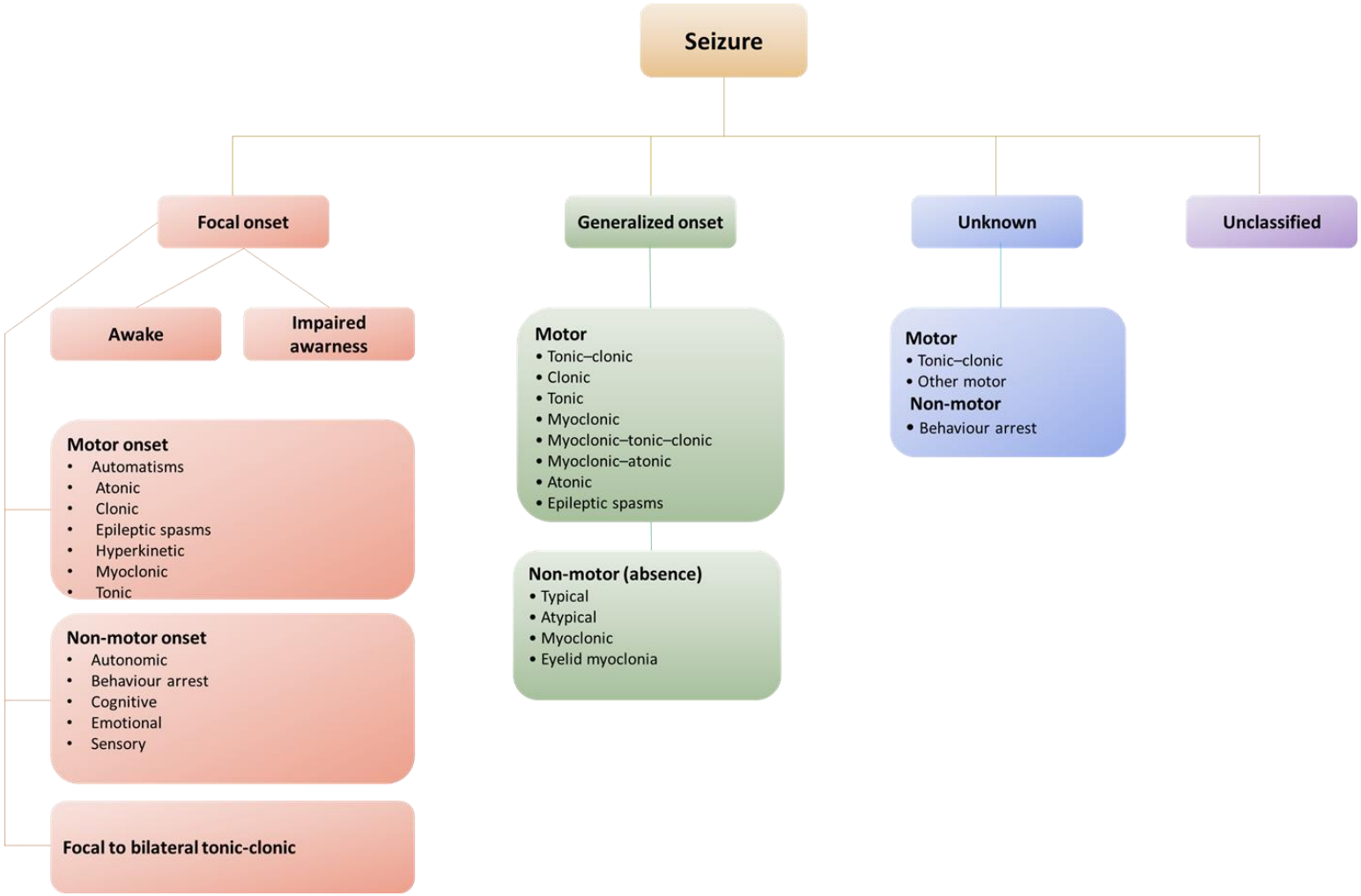


Figure 1: Classification of seizures in accordance with International League Against Epilepsy 2017. The classification is not hierarchical. Modified from (R. S. Fisher et al., 2017)

The second level identifies the epilepsy type. This includes focal, generalized, combined generalized and focal, and an unknown category. Generalized epilepsies have multiple seizure types - absence, myoclonic, atonic, tonic, and tonic-clonic seizures. These are often accompanied by generalized spike-wave activity on EEG. When presented with normal EEG, this class is identified using evidence of family history or other supportive evidence such as myoclonic jerks.

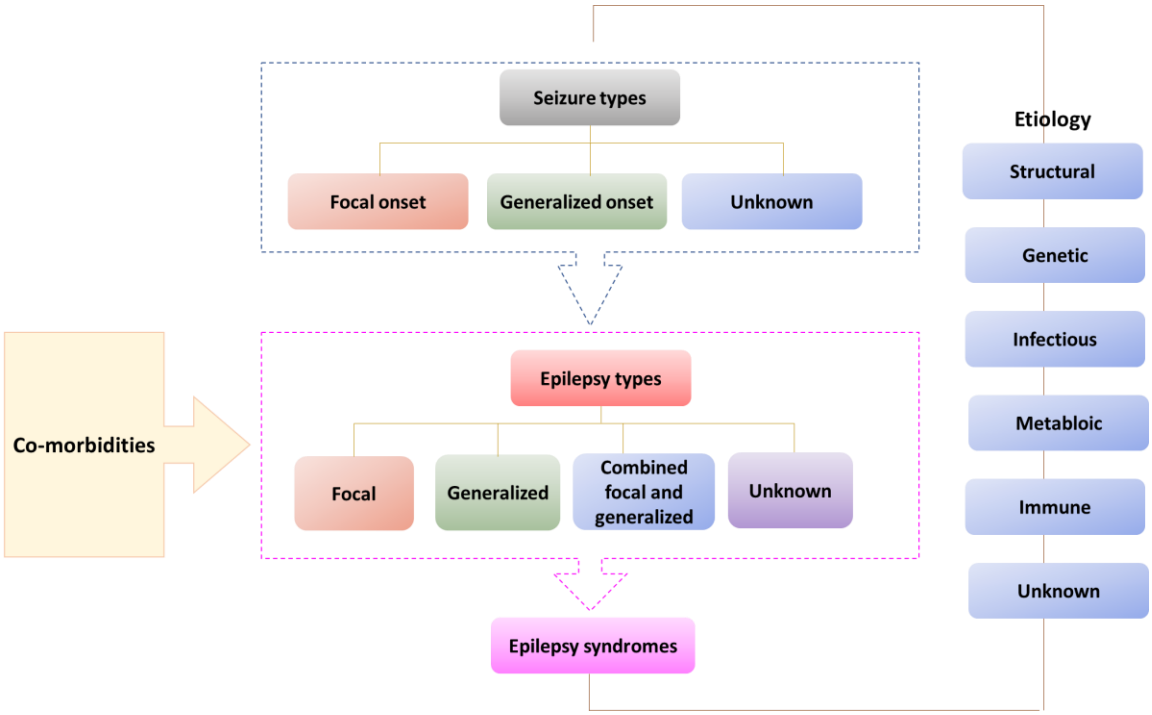


Figure-2: Classification of epilepsies. Adapted from International League Against Epilepsy 2017 classification of the epilepsies (Scheffer et al., 2017)

Focal epilepsies can be unifocal, multifocal or that of one cerebral hemisphere. The seizure types include focal aware seizures, focal impaired awareness seizures, focal motor seizures, focal non-motor seizures, and focal to bilateral tonic-clonic seizures. The phenotype presents with focal epileptiform discharges in EEG.

The combined generalized and focal epilepsies category, incorporated newly, covers patients with both focal and generalized seizures. The diagnosis is made on clinical grounds, supported by EEG findings,

including interictal EEG with generalized spike-wave and focal epileptiform discharges. Epileptiform activity is considered non-essential for a clinical diagnosis of this category.

In a few cases, a definitive diagnosis and thus a classification may be difficult even though individual symptoms may overlap with the categories described above. For all such cases wherein, the clinician cannot determine if the epilepsy type is focal or generalized because there is insufficient information available or the onset of seizures is unreported, the unknown category is used.

The third level of classification covers epilepsy syndrome diagnosis. An epilepsy syndrome is recognized by clustering seizure types, EEG, and imaging features. Over forty different types of epilepsy syndromes are known, and these vary widely with respect to onset, etiology, treatment, psychosocial sequelae, and long-term prognosis (Berg et al., 2010). These are age-dependent and sometimes have identifiable seizure triggers, diurnal variations, and prognoses.

Epilepsy is far from being a unitary construct; there are a large variety of causes, treatments, and prognostic outcomes. As indicated in the classification, aetiologies of epilepsy syndromes are attributable to genetic, structural, immune, infections, and metabolic causes, whilst others are unknown. Epilepsy is considered 'resolved' -- as opposed to being 'in remission or 'cured' --when a patient is seizure-free for ten years and off anti-epileptic drugs (AED) for five years (R. S. Fisher et al., 2014).

1.3 Genetic generalized epilepsies

Genetic generalized epilepsies (GGE) are a common but under-studied cluster of epileptic syndromes of predominantly child and adolescent-onset. It is estimated that 30-40% of all epilepsies are genetic generalized epilepsies (Jallon & Latour, 2005). They constitute a

heterogeneous group of epilepsies with generalized seizures. Electrophysiologically, there is bilateral, synchronous, and symmetrical generalized spike-wave activity on EEG (spikes, polyspike, spike-wave, and polyspike waves) with a frequency greater than 2-3Hz, typically occurring on a normal EEG background. In contrast, symptomatic generalized epilepsies show disorganized and slow backgrounds (Seneviratne et al., 2012).

The GGE syndromes differ primarily in their age of onset, predominant seizure type. The most common GGE syndromes are childhood absence epilepsy (CAE), juvenile absence epilepsy (JAE), juvenile myoclonic epilepsy (JME), and GGE with generalized tonic-clonic seizures only (GTCSO). Additionally, benign familial neonatal seizures and benign myoclonic epilepsy in infancy are rare GGE syndromes, each representing less than 1% of all childhood epilepsies (Jallon & Latour, 2005). Other less frequent types of GGE are epilepsy with myoclonic absences, epilepsy with myoclonic-astatic seizures (Doose syndrome) (Engel & International League Against, 2001). The most frequent class of GGE is JME.

Twin and family segregation studies support the notion of the genetic contribution and heritability of GGE. The rates of GGE in sibling and twin studies are illustrative of this complex genetic contribution: the risk of epilepsy by age 40 in siblings of those with GGE was 7.8 % (Flex et al., 2005). The genetic inheritance of GGE is thought to be non-Mendelian, except for an autosomal dominant form of JME. Autosomal recessive, two-locus and multifactorial models are considered possible in determining the genetic cause (Gardiner, 2005). In this thesis, various aspects of the genetic and molecular attributes of gene variants associated with GGE, specifically JME, are reported. The generation and characterization of mutant mouse lines for two potential JME genes is described.

1.4 Juvenile myoclonic epilepsy

JME was clinically, comprehensively described as a separate syndrome by Janz and Christian in a study of 47 cases from Germany, and they named it "impulsive petit mal" (Janz & Matthes, 1955), which was renamed juvenile myoclonic epilepsy in 1975 and was included as an equivalent diagnosis to impulsive petit mal and a subclass in the epilepsy classification by the ILAE in 1989. JME is the most prevalent subtype of GGE and accounts for about 5-10% (Jallon, 2005). The predominant clinical feature in JME is early morning myoclonic jerks involving the proximal upper extremities. Most patients have occasional generalized tonic-clonic seizures, and about one-third have additional absences (Genton & Gelisse, 2013). The seizures are often precipitated by sleep deprivation, disturbed sleep-wake cycle, fatigue, alcohol intake, and photic stimulation (da Silva Sousa et al., 2005; Dhanuka et al., 2001; Janz, 1985; Welty, 2006).

The current diagnostic criteria for JME were arrived at in 2011. The international experts proposed two sets of criteria – Class I and Class II. For both, patient history of myoclonic jerks predominantly occurring after awakening and an EEG with generalized epileptiform discharges indicative of idiopathic generalized epilepsy are necessary (Kasteleijn-Nolst Trenite et al., 2013).

The typical characters of class I are:

• Myoclonic jerks without loss of consciousness that occur repeatedly within two hours after awakening.

• EEG (routine, sleep, or sleep deprivation EEG) showing normal background activity and ictal generalized high amplitude poly-spikes (or poly-spike and wave) with concomitant myoclonic jerks.

• Normal intelligence.

• Age at onset between 10 and 25 years.

The diagnostic criteria for class II are more generous, and include early morning myoclonic jerks, generalized epileptiform EEG abnormalities with or without consequent myoclonic jerks, and a broader time window for age of onset (6-25 years).

1.4.1 Provocative factors in JME

Patients with JME are highly susceptible to certain triggering factors. A prospective study showed that seizures get triggered by various external factors in 93% of JME patients (Panayiotopoulos et al., 1994). Some seizure triggers in JME, such as stress, alcohol consumption, sleep deprivation, or photostimulation, were already described by Janz (Janz, 1985). Studies have also reported that the tasks combining cognitive function with a motor response are stronger provocative factors for EEG discharges in JME in about 38% of cases (Matsuoka et al., 2005).

Hyperexcitability in JME was worse after sleep deprivation (Badawy et al., 2006; Manganotti et al., 2006) and in the morning hours than in the afternoon (Badawy et al., 2009). These indicate there is a circadian predominance of seizures and spikes occurring mainly in the morning. The other triggers are fatigue (73.7%), photosensitivity (36.8%), menstrual cycle (24.1%), tasks involving concentration (22.8%), and stress, excitement, or frustration (12.3%). Thus, environmental determinants greatly influence the frequency and severity of the disorder.

1.4.2 Neuroimaging in JME

Classical characterization of JME indicates no structural abnormalities associated with the disorder. However, with more advancement in imaging and computation, subtle changes in micro-structural connectivity are being reported. Structural MRI studies using

voxel-based morphometry (VBM), surface-based morphometry, and diffusion tensor imaging (DTI) have indicated abnormalities such as - cortical layers thickening or thinning and volume reduction in the size of the thalamus and hippocampus.

Correlating with the frequency of generalized tonic-clonic seizures in patients with JME, in studies employing diffusion tensor imaging, a reduction in white matter integrity in the anterior limb of the internal capsule. This involves the anterior thalamic radiation to the frontal lobes (Deppe et al., 2008).

Studies using voxel-based morphometry (VBM) indicate subtle structural alterations in the cortex in the areas frontomesial- (Woermann et al., 1999) and frontobasal- (Betting et al., 2006) cortex. An increase in frontomesial cortical grey matter (J. H. Kim et al., 2007; Woermann et al., 1999) and a reduced frontal grey matter volume (Tae et al., 2006) are reported. Neuroimaging evidence is available for the involvement of frontal lobe and thalamocortical connections in the pathophysiology of JME (Koepp 2005). In a voxel-wise meta-analysis of 325 JME patients, an increase in the gray matter in the frontal brain regions, left and right supplementary motor areas are ascribed to JME. Also, a reduction in gray matter volume was seen in left thalamus and left insula (Kazis et al., 2021). These recent findings are in line with the earlier findings of alterations in thalamic networks and the theory of thalamocortical pathways in the causation of juvenile myoclonic epilepsy (Pascalicchio et al., 2007).

A reduction in N-acetyl- aspartate (NAA) concentration has been reported with JME in studies using magnetic resonance spectroscopy. A reduction in NAA in the primary motor cortex, thalamus, and the mesial pre-frontal cortex were seen. These are suggestive of neuronal damage. Also, the ratio of excitatory glutamate-glutamine varied. It was decreased in the same cortical areas but increased in the insula

and the (left) striatum (K. Lin et al., 2009). Functional imaging studies using PET, fMRI, and scans alike for JME are in their early days. So far, they have led to the identification of the inability of JME patients to activate the frontal cortex in a visual memory task (Swartz et al., 1996). There was no difference found in JME patients and healthy controls in the assessment by a verbal and spatial working memory fMRI analysis (Roebling et al., 2009). Cognitive activation beyond a threshold is associated with excessive regional activation of the brain in JME. This load-dependent co-activation of the motor cortex most indicates a probable misconnection and a functional short circuit between cognitive systems and the motor system. This could explain the mechanism of cognitively triggered seizures in patients.

It is suggested that the micro-structural connectivity may represent the anatomical basis for cognitive triggering of motor seizures in JME. Changes in the mesial frontal region occurring with the increased functional connectivity are associated with JME. Although there is no consistent correlation between the structural and functional alterations and their mutual relationship to JME, these structural profiles provide an explanatory framework. They link seizure semiology, neurophysiology, neuropsychology, and imaging findings (Vollmar, 2015).

1.4.3 Treatment and management

JME is treated with anti-epileptic drugs (AEDs). Most patients respond to monotherapy. Treatment of the syndrome is with broad-spectrum anti-seizure drugs. Those effective for the generalized epilepsies include sodium valproate (VPA), phenobarbital (PB), ESM, clobazam (CLB), clonazepam (CLZ), lamotrigine (LTG), levetiracetam (LEV), topiramate (TPM), zonisamide (ZNS) and, more recently, perampanel (PER) and brivaracetam (BRV) (Stephen & Brodie, 2020). Treatment can be lifelong in some individuals, although others may remain seizure

free without medication. Seizure type and tolerability, co-morbidities and age are primary drivers of drug choice. A list of available drugs classified basis modes of action are detailed below.

Drug	Mechanism of action
Sodium valproate	Indirectly increases GABA transmission and various actions on multiple targets
Phenobarbital Clobazam Clonazepam	Activate $GABA_A$ receptors
Ethosuximide	Blocks low-voltage-activated Ca++ channels
Lamotrigine	Blocks fast inactivation of Na+ channel
Levetiracetam Brivaracetam	Modulate synaptic vesicle protein 2A
Topiramate Zonisamide	Various actions on multiple targets

Table 1: Mechanism of action of anti-seizure medication used in the treatment of JME.

Valproic acid is the most effective medication and drug of choice for the treatment of JME (Chowdhury & Brodie, 2016). It is a broad-spectrum medication and treats all seizure types of JME. However, due to its teratogenic potential, it is to be used with caution among pregnant women (Tomson et al., 2016). The EpiPGX consortium retrospective database study of 305 patients in 688 trials of anti-seizure drugs for JME, namely VPA, carbamazepine, LTG, LEV, and TPM, found that VPA offered seizure freedom in around 43% of cases.

In addition to compliance to medication, patients benefit significantly from lifestyle changes and disease management. Proper lifestyle

choices that include a good sleep-wake cycle, avoiding alcohol consumption, and avoiding drug abuse are vital to prevent seizure recurrence. Juvenile myoclonic epilepsy is recognized to have lifelong medical and psychosocial co-morbidities. These include depression, anxiety, sleep disturbance, impulsiveness, social isolation. (Giorgi et al., 2016; Somayajula et al., 2015). In such cases, early counseling regarding the diagnosis, management, and potential outcomes offers realistic patient expectations and treatment options.

1.4.4 JME genetics

Juvenile myoclonic epilepsy is a heritable subtype of genetic generalized epilepsy with Mendelian and polygenic genetic architecture. The expression of mutant alleles and seizure phenotypes is influenced by penetrance and environmental factors (Wolf et al., 2015). Up to 65% of all JME cases were found to be family history positive where both clinical manifestations and EEG patterns were inherited over generations (Panayiotopoulos et al., 1994, Jayalakshmi et al., 2006; Renganathan & Delanty, 2003)

Several genetic studies have led to the identification of genetic loci (Table-2) and mutations in *CACNB4* (Escayg, De Waard, et al., 2000), *EFHC1* (Suzuki et al., 2004), *GABRA1* (Cossette et al., 2002), *CASR* (Kapoor et al., 2008), and *GABA-A* (May et al., 2018), *CILK1* (Bailey et al., 2018), with the majority of the coding for ion channels (Bailey et al., 2018; Delgado-Escueta, 2007; Delgado-Escueta et al., 2013; Noebels, 2002). It has also been suggested that some single nucleotide polymorphism alleles are involved in *ME2, BRD2*, and *CX36* (Delgado-Escueta et al., 2013). In addition, whole-exome sequencing (WES) studies on large international cohorts of sporadic cases (Epi25 Collaborative. Electronic address & Epi, 2019; Epi & Epilepsy Phenome/Genome, 2017) led to the hypotheses that GGE and JME phenotypes are highly heterogeneous, involve rare or ultra-rare

variants, and are determined by a polygenic model of inheritance (Leu et al., 2019)

Locus	Type of study	Reference
2q34	Meta-analysis of genome-wide linkage studies	Leu et al., 2012 (EPICURE Consortium)
2q33-q36	Linkage analysis	Ratnapriya et al., 2010
3q26-3q27.1	Linkage analysis	Sander et al., 2000
5q12-q14	Linkage analysis	Kapoor et al., 2007
5q34	Meta-analysis of genome-wide linkage studies	Leu et al., 2012 (EPICURE Consortium)
5q34	Linkage analysis	Hempelmann et al., 2006
6p21.3	Linkage analysis	Durner et al., 2001
6p12	Linkage analysis	Hempelmann et al., 2006
8q24	Linkage analysis	Zara et al., 1995
9q32-q33	Linkage analysis	Baykan et al., 2004
10p11.22	Linkage analysis	Kinirons et al., 2008
10q25-q26	Linkage analysis	Puranam et al., 2005
14q23	Linkage analysis	Sander et al., 2000
15q14	Linkage analysis	Elmslie et al., 1997
18q21	Linkage analysis	Durner et al., 2001
19q13	Linkage analysis	Hempelmann et al., 2006
21q22.3	Meta-analysis of genome-wide linkage studies	Leu et al., 2012 (EPICURE Consortium)

Table 2: Genetic loci associated with genetic generalized epilepsy

Mutations in *GABRA1* (Gamma aminobutyric acid receptor subunit alpha-1) are associated with JME. First identified in a French-Canadian family with a LOD score of 3.1 at locus 5q34 (Cossette et al., 2002). GABRA1 is a brain-expressed ligand-gated chloride channel, a subunit of the heteropentameric GABA receptor. Human epilepsy *GABRA1* mutations replicated in mammalian cells and mice presented defective

postsynaptic currents affecting amplitude and decay time, absence like seizures at juvenile stages that continued and expanded to myoclonic seizures, increase in pyramidal cell spines density, and defective GABA receptor composition (Arain et al., 2012; Cepeda et al., 2014; Cossette et al., 2002; J. L. Fisher, 2004; Macdonald et al., 2004). *GABRA1* (-/-) GABAergic cells showed reduced innervation field and GABRA1 missense mutation (α1-A322D) induced a striking increase of spine density in pyramidal cells along with an increase in the number of mushroom-like spines (Lachance-Touchette et al., 2014). Zebrafish *gabra1* knockouts show seizures in juvenile stages with photic stimulation and defective inhibitory synaptic networks during development (Samarut, Swaminathan, et al., 2018). The presence of altered synaptic functions and presentation of seizures in *Gabra1* mutants and null animal models explains the phenotype presented by JME patients with *GABRA1* mutations.

Linkage analysis, genome-wide association studies, and candidate gene sequencing studies led to the identification of mutations in ion channel genes and receptors to be associated with JME. These include:

CASR (Calcium-sensing receptor) at 3q13.3-q21 in an Indian family (Kapoor et al., 2008). Calcium-sensing receptor, CASR is a brain-expressed seven-transmembrane G protein-coupled receptor that is important to maintain calcium homeostasis (Gama & Breitwieser, 1998). *CASR* was found to regulate voltage-gated sodium channels (VGSC), Ca2+ activated potassium (BK) channels currents, and non-selective cation channel (NSCC) activity modulating neuronal excitability (Jones & Smith, 2016)Chen et al. 2010; (Mattheisen et al., 2018). *Casr* null mice exhibit parathyroid hyperplasia, reduced growth, osteomalacia, and premature death (Ho et al., 1995; Q. Tu et al., 2003). CASR also regulates dendritic and axonal growth of the prenatal peripheral nervous system and postnatal hippocampal pyramidal neurons (Jones & Smith, 2016; Vizard et al., 2008). Thus, CASR

protein has a multi-modal influence on neuronal growth, excitability, and seizure manifestation.

Mutations and single nucleotide polymorphisms in gamma-aminobutyric acid receptor, delta (GABRD), a protein highly expressed in brain have been identified in individuals with GGE (Dibbens, Harkin, et al., 2009). Mutations in the gene are associated with Rett like syndrome (Okamoto et al., 2015), autism spectrum disorders (Y. Feng et al., 2010; Sesarini et al., 2015). Homozygous *Gabrd* null mice exhibited postpartum depression causing maternal neglect and reduced pup survival. Animals are also reported to have an impaired memory, enhanced fear acquisition. Defects in adult neurogenesis, maturation, dendritic, and plasticity are also reported (Maguire & Mody, 2008; Whissell et al., 2013).

Calcium voltage-gated channel auxiliary subunit beta 4 (*CACNB4*) codes for the auxiliary β4 subunit of the P/Q-type calcium channels. These channels regulate the amplitude, kinetics, and voltage dependence of calcium currents of high-voltage-gated calcium channels. Thus, influence rapid neurotransmitter release in the brain. Variants in the gene were identified in a cohort of GGE patients including 49 JME patients (Escayg, De Waard, et al., 2000). *CACNB4* is also proposed to be a neurodevelopmental disease gene. The missense variant Leu126Pro is associated with severe intellectual disability, seizures, visual impairment, and dystonic and athetoid movements. In cells and hippocampal neurons, this mutation impairs the formation of synaptic P/Q-type calcium channel complexes, prevents activity-dependent nuclear targeting and nuclear functions, and impairs channel and non-channel functions (Coste de Bagneaux et al., 2020)

Mutations in *SCN1A*, a voltage-gated ion channel expressed in the brain, have been identified in patients with GEFS+, and Dravet Syndrome, and JME patients (Bender et al., 2012; Escayg, MacDonald,

et al., 2000; Moulard et al., 1999). Missense mutations in *SCN1A* have been identified in European, Turkish and Malaysian populations in JME patients (Chan et al., 2020; Escayg et al., 2001; Lal et al., 2016). Interestingly, *CACNB4* heterozygous mutation p.Arg468Gln in a *SCN1A* mutation-positive background resulted in severe myoclonic epilepsy in infancy. Cell culture-based assays resulted in increased Ba+2 current density in electrophysiological experiments (Ohmori et al., 2008).

CACNA1G, a low voltage-activated Ca(v)3.1 T-type calcium channel, was selected to be screened in JME patients since mutations in its associate subunit *CACNA1H* have been identified in CAE patients (Heron et al., 2007). *CACNA1G* mutations in Japanese JME patients were identified (B. Singh et al., 2007). Initial findings reported that slightly faster inactivation decay rates were observed in some mutant channels. However, the mechanism by which the susceptibility alleles affect the protein's function and lead to seizure phenotype remains to be elucidated.

These are some examples of ion-channel genes, their roles, and mutations affecting function, structure, and properties, contributing to epilepsy by disturbing neuronal homeostasis, growth, and neurotransmitter release. Another class of molecules functioning along a parallel axis in the causation of epilepsy are components of centrosome and ciliary machinery, such as *EFHC1, CILK1,* and *EFHC2*. These proteins are key players in cell division, migration, signaling, and ciliogenesis. This thesis aims at evaluating the association and roles of microtubule or centrosome-associated proteins to JME. In the following section, I describe key roles of *EFHC1*, *EFHC2*, and *CILK1*.

EFHC1

EFHC1 variants are the most common mutations in inherited forms of juvenile myoclonic epilepsy (JME). Mutations in *EFHC1* were first identified in JME patients in a Mexican family at the 6p12-11 locus,

named EJM1 (Suzuki et al., 2004). Five missense mutations (Asp210Asn, Phe229Leu, Asp253Tyr, Pro77Thr and Arg221His) were identified in six Mexican families and an SNP (Arg182His) in a Belize family with JME. In the last few decades, a total of 54 variants have been identified in *EFHC1* in association with JME. In a recent systematic review, the 54 variants were re-analyzed using statistical and bioinformatic tools. Applying the National Human Genome Research Institute (NHGRI) and American College of Medical Genetics and Genomics (ACMG) guidelines for interpretation of sequence variants, the variants were reclassified. This led to 9 *EFHC1* variants being "pathogenic," 14 variants as "likely pathogenic," and 20 variants "of unknown significance", eight variants were benign and three were likely benign (Bailey et al., 2017). This study offers the most comprehensive evaluation and evidence of EFHC1 as a JME gene.

It is important to note that like in many disorders with complex inheritance and variable penetrance of alleles, some *EFHC1* variants may not be sufficient by themselves to cause epilepsy. The phenotypes may be exaggerated on conjunction with other alleles associated with epilepsy or mutations occurring in proteins of a common conserved pathway.

EFHC1 is an evolutionarily conserved gene expressing two isoforms in humans. *EFHC1* spans about 72Kb in the human genome and encodes a protein of 640 amino acids called myoclonin1 that harbors three tandemly repeated DM10 domains of unknown function and an EF-hand, a Ca2+ -binding motif at the C-terminus. EFHC1 is a pro-apoptotic protein. In brain hamster kidney BHK-cell lines, EFHC1 interacts with the R-type voltage-gated calcium channel (Cav2.3) and leads to increase in the Ca2+ current (Suzuki et al., 2004).

EFHC1 is a microtubule-associated protein (MAP) that is crucial for regulating cell division and neuronal migration during cortical

development (de Nijs et al., 2009). Mammalian EFHC1 is orthologous to Rib72, an axonemal protein of *Chlamydomonas reinhardtii* (Ikeda et al., 2005; King, 2006). In *Tetrahymena, C.elegans,* and in mouse brain ependymal cells, EFHC1 localizes to the ciliary basal body and cilia (Loucks et al., 2019; Stoddard et al., 2018). Its presence determines the ciliary beat frequency and functioning of cilia (Stoddard et al., 2018; Suzuki et al., 2009). In *C. elegans*, EFHC1 was recently observed in specialized non-motile mechanosensory cilia, regulating neuronal activation and dopamine signaling. The protein modulated dopamine signaling at the synapse in cooperation with the orthologue of an R-type voltage-gated calcium channel. Thus, EFHC1 is involved in the dual-regulation of neuronal excitability at neuronal sensory input (cilium) and neuronal output (synapse) sites. Extrapolating these observations to mammalian neurons suggests that in neurons where non-motile cilia are present, EFHC1 in cilia may modulate dopamine-mediated signaling and neuronal excitation thresholds (Loucks et al., 2019).

Further, in mice *Efhc1*-deficiency causes spontaneous myoclonus and increased seizure susceptibility. *Efhc1* null mice have significantly enlarged ventricles of the brain and reduced ciliary beating frequency as compared to animals heterozygous for the mutation. In both heterozygous and homozygous mutants, the threshold of seizures induced by pentylenetetrazol was decreased significantly (Suzuki et al., 2009). Interestingly, *Defhc*1.1 knockout flies exhibited increased dendritic arbors, satellite boutons resulting in excess spontaneous neurotransmitter release (Rossetto et al., 2011). Taken together, we know that EFHC1 plays critical roles in cell division, migration, signaling, and mutant proteins disrupt one or more of these functions and cause seizures. However, the exact mechanism that leads to epilepsy remains unelucidated.

EFHC2

EFHC2 is situated on the X-chromosome at Xp11.3 and codes for EF-hand domain containing 2. This protein is a paralog of *EFHC1*. EFHC2 is the second protein within the DM10 domain family which consists of three DM10 domains and one C-terminal EF-hand motif. Pairwise sequence alignment showed that EFHC1 and EFHC2 share an overall similarity of 41.6% at the amino acid level. Though detailed genetic studies on human *EFHC2* are not reported, an earlier association study implicated *EFHC2* as a genetic risk for JME in German population (Gu et al., 2005). Because of its similarity to EFHC1 and association with JME, it is a potential JME gene. In chapter 2 of this, thesis, I consolidate the identity of variants in *EFHC2* associated with JME.

EFHC2 is a conserved ciliary protein. In *C.elegans, Tetrahymena* and in *Xenopus* embryos *EFHC2* has been shown to the cilia in multiciliated cells (F. Tu et al., 2018). EFHC2 is expressed in motile cilia containing human bronchial epithelial cells (HBEC) (X. Sun, 2009). Considering that in *C. elegans*, EFHC1 has been shown to have dual roles at the cilia and at synapses, it would be important to understand the significance of ciliary localization of EFHC2. In mammalian cells, the cellular localization, functions, and properties of the domains comprising EFHC2 are less explored. In this thesis, I have examined the cellular attributes of *EFHC2* and its variants.

Along with its association to JME, *EFHC2* SNPs and variants have been identified with social, cognitive, and behavioral deficits (Blaya et al., 2009; Weiss et al., 2007). To validate these associations, I have studied *Efhc2* in mouse and generated two mutant mouse lines as detailed in chapter 3.

CILK1

Ciliogenesis associated kinase-1 *(CILK1),* is also known as intestinal cell kinase (ICK), encodes an evolutionarily conserved serine/threonine kinase. In the human genome, it is a gene present on the 6th chromosome, 0.58Mb away from *EFHC1*. CILK1 is critical for the development of nervous, endocrine, and skeletal systems. *CILK1* and its homologs are components of several cellular processes such as - cell proliferation and cycle progression, ciliogenesis, hedgehog signaling inter-flagellar transport (IFT), and in autophagy (Fu et al., 2019).

CILK1 mutations were first reported in a multisystem, neonatal lethal condition called Endocrine-Cerebro-Osteodysplasia (ECO) syndrome in an Amish family (Lahiry et al., 2009). Homozygous mutations in the catalytic domains of *CILK1*, such as p.Gly120Cys and p.Arg272Gln, p.Glu80Lys, lead to neonatal lethality with clinical phenotypes correlating to ciliopathies such as short rib polydactyly syndrome (Moon et al., 2014; Oud et al., 2016; Paige Taylor et al., 2016). Using mammalian cell lines and patient-derived fibroblasts, it was shown that *CILK1* mutations affect localization and ciliogenesis.

In a recent study, *CILK1* was identified as a causative JME gene (Bailey et al., 2018). Twenty-one pathogenic heterozygous variants in *CILK1* were identified in patients with juvenile myoclonic epilepsy of Latin American and Japanese origin. A few of these variants studied using ex vivo electroporation in the mouse brain at E14.5 caused mitotic defects, delayed cell-cycle exit, and impaired radial neuroblast migration while promoting apoptosis. In replicative studies with 1149 individuals with GGE, no pathogenic CILK1 variants were identified. This led to the speculation that the contribution of *CILK1* variants to JME could be a population- specific observation. In this thesis, I have examined a set of Indian JME patients for *CILK1* variants; findings are reported in chapter 4 of the thesis.

CILK1 knockout and knock-in mutations recapitulate human ciliopathies. CILK1 null and mice harboring inactivating mutation Arg272Gln are perinatal lethal (Chaya et al., 2014; Moon et al., 2014; Tong et al., 2017). In a *Cilk1*-deficient mouse model, embryos displayed enlarged cerebral cortexes and hydrocephalus (Chaya et al., 2014). These embryonic phenotypes are associated with reduced number and activity of neuronal progenitor cells, aberrant ciliogeneisis, and disrupted sonic hedgehog signaling (Chaya et al., 2014). The aberrant sonic hedgehog signaling in *Cilk1* null mice leads to the dramatically shortened intestine and decreased cell proliferation (Y. Yang et al., 2021). Evaluation of heterozygous *Cilk1*$^{+/-}$ mice for spontaneous and isoflurane-provoked seizure phenotype concluded that the mice are epileptogenic (Bailey et al., 2018). However, replicative studies with *Cilk1*$^{+/-}$ and *Cilk1*$^{R272Q/+}$ did not generate spontaneous or isoflurane-induced seizures (Salvati et al., 2021) refuting the idea that in mice, heterozygous mutations lead to epilepsy.

Along with the above-listed genetic susceptibility factors, several copy number variations and leads from GWAS, meta-analysis, case-control, and family-based association studies and large samples analyzed through the international epilepsy consortia have provided key pointers. These are discussed briefly in the following section.

Association studies

For JME, many candidate genes chosen based on their physiological role or their localization in and around JME-linked loci have been analyzed. The first locus for JME was found to be in the proximity of the HLA region on chromosome 6. Preliminary evidence for the association between JME and alleles in the HLA region was reported in German (Durner et al., 2000) and Saudi Arabian (Obeid et al., 1994) populations. The frequency of HLA-DR13 and HLA-DQB1 alleles was significantly higher in JME patients than non-JME patients, indicating

that the JME locus probably lies within the HLA region (Greenberg et al., 1996). However, other groups found no significant evidence of association of HLA alleles with JME in various populations (Le Hellard et al., 1999; Sander et al., 1997).

There is an association between JME and a core haplotype of 5 SNP and microsatellite markers in the 6p21 critical region, with LD peaking at the *BRD2* gene at 6p21.3 (EJM3) (Pal et al., 2003). Two JME-associated SNP variants in the promoter region of BRD2 were found on sequencing; however, no other potentially causative coding mutations were identified in 20 probands in families showing positive linkage to the locus. Subsequently, a multicenter study examining the role of *BRD2* as a risk factor for JME replicated the result in the British cohort, found suggestive evidence for its role in an Irish cohort, but were not replicated in the German, Australian, and Indian cohorts studied (Cavalleri et al., 2007)

A significant association between JME and the c.588C>T variant, located in exon 2 of the *CX36* gene at 15q14 (Elmslie et al., 1997), was observed in a group of 46 JME patients, including 29 individuals from families (Mas et al., 2004). *CX36* is also known as *GJD2*. Bioinformatics analysis of the associated SNP suggested that it perturbed the expression of CX36 by affecting exonic splicing enhancers.

For the GGE-linked locus at 18q21, case-control and family-based association methods identified *ME2* (malic enzyme 2) as a gene predisposing to GGE. An increased risk to GGE was found when ME2-centered nine-SNP haplotype was present in the homozygous state (Greenberg et al., 2005). However, this association was not reproducible in a population-based association study involving 660 GGE patients and 666 controls of German descent (Lenzen et al., 2005).

In a case of benign familial neonatal convulsions, it was found that mutations in the potassium voltage-gated channel, KQT-like subfamily, member 2 (*KCNQ2*) and member 3 (*KCNQ3*) genes caused this monogenic subtype of IGE (N. A. Singh et al., 1998); (Charlier et al., 1998). Suggestive evidence for an allelic association between *KCNQ3* and JME using a family-based association study; and the presence of protective and susceptibility effects of *KCNQ1* allelic variants on JME have also been reported (Vijai et al., 2005).

Copy number variations

Structural variants resulting from deletions or duplications, ranging from tandem to complex multisite chromosomal rearrangements larger than 1Kb running into to several megabases in size, are referred to as CNVs. CNVs are found on all chromosomes without any clinical significance and are often distributed throughout the genome impartially, encompassing several genes. CNVs can be inherited or be of de novo origin and are associated with both simple and complex disorders. CNVs are increasingly found in epilepsy patients when often associated with developmental delay, autism spectrum disorder (ASD), intellectual disability, and dysmorphic features (Mefford, 2015; Sharp et al., 2008). Although micro-deletions investigated are individually rare in patients with GGE, they do collectively account for a sizable fraction of the genetic variance.

In the case of GGEs, the most frequently CNV is microdeletion at 15q13.3. This deletion is found in 1% of patients with IGE (Dibbens, Mullen, et al., 2009; Helbig et al., 2009). Recurrent micro-deletions at 15q11.2 and 16p13.11 have also been reported in patients with JME. These three 'hot spot' CNVs are seen in an estimated 3% of patients with generalized epilepsy (Mefford et al., 2010).

1.5 Epilepsy consortia studies

Next-generation sequencing approaches have rapidly expanded the field of epilepsy genetics. NGS offers powerful tools such as exome, genome, and gene-panel-based examination to study genetic variants' contribution to a disorder in a single patient or in a large cohort. In GGE, one of the earliest studies examined 118 European GGE patients by whole-exome sequencing. Further, potential candidates identified in the screen were examined in 878 GGE cases (Heinzen et al., 2012). Among the top-ranking variants, heterozygous *GREM1* Pro35Ala was the most enriched GGE associated variant located near the epilepsy CNV hotspot 15q13.3, while heterozygous missense variants in *PSME2, BTD,* and *PEX6* and homozygous variant in *AGPAT3* were absent in controls and found in more than three epilepsy cases (Heinzen et al., 2012). Although this did not result in the identification of a single gene variant as causative for GGE, it fueled further explorations in the area using such approaches. Considering the identification of gene variants in ion channel genes for GGE, targeted Sanger sequencing of 237 channel genes in 152 GGE cases was carried out. This effort also did not identify any significantly associated variant or gene or gene family (Klassen et al. 2011). In these cases, the influence of common and rare variants and their complex interactions were inferred to contribute to disease causation.

Several AEDs target the *GABA-A* receptors and are efficient in disease management. Linkage and candidate gene sequencing studies have identified several gene variants in *GABA-A* receptors. In line with this, in the WES of 152 European GGE patients, while there was no enrichment of a single gene variant, gene set analysis identified missense variants in *GABA-A* receptor genes to exhibit significant enrichment that was recapitulated in two other cohorts of 357 and 583 GGE European patients. These variants were equally distributed among the *GABA-A* receptor genes (Klassen et al., 2011)

To date, the most powerful genetic analysis for epilepsies has come from worldwide collaborative studies involving large sample sizes of patients and controls. These data sets also feature samples across many geographic and ethnic backgrounds, thus eliminating sample set biases. The first study of this kind analyzed 4,000 epilepsy genomes, including epileptic encephalopathy, genetic generalized epilepsy, and focal epilepsy patients from multi-generational families, pairs, and sporadic individuals (Epi, 2012). Analysis of familial genetic generalized epilepsy patients of European ancestry by a case-control study by exome sequencing identified enrichment of ultra-rare variations in known 43 dominant epilepsy and 33 epileptic encephalopathy genes (Epi & Epilepsy Phenome/Genome, 2017). No single gene was study-wide significantly associated with the disorders, while the top ten ranking genes with functional variants included three known epilepsy genes (*KCNQ2, GABRG2*, and *SCN1A*) and the rest were *ATP1A3, CACNA1B, COPB1, KEAP1, SLC9A2*.

In the ongoing largest study of whole exome sequencing of 25,000 epileptic individuals through the Epi25 collaborative, three primary epilepsy types are being examined. These are developmental and epileptic encephalopathy (DEE), genetic generalized epilepsy (GGE) and non-acquired focal epilepsy (NAFE). So far, analysis of 13,487 epilepsy-affected individuals and 15,678 control individuals has been performed. In comparison to controls, individuals with epilepsy had an excess of ultra-rare, deleterious variants in constrained genes and genes previously associated with epilepsy. No single gene exhibited study-wide statistical significance while known or potential epilepsy associated genes exhibited higher burden of ultra-rare variants whose allele count is not more than three. However, inhibitory $GABA_A$ receptor genes showed enrichment for missense variants across all three classes of epilepsy. Among individuals with GGE, ultra-rare nonsynonymous variants, notably protein truncating variants were

present in genes intolerant to loss of function changes. In the analysis of missense ultra-rare variants in known epilepsy and brain expressed genes at least 2-fold enrichment of URVs were seen in GGE patients over controls (Epi25 Collaborative. Electronic address & Epi, 2019). Among individuals with GGE or NAFE, no single gene surpassed exome-wide significance. Singleton ultra-rare variants in *CACNA1G* and *UNC79* were top ten ranking GGE-associated, brain-enriched and novel epilepsy genes that function as ion channels or ion channel interactors. The rest top-ranking associations for GGE were found with *CACNA1G, EEF1A2*, and *GABRG2* for GGE and *LGI1, TRIM3*, and for NAFE, *GABRG2* was among the top associations identified (Epi25 Collaborative. Electronic address & Epi, 2019). A list of genes with the burden of ultra-rare variants was published for each of the three primary epilepsies. In chapter 5 of this thesis, I present our preliminary findings on analyzing WES data of small families with two or more affected individuals with GGE against this list.

1.6 Modeling epilepsy

Genetically and pharmacologically manipulable model systems that recapitulate the disease phenotype are potent tools in understanding the pathophysiology and mechanism of diseases, especially of heritable nature. In the case of epilepsy, while over a hundred single gene causes are known, their mechanism remains to be identified. Epilepsy is a phenotypically complex disorder with a significant overlap of causative genes also being implicated in autism, intellectual disability, and other neurodevelopmental disorders. The inaccessibility of human brain tissues and the lack of robust computational models for these disorders coupled with the complex nature of the disease makes understanding biological mechanisms involved in humans challenging. Models of epilepsies in other organisms, and more recently using iPSCs, brain organoids, therefore, provide an opportunity to overcome some of these limitations. They are proving helpful in elucidating the

neurological basis of seizure generation and identification of novel genes and therapeutics.

The pre-requisites for a successful animal model are (i) construct validity, (ii) recapitulation of the features and etiology of a human neurological disorder (iii) face validity, with phenotypes comparable to that in patients; and (iv) predictive validity, specificity of response to treatments used in patients with the modelled disorder (Garner, 2014; Willner, 1984). Such animal models that meet the above criteria offer a robust and an indirect window into human pathology. In epilepsy research, *C.elegans, Drosophila*, zebrafish, mouse, rats, human iPSCs and brain organoids are used.

1.6.1 Caenorhabditis elegans

The transparent, microscopic round worm *C.elegans* has a simple and well-characterized nervous system with 302 neurons. Morphologically, these neurons resemble mammalian neurons. The neurotransmitters are also conserved between *C.elegans* and mammals. Also, around 80% of protein-coding genes have human homologs (Lai et al., 2000). They are a robust, inexpensive, and malleable system in which electrophysiological recordings are possible. Pentylenetetrazol (PTZ), a GABA-A receptor antagonist a common proconvulsant in mammals is also effective in *c elegans*. The most striking difference is that *C.elegans* lack voltage-gated sodium channels, which are found in mammals (Bargmann, 1998).

The phenotypes assessed are convulsions or head bobbing in terms of modeling seizures. Convulsions have been reported in the genetic models of lis-1, unc-49, unc-47, or unc-25 (Calahorro & Ruiz-Rubio, 2013; Williams et al., 2004). On exposure of PTZ in these mutants, pharyngeal pumping stops, and a dose-dependent paralysis is seen. The pnm-1 (a nonsense allele of lis-1) that escaped the early

developmental lethality associated with LIS1 inactivation demonstrated 'head-bobbing' convulsions on exposure to PTZ (Williams et al., 2004). Furthermore, worms depleted for the LIS1 pathway components show genetic interactions that greatly enhance sensitivity to convulsions.

C.elegans is a genetically tractable model whose complete neuronal connectivity has been defined. Also, the worms present seizure-like behavior. Several previously isolated mutants, a range of phenotypic assays, and the availability of GFP markers specific to neuronal classes and processes make them a very attractive model system in this field.

1.6.2 Drosophila melanogaster

The rich history of *Drosophila* research offers genetic, molecular, and transgenic experimental capabilities to uncover the underlying pathophysiology of human disorders. The ease of handling, low maintenance cost, and the availability of a comprehensive fly repository are some of the benefits of this model system. Also, the availability of behavioral assays comparable to seizure disorders make them a tenable model system.

The mammalian Voltage-gated and ligand-gated signaling ions Na+, K+, and Ca2+ channels, and acetylcholine, glutamate, and gamma-aminobutyric acid (GABA) transmitter receptors all well conserved or have homologs in the fruit fly. The significant difference is in the organization of the central nervous system. In flies' ganglionic structure with synaptic neuropilar regions are present. In contrast, the mammalian cortex is organized into synaptic layers. However, flies are susceptible to electrical shock delivered to the brain and elicit seizure-like in appearance and neuronal spiking activity (Kuebler & Tanouye, 2000; J. Lee & Wu, 2006). With severe seizure-like behaviors and electrical abnormalities as readouts, seizure suppressor and activator mutations have been identified. Examples of seizure suppressors are -

para^{bss} mutations (human ortholog is SCN1A) mapped to the *Drosophila* Na+ channel gene para (Parker et al., 2011), and top1^{JS} mutation (human ortholog - DNA Topoisomerase I) is a general seizure suppressor correcting phenotypes of sda (slamdance), eas (easily shocked), and bss (bang senseless) (J. Song et al., 2007); fly prickle^{sple} (pk^{sple}) displayed seizures that were corrected by anti-epileptic medication (Tao et al., 2011). This model system has also been important for the discovery of ion-channel mutations like seizure (hERG), shaker (Kv1), and slowpoke (bk). These mutants were identified using temperature-sensitive (ts) or ether-induced - leg-shaking phenotype (Baraban, 2007). The bang-sensitive (bs) mutants, displaying several genetic modifiers for seizure activity, have helped explore the heterogeneous nature of epilepsy (Kuebler et al., 2001) as well as for high throughput screening to identify novel anticonvulsants and seizure susceptibility loci (Stilwell et al., 2006).

Drosophila allows the study of human mutated genes using CRISPR/Cas9 to generate knock-in mutants where the endogenous copy of the gene is replaced by a mutated version of the human homolog. Such flies can then be used for large-scale screening behavior and physiology and drugs for therapy. Maintenance costs for flies are low. They have a short life cycle. They can be administered with drugs via food. These factors make them suitable for high-throughput drug screening. The availability of tools and methods to perform electrophysiological recording and stimulation are at various stages of the life cycle are benefits of this system (Parker et al., 2011; Ugur et al., 2016).

1.6.3 Zebrafish

Zebrafish (*Danio rerio*) is a powerful tool in neuroscience research. The genetic tractability, conservation of genes and physiological pathways, small body size, ease of experimental manipulations in vivo, and rich

behavioral repertoire make them a popular choice (Fontana et al., 2018). Both larvae and adult zebrafish are susceptible to seizures, to kindling by PTZ, kainic acid, and to EEG recordings which have been developed in both larval (Y. Lee, Seo, et al., 2020) and adult (Y. Lee, Lee, et al., 2020) forms.

For epilepsy modeling in zebrafish, forward genetics approaches using (ENU)-induced model (with mutations in scn1, scn8, and hcn1) for Dravet syndrome (Marini et al., 2011), scn1a mutants haploinsufficiency for the voltage-gated sodium channel (Escayg & Goldin, 2010; Griffin et al., 2016)) and the zebrafish models prickle1/prickle 2(Tao et al., 2011) are classic examples. Zebrafish are also responsive to photic stimulation – GABRA1 knockouts show seizures in juvenile stages with photic stimulation and defective inhibitory synaptic networks during development (Samarut, Dona, et al., 2018).

1.6.4 Rat

With increasing efforts of referencing the rat genome, transgenesis and site-specific editing are now feasible in rats (Guan et al., 2014). The rat brain anatomy is similar to that of humans. Also, the neural circuits, genes associated with neurological disease, and network activity patterns are homologous to those in humans (Huang et al., 2004)

Over the last 30 years, a number of models have been developed that display varying degrees of pharmacoresistance, such as the phenytoin- or lamotrigine-resistant kindled rat, the intrahippocampal kainate models and rats in which spontaneous recurrent seizures develop after inducing status epilepticus by chemical or electrical stimulation. (Loscher, 2017). The most popular and well characterized rat epilepsy models are inbred strains with spontaneous mutations. These include GAERS and WAG/Rij strains (Danober et al., 1998). Abundant data

have been collected on the Generalized Epilepsy Prone Rat (GEPR) which presents convulsive seizures (Jobe et al., 1991). Rats have also been used to examine seizure susceptibility and epileptogenesis in vitro and in vivo. Tethered (Medlej et al., 2019) and wireless (McGuire et al., 2019) EEG setups are available. *ex vivo* brain slice preparations protocols are also available (Accardi et al., 2018).

1.6.5 Mouse

Mus musculus, are strong candidate organisms for modeling human neurogenetic disorders since (i) over 99% of mouse genes have human homologs (Mouse Genome Sequencing et al., 2002), (ii) the spatio-temporal gene expression patterns are well conserved in the mouse and human brain (Strand et al., 2007), and (iii) mouse models of epilepsy mimic patients in seizure patterns and involvement of brain structures (Riban et al., 2002).

The ease of genetic manipulation using homologous recombination and CRISPR/Cas9 approaches have propelled their usage in studying neurological disorders. Broadly, mice models can be categorized into genetic models and induced models. In the case of genetic models, mutations (spontaneous or engineered) result in spontaneous seizures, and in induced models, chronic seizures are inducted by chemical, electrical or acoustic stimulation.

Several examples of the use of genetic mouse models for epilepsy are now available. For example, it is known that in humans, a *de novo* mutation in the *SCN1A* gene causes Dravet syndrome due to a loss of function of the type I voltage-gated sodium channel (Nav1.1) (Claes et al., 2001). With this information, many mouse models of the Dravet syndrome have been developed. Spontaneous and hyperthermia-induced seizures are observed in Scn1a$^{+/-}$ mice (Oakley et al., 2009).

KCNA1 gene encoding the Kv1.1 voltage-gated potassium channel is critical for regulating numerous features of neuronal function. In humans, partial temporal lobe seizures and generalized tonic-clonic seizures are observed due to dominant-negative missense mutation in a potassium channel gene (*KCNA1*) (Zuberi et al., 1999). *KCNA1* null mice exhibited the following features - 50% of the mice showed generalized seizures and died suddenly at 3-5 weeks of age. The mice that live continued to display sporadic spontaneous seizures (Smart et al., 1998).

To study JME, *Efhc1* null mice and *Cilk1* knockout and knock-in mouse models are available. *Efhc1*-deficiency in mice causes spontaneous myoclonus and increased seizure susceptibility. *Efhc1* null mice have significantly enlarged ventricles of the brains and reduced ciliary beating frequency compared to animals heterozygous for the mutation. Furthermore, the threshold of seizures induced by pentylenetetrazol was reduced considerably in both heterozygous and null mutants (Suzuki et al., 2009). Although, no prominent abnormal EEG activity was observed in the null or heterozygous mutants during myoclonus, the twitching was confirmed by electrocorticograms (EcoGs) during involuntary movements. The lack of EEG abnormality during spontaneous myoclonus and evoked potentials in the EcoG recordings during involuntary but not voluntary movements were attributed to the origin of the seizures being thalamic or involving the brain stem rather than the cortex (Suzuki et al., 2009).

The *Cilk1* mouse, while offering phenotypic and construct validity in the context of ciliopathies, is a subject of debate concerning the epileptic phenotype and JME. *Cilk1* homozygous null and Arg272Gln mice are perinatal lethal. In a genetic study, evaluating the heterozygous *Cilk1* null (*Cilk1*+/-) mice for spontaneous and isoflurane provoked seizure phenotype concluded that the mutations are epileptogenic (Bailey et al., 2018). However, in a more recent re-examination of heterozygous null

Cilk1 $^{(Cilk1+/-)}$ mice and mice harboring the loss-of-function Cilk1 variant, Arg272Gln (*Cilk1*$^{R272Q/+}$) failed to reproduce the phenotype. Video capture with electrocorticogram and electromyogram (ECoG/EMG) recordings were performed under isoflurane induction and control conditions. No seizure activity or abnormal EEGs were captured in any of the conditions tested, challenging the conclusion that *CILK1* haploinsufficiency contributes to JME. These differences are suggested to be an effect of strain differences between BL6J versus BL6N impressing upon varied seizure susceptibility across different mouse strains as has been reported earlier in *Scn2a*^{Q54} (Thompson et al., 2017); genetic modifiers affecting the severity of epilepsy caused by mutation of sodium channel *Scn2a* (Schauwecker, 2011).

Another kind of mouse models are induction models. The pharmacological induction models offer the advantage of control over timing and severity of seizures basis dosage. Kainic acid, pilocarpine and pentylenetetrazole (PTZ) are widely used as seizure-inducing drugs. PTZ is a gamma aminobutyric acid (GABA)-A receptor antagonist. PTZ suppresses the function of inhibitory synapses, leading to increased neuronal activity and causes seizures. Kainic acid- or pilocarpine-mediated SE promotes chronic spontaneous and recurrent seizures. In the evaluation of seizure susceptibility, a combination of genetic models with induction through PTZ is also used.

1.6.6 Human iPSCs

Induce pluripotent cells (iPSCs) are derivatives of human tissues such as skin biopsies or blood samples that are subsequently be differentiated into cell types of interest such as neurons or glia. These cells contain the same genetic material as the original human tissue sample. Patient-derived iPSC lines are useful for modeling disorders with complex genetic architecture. Such disorders would be impractical to model in whole animals. In the case of monogenic

diseases, mutations identified in patients can be engineered into control iPSC lines using tools such as CRISPR/Cas9.

iPSCs can be grown in two-dimensional cultures or in three-dimensional scaffolds to generate cerebral organoids. These structures have physiologically realistic brain architectures (Benito-Kwiecinski & Lancaster, 2020; Niu & Parent, 2020). The organoids comprise diverse neuronal and glial cell types, with limited self-organization. They have spontaneous, synchronized neural activity (Izsak et al., 2019). They offer the most plausible alternative to animal models of epilepsy for in vitro functional studies and aid the drug discovery process.

In case of epilepsy, iPSCs have been derived from patients with Dravet syndrome. Although the excitatory neurons appear unaffected, their interneuron derivatives show hypoactivity (Y. Sun et al., 2016). Two-dimensional cultures of iPSCs carrying a KCNT1 mutation display network excitability and synchrony (Quraishi et al., 2019). Brain organoids cultured for >8 months exhibit spontaneous neuronal network activity (Quadrato et al., 2017). To date, there is limited evidence that brain organoids can generate spontaneous epileptiform seizures. Currently, this is a limitation towards the face validity of protocols for modeling epilepsy in organoids (Morris et al., 2021). Assessment of the predictive validity of iPSC models in testing anti-seizure therapies remains a challenge. This is because seizures are not well recapitulated in current versions of these models. Another limitation of the use of iPSCs is the heterogeneity in electrical properties across cells generated from different donors (Kyttala et al., 2016).

1.7 Objectives

The broad focus of the thesis has been genetic generalized epilepsy. The studies conducted aim to identify and characterize mutations identified in a set of individuals with GGE from India. Specifically, in patients with JME, the contributions of mutations in genes *EFHC2* and *CILK1* were examined. I used cell biological assays and in silico analysis on the variants identified to assess their potential pathogenicity and significance. The expression, subcellular localization, and properties of mouse *Efhc2* were examined . Mutant mouse lines for *Efhc2* were generated and some of their cell biological properties studied. In an ongoing study, 55 individuals with GGE were analyzed using whole exome sequencing and variants in certain ciliary genes were found. This study has been carried out with the following objectives:

i) Identification of potentially pathogenic variants underlying JME/GGE,

ii) Generation and characterization of mice carrying *Efhc2* gene mutations.

Chapter 2

Functional characterization of *EFHC2* and its variants

Summary

Previous studies have identified *EFHC2* as a candidate gene for juvenile myoclonic epilepsy patients (JME). *EFHC2* encodes a 2.2 kb transcript and a 749 amino acid protein. I examined a set of JME patients to identify gene variants that may be functionally relevant.

In cultured mammalian cells, EFHC2 was observed to localize to the spindle poles, midbody and in a subpopulation of cells, it decorates the midzone microtubules. In co-immunoprecipitation studies, EFHC2 interacts with α- and γ-tubulin and is a microtubule-associated protein. EFHC2 consists of three DM10 domains of unknown function. I examine the role of these domains for the cell -localization, -cycle, and interaction with microtubules. Employing *in utero* electroporation, we examined the effect of *EFHC2* alleles in the developing mouse brain at e14.5-15.5.

2.1 Introduction

EFHC2 is a gene situated on the X-chromosome at Xp11.3 (ID 80258). It is 239,908 nt long and comprises 15 coding exons. EFHC2 expresses a 3269 bases mRNA, has a 2250 bp ORF, and a predicted protein with 749 amino acids (NP_ 079460.2). It is a paralog of *EFHC1*, a gene associated with juvenile myoclonic epilepsy (Suzuki et al., 2004). EFHC1 and EFHC2 are evolutionarily conserved proteins. They have similar protein architecture consisting of three DM10 domains. EFHC and their orthologs have been identified in cilia, flagella and are associated with the doublet microtubules in axonemes (Ikeda et al., 2005; Kilburn et al., 2007; Stoddard et al., 2018). Based on a detailed analysis of the *EFHC1* variants in accordance with guidelines of the National Human Genome Research Institute (NHGRI) and American College of Medical Genetics and Genomics (ACMG), *EFHC1* was established to be a gene causative of JME (Bailey et al., 2017). EFHC1 deficient mice show brain ventricular enlargement, reduced ciliary beating frequencies, and develop spontaneous myoclonic jerks (Suzuki et al., 2009). In dividing cells, EFHC1 has been observed at the mitotic spindle and the midbody. Knockdown of *Efhc1* in cultured cells and in rat embryonic brain leads to disruption of the mitotic spindle, M-phase progression defects, and impaired cell cycle exit of cerebrocortical progenitors. Impairment of EFHC1 affects the extension of radial glial cell processes and, in turn, the accumulation of cells in the ventricular and sub-ventricular zones (VZ/SVZ) of the brain. Additionally, the mutant EFHC1 also affects the morphology of cells in the VZ/SVZ leading to multipolar morphology and short leading process. The misoriented mitotic spindles, short leading process of

cells are suggestive of perturbations in centrosome and microtubule organization (de Nijs et al., 2009).

There are suggestions for the role of *EFHC2* in neurological traits such as JME (Gu et al., 2005), fear recognition (Weiss et al., 2007), and social cognitive abilities (Startin et al., 2015). Work examining gene variants in *EFHC2* occurring in a set of JME patients strengthens this association. Examination of around 480 JME individuals from India has led to the identification of seven rare and potentially pathogenic variants in *EFHC2* (Raju, 2014). In this thesis, I extend this study by sequencing the *EFHC2* coding exons in 72 patients and in controls.

EFHC2 is a conserved ciliary protein. EFHC2 is expressed in motile cilia containing human bronchial epithelial cells (HBEC) (X. Sun, 2009). In sea urchin sperm flagella, the ortholog of EFHC2 is found associated with the lumen of the DMTs (Linck et al., 2014). In the *Xenopus* embryos, EFHC2 is present at the axonemes of multiciliated cells (F. Tu et al., 2018).

In humans, EFHC2 comprises 749 amino acids, with three tandemly occurring conserved DM10 domains. These conserved domains are ~105 amino acids in length and are poorly understood for their function(s). These domains have been reported in two distinct classes of proteins in *Chlamydomonas* and mammals. The first class of these proteins has a single copy of the DM10 domain positioned at the N-terminus, as in the case of nucleoside diphosphate kinase, NDK7. The second class has three DM10 domains occurring tandemly, as in the case of EFHC1, EFHC2 in humans, and RIB72, in *Chlamydomonas*. Notably, both classes of proteins with the DM10 domains have members that associate with microtubules – NDK7 directly binds microtubules (Ikeda, 2010). In view of these studies and our observations of EFHC2 localization to the spindle

apparatus, we tested for its association with α- and γ-tubulin, as well as with other cell division associated proteins.

In this chapter, I present, analysis of the contribution of *EFHC2* gene variants to JME. Towards characterizing the protein, I report the cellular localization of EFHC2 across different cell lines and its association with microtubules. Also, evidence to the contribution of the DM10 domains to the localization and function of EFHC2 are provided here.

2.2 Materials and Methods

2.2.1 Subjects

A set of 72 JME patients were ascertained from the National Institute of Mental Health and Neurosciences (NIMHANS), Bengaluru, and Sri Chitra Tirunal Institute for Medical Sciences and Technology (SCTIMST), Thiruvananthapuram. Inclusion criteria for JME were histories of occurrence of frequent, early morning myoclonic jerks without alteration of consciousness, often accompanied with generalized tonic-clonic seizures (GTCS) starting with bilateral myoclonic jerks, fast 4–6 Hz generalized spike-and-wave and polyspike-and-wave discharges on interictal EEG, and age-of-onset between 8-25 years. Patients manifesting partial seizures of any form, atonic/astatic and tonic jerks, stimulus-induced jerks only, and those with family histories of progressive myoclonic epilepsy (PME) were not included in the study. These criteria were chosen as per the current consensus on JME diagnosis and management (Kasteleijn-Nolst Trenite et al., 2013). For controls, 480 healthy individuals without any history of neurological conditions phenotype were examined. Genomic DNA was extracted from the peripheral blood samples of the participants using the phenol-chloroform method. This study had the approval of

Institutional Ethics Committees of the NIMHANS and JNCASR. All individuals –patients and –controls provided written informed consent.

2.2.2 Mutational analysis of *EFHC2*

Complete *EFHC2* transcript structure, including its 15 exons, exon-intron boundaries, and 5'-UTR and 3'-UTR sequences, as obtained from the Human Genome Data Viewer database at National Center for Biotechnology Information (NCBI), National Institute of Health (NIH), USA, and Ensembl Genome Browser, Ensembl Release 75, were examined. For information about the primer pairs used for PCR amplification, please see Table 1, Appendix 2. Amplification was carried out under the following conditions: initial denaturation at 94 °C for 5 minutes, followed by 40 cycles of denaturation at 94 °C for 30 seconds, annealing at 55-60 °C for 30 seconds and extension at 72 °C for 30 seconds, followed by a final extension at 72 °C for 10 minutes. PCR reactions of 20 µl-volume containing 100 ng of genomic DNA, 5 pmol of each forward and reverse primers, 200 µM of each dNTP, 1.5 mM MgCl2, buffer, and 1U of Taq DNA polymerase (New England Bio Labs, Massachusetts, USA). Amplified products were checked using agarose gel electrophoresis. Before sequencing, the PCR products were purified using MultiScreen-PCR96 filter plates. DNA sequencing was carried out by cycle sequencing using BigDyeTM Terminator Cycle Sequencing kit v3.0 (Applied Biosystems, Massachusetts, USA). The cycle sequencing was performed at following conditions: initial denaturation (96°C, 1 minute) followed by 25 cycles of denaturation (96°C, 10 seconds), annealing (50°C,5 seconds) and extension (60°C, 4 minutes). After cycle sequencing, the products were purified by ethanol precipitation. The pellet was dissolved in 10 µl of Hi-Di formamide (Applied Biosystems), denatured at 95°C for 5 minutes, and sequenced on an automated 48-capillary sequencer (ABI 3730

Genetic analyzer, Applied Biosystems). Sequences were analyzed on an ABI3100 Genetic Analyzer (Applied Biosystems). The sequences thus obtained were compared with the gene sequence in the GenBank database (GenBank database at NCBI, NIH, USA), using SeqMan5.01 (DNASTAR, Madison, WI). Variants obtained were examined in the unaffected, population-matched control individuals to estimate their allele frequencies.

2.2.3 Bioinformatic analysis of the variants identified

The complete nucleotide sequence of the *EFHC2* gene is available (NCBI accession number, NT_079573.5). The nucleotide sequence corresponding to accession number NT_079573.5 encompasses 197.37 kb and contains the reverse complement of the *EFHC2* coding sequence from nucleotides 44146303 to 44343677. Potential pathogenic effects of the variants were examined using SIFT (P. Kumar et al., 2009), PolyPhen2 (Adzhubei et al., 2010), and Mutation Taster (Schwarz et al., 2010) and using VEP annotation. In addition to examining control individuals from Southern India, allele frequencies of the variants were checked in the dbSNP, 1000 Genomes (Genomes Project et al., 2012) and exome variant server (NHLBI Exome Sequencing Project (ESP), Seattle, WA, December 15, 2012) databases.

2.2.4 Clones and constructs

Full-length cDNA of *EFHC2* was cloned into pEGFP-C2 and p3X-FLAG-CMV-10 vectors. Using pcDNA-3.1-EFHC2 clone as a template, the cDNA was amplified with primer pairs 5'-ACTCTAAGATCTATGGCCCTGCCTCTGCTGCCG-3' and 5'-GGTACCGTCGACTTATTCCTCCTCTAAGCCAAA-3', harboring BglII and SalI. The amplified, digested, and purified cDNA was ligated into pEGFP-C2- backbone as a BglII-SalI fragment. The sequence of

the fragment and the frame of integration were confirmed by using Sanger sequencing with cDNA-specific primers (Table 2, Appendix 2). For cloning of the cDNA insert into p3X-FLAG-CMV-10, the primer pairs, with 5'- CGGGGTACCGCCACCATGGCCCTGCCTCTGCTG-3' and 5'-CGCGGATCCTTATTCCTCCTCTAAGCCAAAC-3', with as KpnI and BamHI restriction sites were used. The sequences of the insert and the reading frame were validated by Sanger sequencing. Using pEGFP-C2- *EFHC1* full-length clone as the template (Raju, 2014), cDNA of EFHC1 was amplified using primer pairs 5'-CTTGCGGCCGCGATGGTGTCCAATCCCGTGCAT-3' and 5'-CGCGATGAATTCTCAGTTTGAGAAAGCACGAAC-3' as NotI and EcoRI fragment.

Site-directed mutagenesis was used to generate domain truncation alleles. The cDNA construct with mutations incorporated were generated by site-directed mutagenesis, carried out using QuikChange® II XL mutagenesis reagents (Stratagene, San Diego, California, USA). The mutagenic oligonucleotide primer pairs mentioned were designed to incorporate point mutations in the cDNA (Table 3, Appendix 2).

2.2.5 Cell culture and synchronization

HEK293T cells lines were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% heat-inactivated fetal bovine serum, 2 mM L-glutamine, 100 U/ml penicillin, and 0.1 mg/ml streptomycin in a humidified atmosphere of 5% CO2 at 37°C. For whole-cell lysate preparations, cells were cultured in 60mm dishes. For imaging, cells were seeded on poly-L Lysine coated glass coverslips in 35mm dishes or in 6-well flat bottom dishes. To synchronize cells, untransfected and transfected HEK293T cells were grown to 60% confluence and incubated in culture medium containing 2 mM thymidine (Sigma) for 14 hours. After two washes

in plain DMEM, cells were cultured further for 9 hours 10% complete DMEM containing 24 μM deoxycytidine (Sigma) to achieve mitotic enrichment.

HEK293T cells at 60-70% confluence were transfected as follows: An hour before transfection, 10% complete DMEM was replaced with DMEM with 2 mM L-glutamine. Transfections were carried out using Lipofectamine®2000 (Invitrogen, Thermo Scientific, Massachusetts, USA) at plasmid DNA: Lipofectamine ratio of 1:1.5. After 5-6 hours, transfection media was replaced by 10% complete DMEM. Cells were harvested for lysate preparation or fixed for imaging after 24-48 hours of transfection.

2.2.6 Immunocytochemistry

Coverslips with HEK293T cells, transfected with plasmid constructs, and untransfected controls were washed three times in 1X PBS (Phosphate Buffer Saline). Cells were fixed in 2% PFA for 15 minutes at room temperature or in ice-cold methanol for 7-15 minutes at -20°C. Cells were permeabilized in 0.1% Triton X-100, followed by blocking for one hour with 3% BSA (Sigma) in 1X PBS. Cells were incubated with primary antibodies for one hour. Primary antibodies used in these experiments were (i) mouse monoclonal anti-α–tubulin antibody (1:5000, Sigma) and (ii) Mouse monoclonal anti-FLAG antibody (1:2000, Flag-M2, Sigma) (iii) mouse monoclonal anti-γ-tubulin antibody (1:5000, Sigma). Cells were then washed with 1X PBS and incubated for one hour with 1:500 dilution of secondary antibodies conjugated to Alexa 468 or Alexa 568 (Molecular Probes, Oregon, USA), followed by DAPI staining and mounting. Images were captured using Zeiss LSM 880 confocal microscope and processed using ImageJ (NIH, USA) and Adobe® Photoshop® software.

2.2.7 Western blotting

Whole-cell lysates of HEK 293T cells were prepared in RIPA buffer (150mM NaCl, 10mM Tris pH 7.5, 0.1% SDS, 1% Triton X-100, 1% sodium deoxycholate, 5mM EDTA) with Protease Inhibitor Cocktail (Sigma). Adherent cells in culture dishes were washed twice in 1X PBS, incubated with RIPA buffer for 30 minutes on ice with intermittent scraping. The suspension was collected and homogenized by passing through 1ml, 26G syringe. This suspension was centrifuged at 13000 RPM at 4°C for 30 minutes. The supernatant was collected as whole cell lysate.

Total protein concentration in the lysates was estimated using the bicinchoninic acid assay (BCA) reagents (Sigma). For Western analysis, aliquots of 50- 70 µg of protein were mixed with SDS gel-loading buffer, boiled for 5 minutes, and then resolved electrophoretically on a 10% or 8% SDS-polyacrylamide gel. The proteins on the gel were subsequently electrotransferred (Bio-Rad, California, USA) to equilibrated nitrocellulose membrane (Amersham Protran, GE or BioTrace™ NT, Pall life sciences) at constant 400 mA for one hour and fifteen minutes. Transfer was confirmed by Ponceau S staining followed by washes in distilled water. Blocking was performed using 3% Skimmed milk powder (Sigma) and 2% Bovine Serum Albumin (BSA, Sigma) in 1X PBS with 0.05% Tween 20, for 12 hours at 4°C. The membrane was probed with a primary antibody for 10 hours at 4°C. The primary antibodies used in this study were anti-GFP antibody, A6455 at 1:3000 in 0.05% PBSTween-20 (Thermo Fisher Scientific, MA, USA); anti-Flag antibody 1:3000 in 1% BSA in PBS (Flag-M2, Sigma), anti-alpha tubulin antibody, and anti-gamma tubulin at 1:5000 (Sigma). Post incubation with primary antibody, the membrane was washed in 1X PBS with 0.05% Tween-20 at room temperature for 5 minutes. The blot was further incubated for one hour at room temperature with a

1:5000 dilution of horseradish peroxidase (HRP)-conjugated, goat anti-rabbit IgG (Bangalore Genei, Bangalore, India) or goat anti-mouse IgG (Bangalore Genei, Bangalore, India) in 1X PBS containing 1% BSA, 1% skimmed milk powder, and 0.05% Tween 20. The membrane was washed twice for 10 minutes each at room temperature in 1X PBS and 0.05% Tween 20. The protein bands were detected using an enhanced chemiluminescent substrate for HRP (Pierce-Thermo Fischer Scientific, USA, or Clarity™ Western ECL Substrate - Bio-Rad).

2.2.8 Co-immunoprecipitation

HEK293T cells were transfected with p3X-FLAG-EFHC1, p3X-FLAG-EFHC2 or co-transfected with pCDNA-EFHC1; and with p3X-FLAG-EFHC2, pCDNA-EFHC2 and p3X-FLAG-EFHC1. For the analysis of the interaction of domain- truncation alleles, p3X-FLAG-EFHC2, and its variants were transfected in HEK293T cells. To test the interaction of EFHC2 with other cell division-associated proteins, co-transfections of pEYFP-PLK1 (Addgene, ID no. 39843) and pEGFP-N1-CDC20B and pEGFP-N1-Centrin2 (Pooja Barak, 2021, PhD Thesis) constructs were also used. 36-48 hours after transfection, lysates were prepared in the following immunoprecipitation (IP) lysis buffer – 50mM Tris at pH-7.4, 150mM- NaCl, 1% Triton-X 100, 10mM NaF, 1mM Na_3VO_4, PIC ultra-complete (Roche, Sigma). The lysates were clarified at 16000g at 4°C for 30 minutes. Supernatants were collected as lysate for IP. Dynabeads (Protein-G, Thermo Fisher Scientific) were washed three times in PBS, incubated with 1µg of the antibody or IgG isotype control. Antibody binding was done at room temperature for 10 minutes on an end-to-end rotor. Antibody-bound beads were incubated with corresponding lysates, incubated on an end-to-end rotor at room temperature for 30 minutes. Supernatants were collected and stored. The beads were washed two times with 0.05% PBST for 2 minutes each on the

end-to-end rotor. Each wash was collected and stored. For the final wash step, wash buffer was added, and beads were carefully transferred to a fresh tube, the supernatant was collected and eluted with 20µl of elution buffer, or 20µl of 2X-SDS loading dye without β-ME or with 6X-loading dye with β-ME. The elute was collected by boiling the beads directly in the dye. 5-10% of the inputs, washes, and elutes were loaded on a 10% SDS PAGE for Western blotting.

2.2.9 *In utero* electroporation

EFHC2 wildtype and EFHC2 variants p.Arg246His and p.Arg486Cys were cloned in pEGFP-C2 vector. Endotoxin-free plasmids were generated using Qiagen (Hilden, Germany) Endofree®plasmid purification kit following the manufacturer's protocol. 1-2µg of the plasmid was mixed with 0.1% Fast Green solution. E14.5-15.5 pregnant mice were anesthetized with isoflurane (1.5% for induction and 1-1.5% during surgery. 2-3µl of the DNA solution was injected through the uterine wall into one of the lateral ventricles (hippocampus and cortex) using a glass capillary tube. After soaking the uterine horn with a phosphate-buffered saline (PBS) solution, the embryo's head was carefully held between tweezer-type platinum electrodes (5mm diameter), and electroporation was performed using ECM 830, BTX, Harvard Apparatus (Holliston, Massachusetts, USA). 30-48 hours after electroporation, the embryonic brains were harvested and coronally sectioned and stained with anti-GFP antibody (1:1000, A6455, Invitrogen), anti-ki67 antibody (1:100), and DAPI. Sections were imaged at 20X and 63X. Localization, migration, and filament length analysis were studied using FIJI image analysis software.

2.2.10 Statistical analysis

All the statistical comparisons were made using GraphPad Prism5. Statistical tests performed were one-way ANOVA followed by Dunnett's test for multiple comparisons or the Kruskal-Wallis test. Results are shown as mean ± standard error of the mean (SEM). Differences between groups were considered statistically significant for $P \leq 0.05$.

2.3 Results

2.3.1 *EFHC2* variants among JME patients

Among the seventy-two JME patients examined, we report the occurrence of three missense variants p.Arg135Gln (R135Q), p.Ser430Tyr (S430Y), and p.Glu506Lys (E506K) (Table 1). p.Arg135Gln, p.Ser430Tyr variants have been found commonly among patients and control populations with minor allele frequencies greater than the operational cut-off 0.005. For the representative electropherograms confirming the variants, please see Figure 1. The variant p.Arg135Gln occurs at a conserved amino acid residue within the first DM10 domain while serine at 430 and glutamic acid at 506 are less conserved (Figure 2). While p.Glu506Lys is relatively rare in the patient and control cohort sequenced in this study, it occurs at a minor allele frequency of 0.0084 among the South Asian (Thakran et al., 2020) population as reported in 1000genomes: phase 3 database. This frequency is greater than cut-off 0.005. Thus, no novel, rare, missense variants were identified in the 72 JME patients sequenced in this study. Three intronic changes at genomic locations - X:44272669 G>A, X:44035673 G>A and X:44176415 T>A were identified (Table 2).

In a total of 550 JME patients from south India, seven potentially pathogenic variants were identified (Table 3). Of these variants,

p.Arg135Gln occurs commonly in patients as well as controls. However, this variant leads to significant chromosome segregation defects and abnormal mitotic spindles (Raju, 2014).

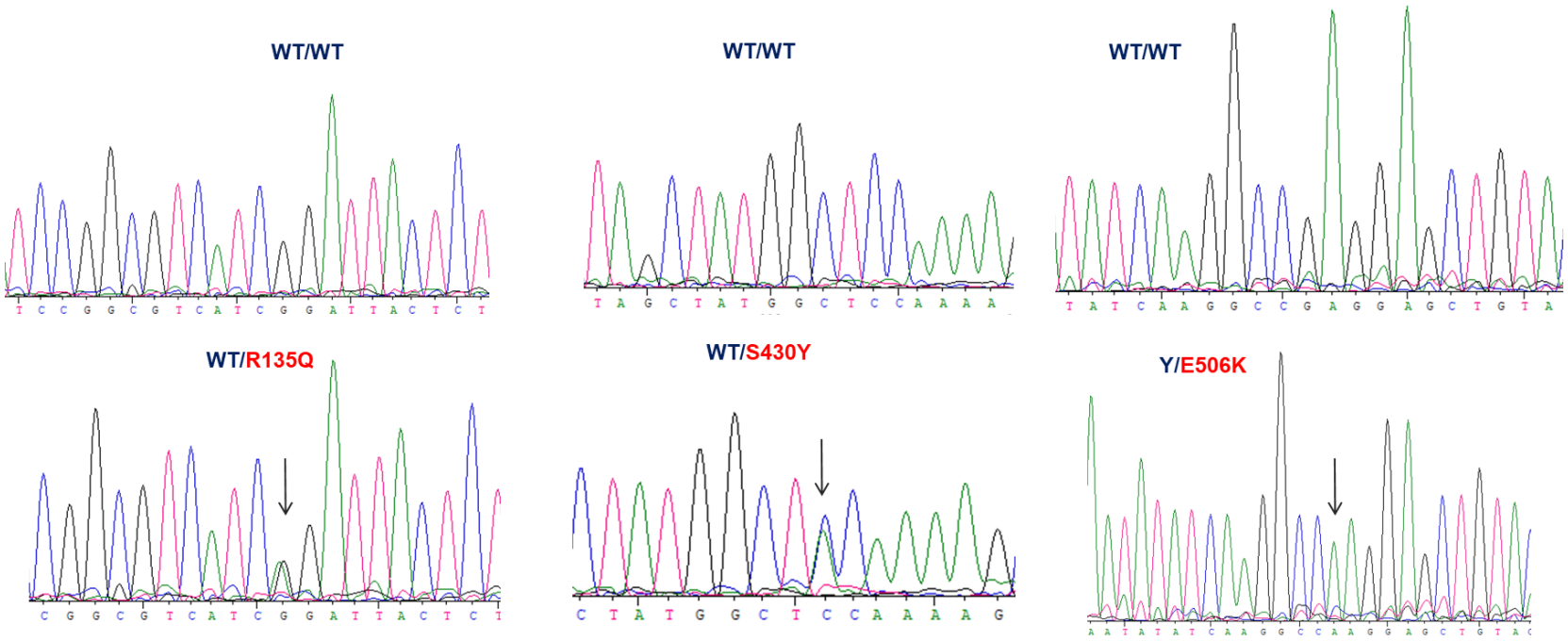


Figure 1: Electropherograms of wildtype EFHC2 and variants c.404 g>a (p.Arg135Gln), c.1289C>A (p.Ser430Tyr), and c.1516G>A (p.Glu506Lys).

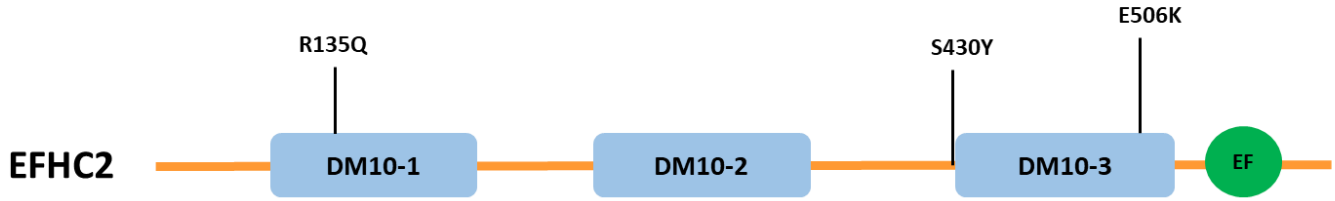
Nucleotide change	Amino Acid change	Occurrence in patients (n= 72)	MAF in control chromosomes (n=568)	MAF in 1000G, ExAC, or dbSNP 153	SNP ID	SIFT score	Polyphen 2 Hum Var score	PROVEAN Score	Mutation Taster crobability value	Mutation Assessor pred and MAori	FATHMM
c. 404 g>a	p. Arg135Gln	5	0.017	1000G (0.0029) ExAC (0.0028) dbSNP (0.00159)	rs201560745	Deleterious (0.03)	Probably damaging (0.91)	Deleterious (-3.4)	Disease causing (1.0)	Medium (2.59)	Tolerated (0.27)
c.1289C>A	p. Ser430Tyr	7	0.08	1000G (0.08) ExAC (0.08) dbSNP (0.08)	rs2208592	Tolerated (1.0)	Benign (0.0)	Neutral (-0.33)	Polymorphism	Neutral (0.08)	Tolerated (-0.30)
c. 1516G>A	p. Glu506Lys	1	0.001	1000G (0.0023) ExAC (0.0037) dbSNP (0.005)	rs192827775	Deleterious (0.04)	Benign (0.268)	Neutral (-1.81)	Disease causing (0.99)	Low (1.35)	Tolerated (-0.26)

Table 1: Missense variants identified among 72 JME patients.

Genomic position GRCh38	Nucleotide change	Consequence	Occurrence in patients (n= 72)	MAF in 1000G, ExAC or dbSNP 153	SNP ID
X: 44176428	C>T	c.2034-44	12	0.09	rs7887384
X:44272669	G>A	c.466+17	3	Not reported	-
X:44176416	T>A	c.2034-32	6	Not reported	-

Table 2: Intronic variants among 72 JME patients.

a)



b)

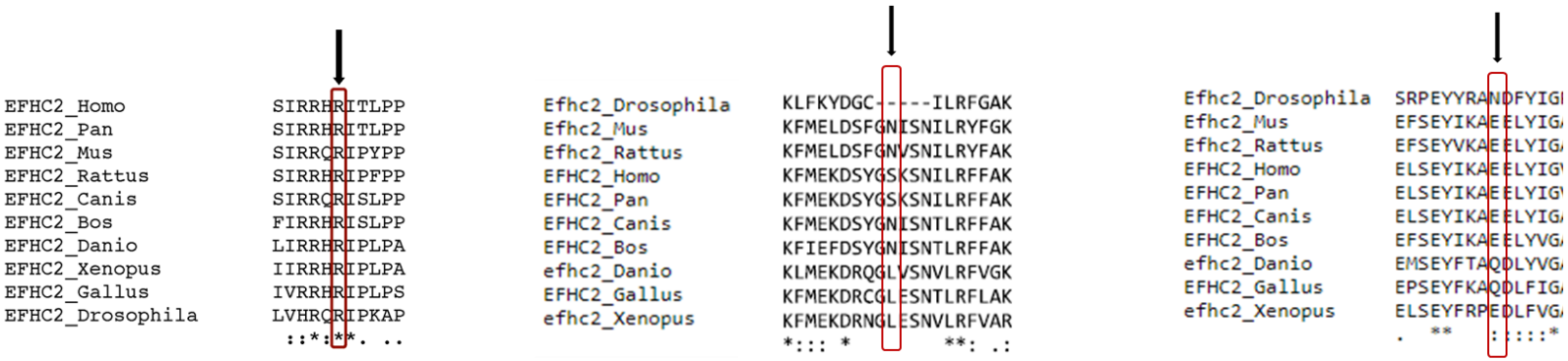


Figure 2: **(a)** EFHC2 domain architecture with the position of variants identified. **(b)** Alignment showing the location and conservation of amino acid residues, p.Arg135Gln, p.Ser430Tyr and Glu506Lys

Nucleotide Change	Amino Acid Change	Occurrence in Patients (n= 550)	MAF in Control X-Chromosomes (n=568)	MAF in 1000G, ExAC or dbSNP 153	Conserv-ation Score[a]	SIFT[b]	Polyphen2 Hum Var[c]	PROVEAN Score[d]	Mutation Taster Probability Value[e]	Mutation Assessor pred and MAori[f]	FATHMM[g]
c. 404 g>a	p. Arg135Gln	18	0.017	1000G (0.0029) ExAC (0.0028) dbSNP (0.00159)	-1.07	Deleterious (0.03)	Probably Damaging (0.91)	Deleterious (-3.4)	Disease Causing (1.0)	Medium (2.59)	Tolerated (0.27)
c. 528 c>g	p. Asn176Lys	1	0	Absent	-0.128	Deleterious (0.05)	Benign (0.29)	Deleterious (-4.6)	Disease Causing (0.999)	Medium (2.4)	Tolerated (-0.28)
c. 587 a>t	p. Tyr196Phe	1	0	1000G (2.65E-04) ExAC (8.28E-06) dbSNP (0.0003)	-0.624	Deleterious (0)	Probably Damaging (0.99)	Deleterious (-3.5)	Disease Causing (0.999)	Medium (2.05)	Tolerated (-0.47)
c. 737 g>a	p. Arg246His	1	0	ExAC (2.3E-04) dbSNP (0.0008)	2.05	Tolerated (0.6)	Possibly Damaging (0.54)	Neutral (-1.7)	Disease Causing (0.501)	Medium (2.5)	Tolerated (-0.36)
c. 1156 a>g	p. Lys386Glu	1	0	ExAC (8.09E-06) dbSNP (0.00001)	0.55	Tolerated (0.11)	Benign (0.13)	Neutral (-2.4)	Disease Causing (0.999)	Medium (2.5)	Tolerated (-0.39)

c. 1456 c>t	p. Arg486Cys	1	0	ExAC (3.35E-05)	-0.633	Deleterious (0.02)	Probably Damaging (1.0)	Deleterious (-7.1)	Disease Causing (0.999)	High (3.76)	Tolerated (0.22)
c. 1828 c>t	p. Pro610Ser	1	0	Absent	0.913	Tolerated (0.5)	Benign (0.02)	Neutral (-1.3)	Polymorphism (0.999)	Medium (2.2)	Tolerated (-1.19)

Table 3: *EFHC2* mutations among JME patients

Abbreviations are as follows: dbSNP153 = database of single nucleotide polymorphism database; **1000G**= 1000 Genome project; **ExAC Browser** = The Exome Aggregation Consortium; **SIFT =** Sorting Intolerant From Tolerant; **FATHMM** = Functional Analysis through Hidden Markov Models; **MAF** = Minor allele frequency; **PROVEAN** = Protein Variation Effect Analyzer.

[a]ConSurf Server; normalized conservation score for the amino acids based on the phylogenetic relationship from 150 homologous sequences. [b]Damaging if score is smaller than 0.05. [c]Probably damaging if score is 0.909-1; possibly damaging 0.447-0.908; benign 0.0-0.446. [d]Variants with a score equal to or below -2.5 are considered deleterious, Variants with a score above -2.5 are considered neutral. [e]Probability of causing disease; values close to 1 indicate a high confidence of the prediction. [f]Score predicts functional impact of SNV. [g]SNV is damaging if score is <=-1.5, otherwise tolerated

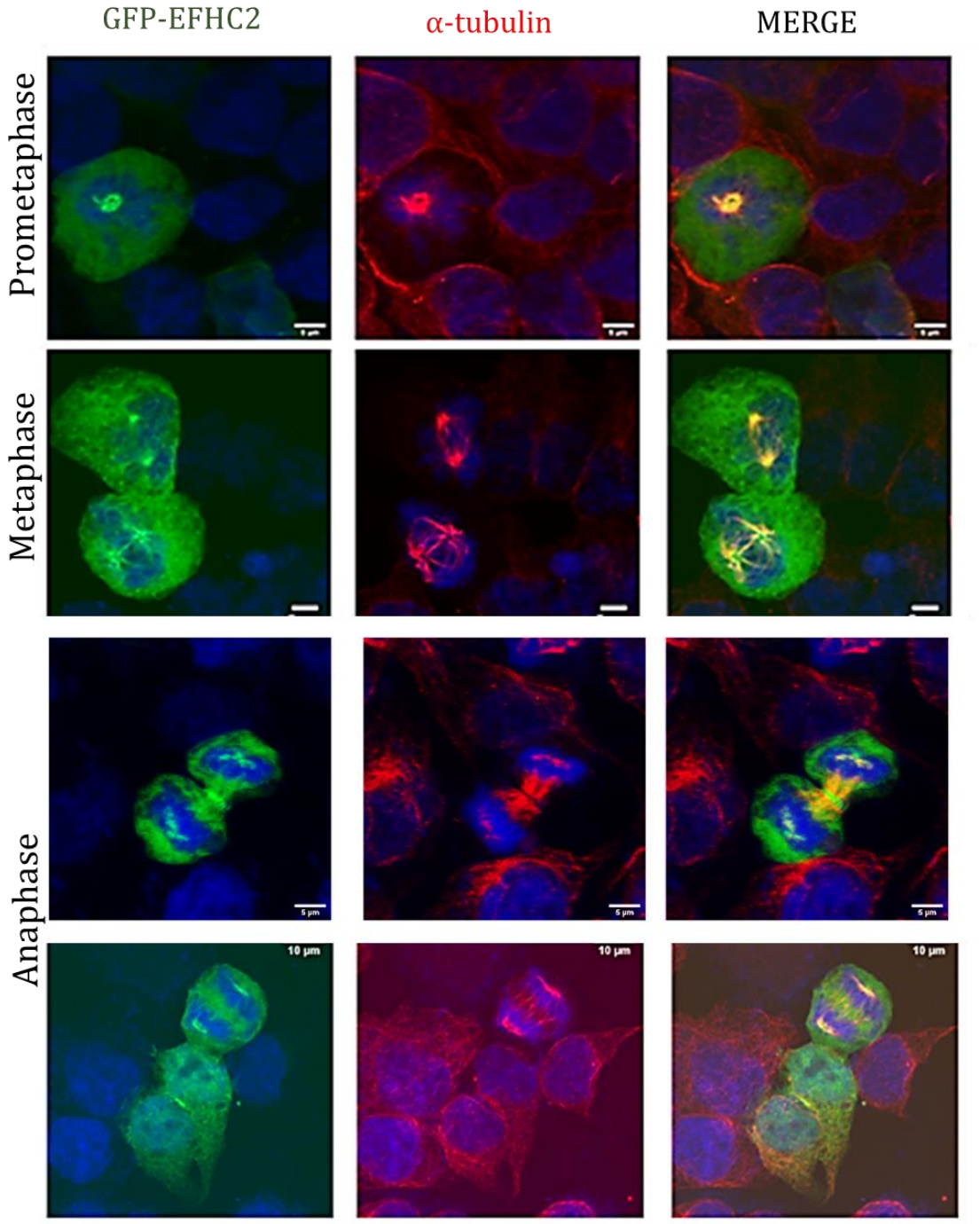
480 JME patients (Raju, 2014); 72 additional JME patients (this study).

2.3.2 EFHC2 localizes to the spindle apparatus during cell division

On staining of HEK293T cells with anti-EFHC2 antibody, we did not detect endogenous expression of EFHC2. However, the subcellular localization of the over-expressed protein was in the nucleus and centrosome. It changed location sequentially to spindle poles, mitotic spindle, and spindle midzone during mitosis, and finally transferred to midbody at cytokinesis. Our observations are in line with a report on EFHC2 to be a ciliary protein localizing to the nucleus and cytoplasm (McClure-Begley & Klymkowsky, 2017). During mitosis, the protein was observed at the spindle poles during prometaphase, metaphase, and anaphase. In a subpopulation of cells, the protein was present at inter-polar microtubules and astral microtubules. During cytokinesis, localization was observed at the midbody ring, intercellular bridge (Figure 2). EFHC2 staining overlapped with γ- tubulin at the centrosome, spindle poles, and midbody (Figure 3).

In U-87 MG cells, a glioblastoma cell line, and SH-SY5Y cells, an astrocytoma cell line, EFHC2 is endogenously expressed (Figure 4,5). Immunofluorescence analysis using anti-EFHC2 antibody stains the cytoplasm, nucleus, and centrosome. The pattern of expression was similar under native and over-expression conditions. The antibody staining pattern of EFHC2 was reproducible with tagged constructs of EFHC2 such as GFP-EFHC2 and FLAG-EFHC2. The localization of EFHC2 to centrosomes was further confirmed by co-staining using γ-tubulin. During cell division, we find that EFHC2 localized to the spindle pole and midbody. EFHC2 was not seen in the entire microtubule meshwork in the cytoplasm of the interphase cells. However, during prometaphase to telophase, EFHC2 was localized to mitotic spindle poles and spindle microtubules stained with α-tubulin.

During cytokinesis, it was colocalized with γ-tubulin in the midbody.



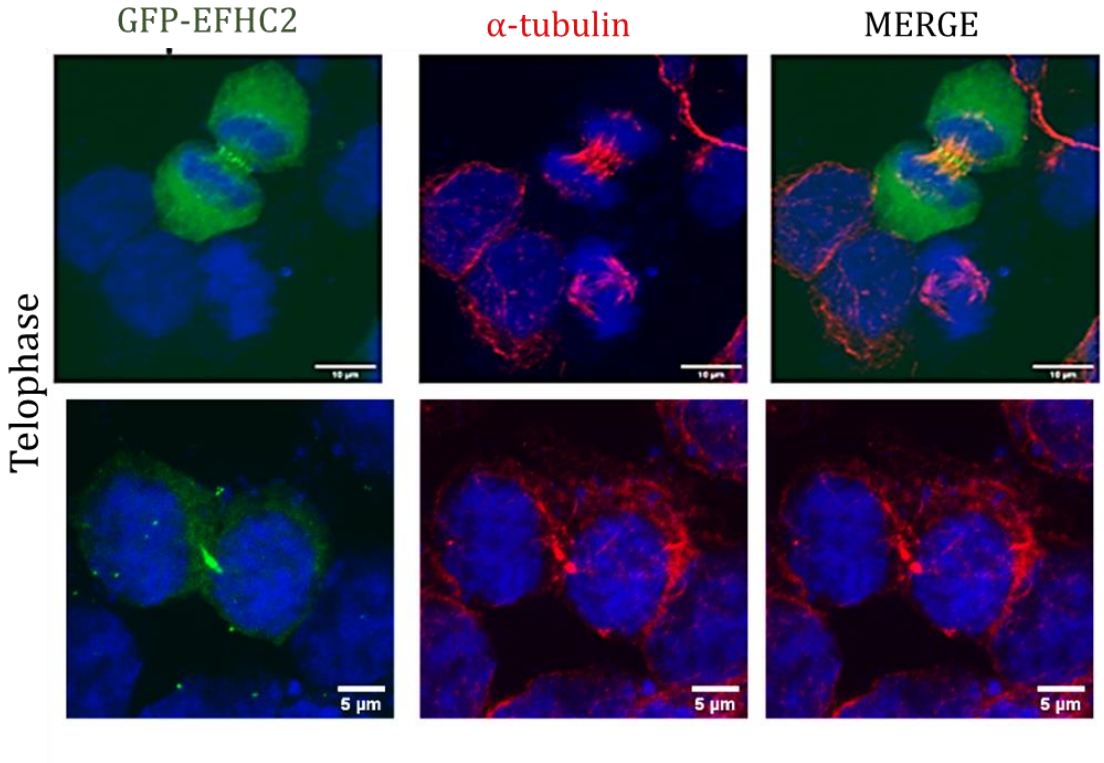


Figure 2: **Localization of EFHC2 during mitosis**. HEK293T cells expressing GFP-EFHC2 co-stained with α-tubulin. EFHC2 (green), α-tubulin (red) and nucleic acid (DAPI).

The M-phase specific localization of the protein and its association with mitotic spindles suggest the role of EFHC2 in the organization of mitotic spindles and cell division.

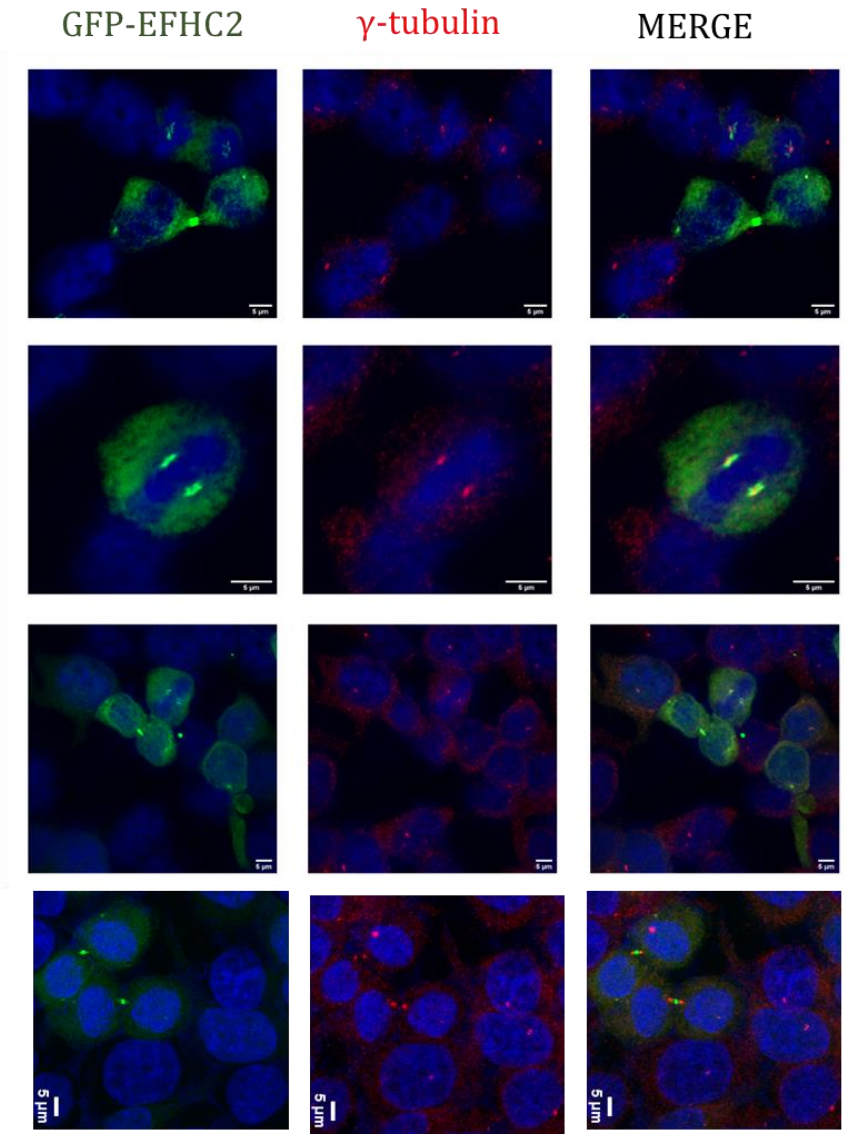


Figure 3: **Localization of EFHC2 with γ-tubulin**. HEK293T cells expressing GFP-EFHC2 co-stained with γ-tubulin. EFHC2 (green), γ-tubulin (red) and nucleic acid (DAPI).

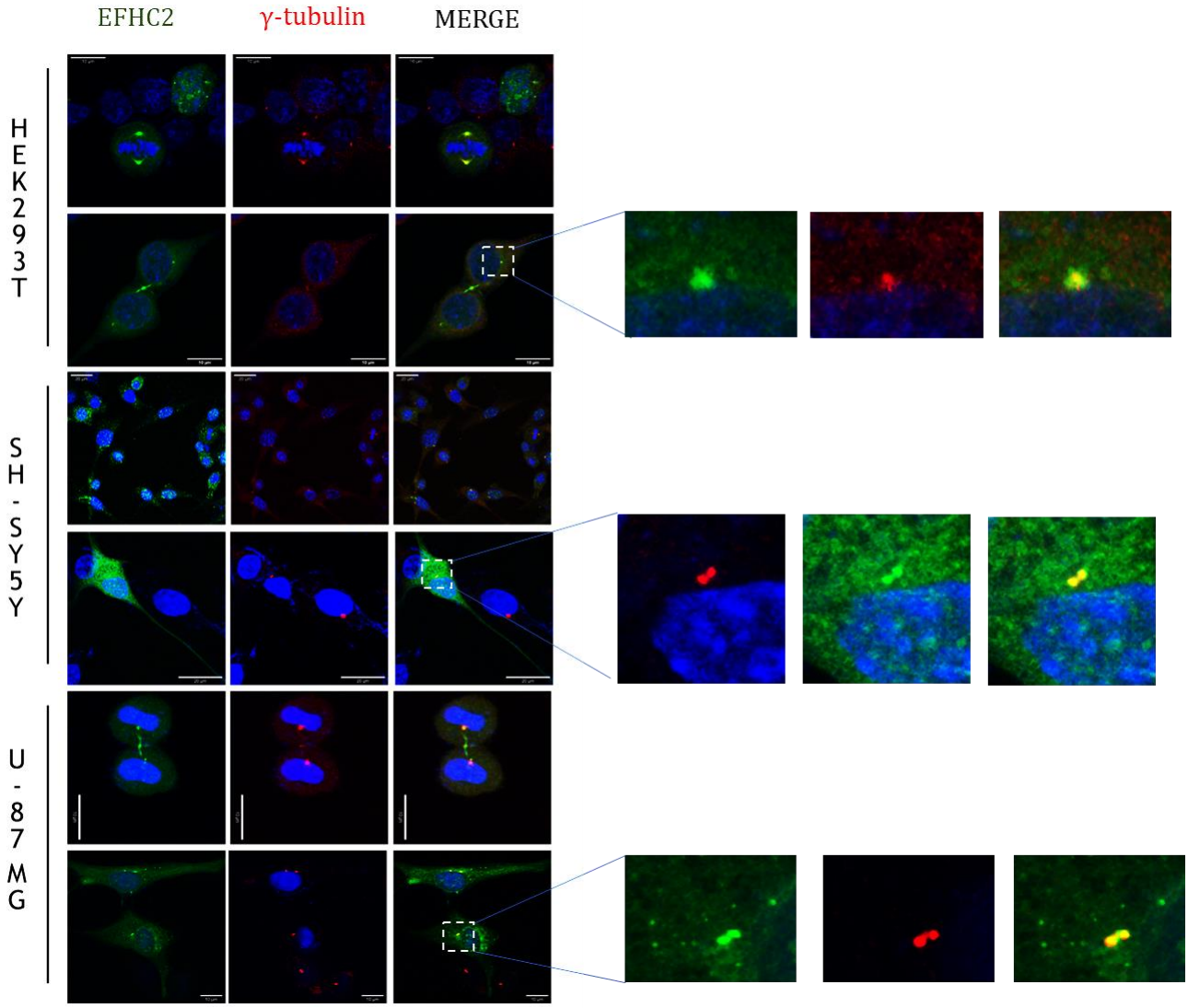


Figure 4: Expression of EFHC2 across cell lines. HEK293T, SH-SY5Y and U-87 MG stained for EFHC2 (green), gamma-tubulin (red) and nucleic acid (DAPI). EFHC2 localized to centrosomes and spindle poles and midbody in these cell lines.

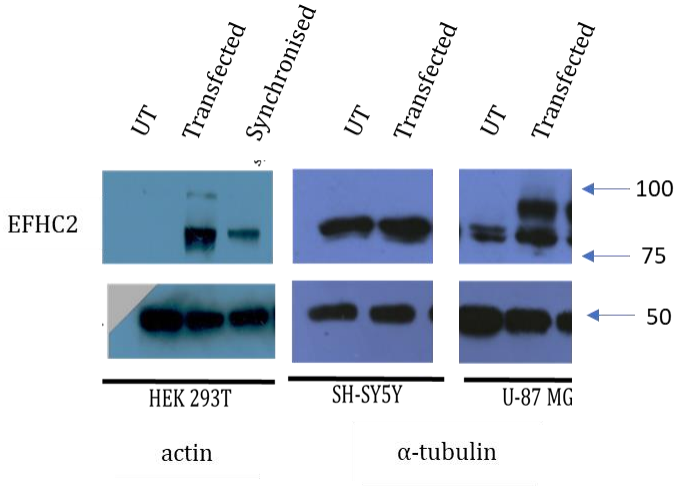


Figure 5: Western analysis of expression of EFHC2 across cell lines. HEK293T, SH-SY5Y and U-87 MG lysates were probed with anti-EFHC2 antibody. Endogenous expression observed in SH-SY5Y and U-87 MG cells

Data mining for EFHC2 supported our observations. Bioinformatic prediction tools such as eukaryotic linear motif resource (ELM) identified motifs in EFHC2 that overlap with proteins localizing to Golgi, endoplasmic reticulum, nucleus, cytosol, among others. Among these motifs, we found several peptide stretches that are observed in proteins associated with microtubules with roles in cell division. CiliaCarta database assigns EFHC2 to be a cilia-associated protein with a rank of 88 and a score of 5 (van Dam TJP, 2019). The localization pattern of EFHC2 during mitosis is like that of EFHC1 and CILK1, which are also ciliary proteins associated with JME.

2.3.3 EFHC2 is a microtubule-associated protein important in cell division

The localization pattern observed for EFHC2 during cell division overlapped with that of EFHC1, tubulins -α and -γ. This prompted us to check the interaction of EFHC2 with microtubules and other cell

division-associated proteins. In co-immunoprecipitation experiments, tubulins -α and -γ were found in the elute fractions suggesting their association and providing evidence of EFHC2 as a microtubule-associated protein (Figure 6).

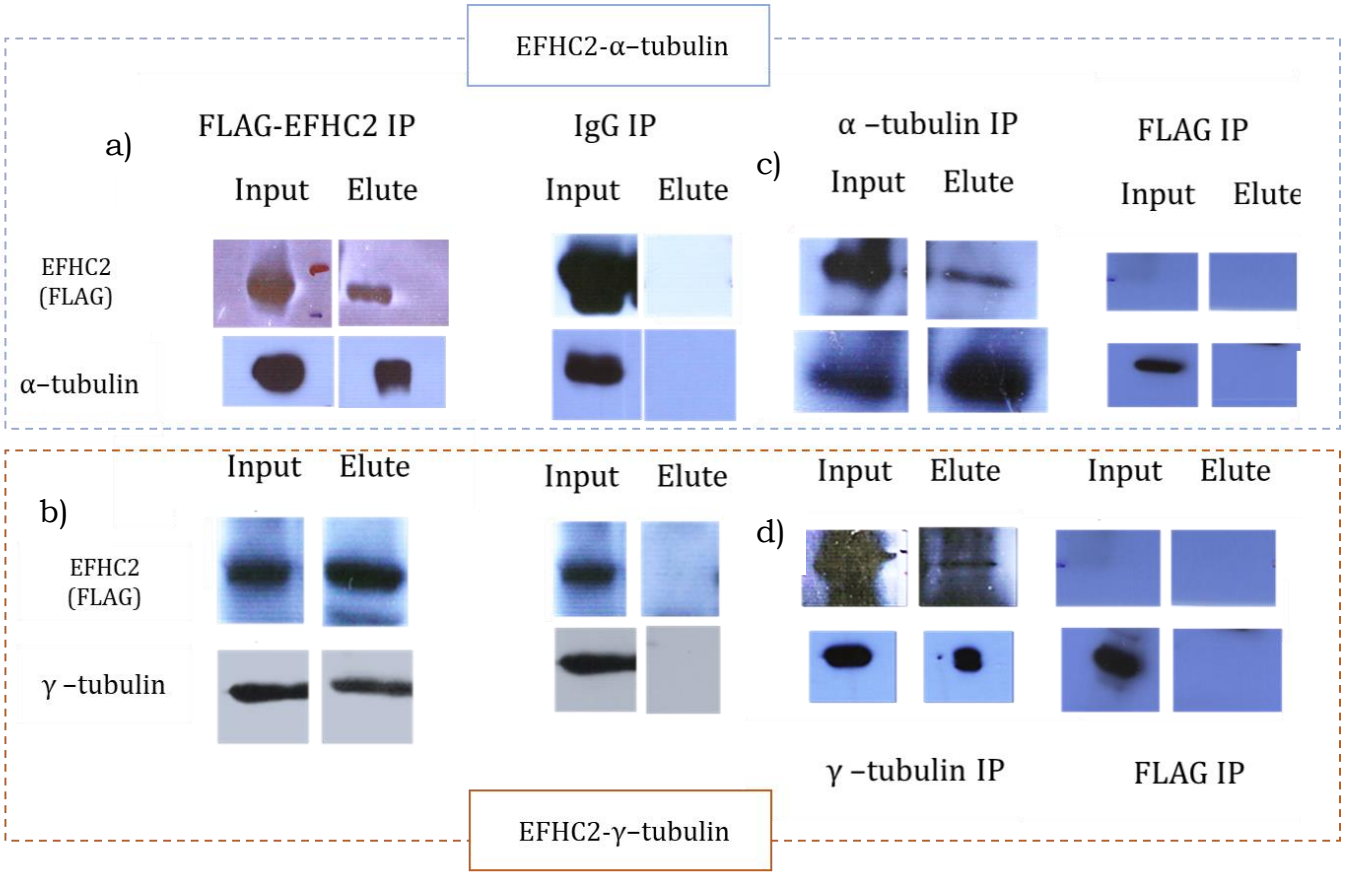


Figure 6: Co-immunoprecipitation of EFHC2 and tubulins –α and –γ. **(a)** FLAG-tagged EFHC2 pulldown, probed with anti-FLAG and anti-α-tubulin shows that these two proteins co-immunoprecipitate. **(b)** FLAG-tagged EFHC2 pulldown, probed with anti-FLAG and anti-γ-tubulin shows that these two proteins co-immunoprecipitate. **(c)** IP with α-tubulin shows coprecipitation of EFHC2 and **(d)** γ-tubulin coprecipitates with EFHC2

To gain initial insights into the role of EFHC2 as a cell division protein, the interaction of EFHC2 was tested with Centrin2, Polo- Like Kinase1 (PLK1), and EFHC1.

Polo Like Kinase 1, PLK1 – is a member of the polo-like family of serine/threonine kinases. PLK1 is important for cell division. It regulates spindle formation, spindle checkpoint initiation. During cell division, it localizes to the mitotic centrosomes, kinetochores, and midbody. These localizations allow PLK1 to phosphorylate specific downstream targets to regulate mitosis (Colicino & Hehnly, 2018). It governs the recruitment of γ-tubulin and pericentriolar material on centrosomes, and PLK1 is enriched in the midbody, where it ensures faithful abscission by recruiting midbody components in an orderly manner and through the phosphorylation of substrates (Colicino & Hehnly, 2018). EFHC2 is also localized to the midbody. During cytokinesis, in a population of cells transiently transfected with EFHC2, we observed the protein to be distributed at the intercellular bridge, cleavage furrow, and in the midbody ring. We thus wondered if PLK1 may sequester EFHC2 to the spindle and midbody and if they physically interact. We evaluated the interaction on EFHC2 and PLK1 *in vitro* using immuno-pull downs and observed that EFHC2 eluted with PLK1 (Figure 7a).

EFHC1 and EFHC2 have similar subcellular localization and protein domain architecture. EFHC1 is reported to bind to α-tubulin. In high throughput two-hybrid interaction studies, EFHC1 is reported to be an interactor of EFHC2 (**https://thebiogrid.org/123203/summary/homo-sapiens/efhc2.html**). We tested for the interaction between EFHC1 and EFHC2 using immuno-pulldown assay. Under the experimental conditions, we did not detect their interaction (Figure 7c)

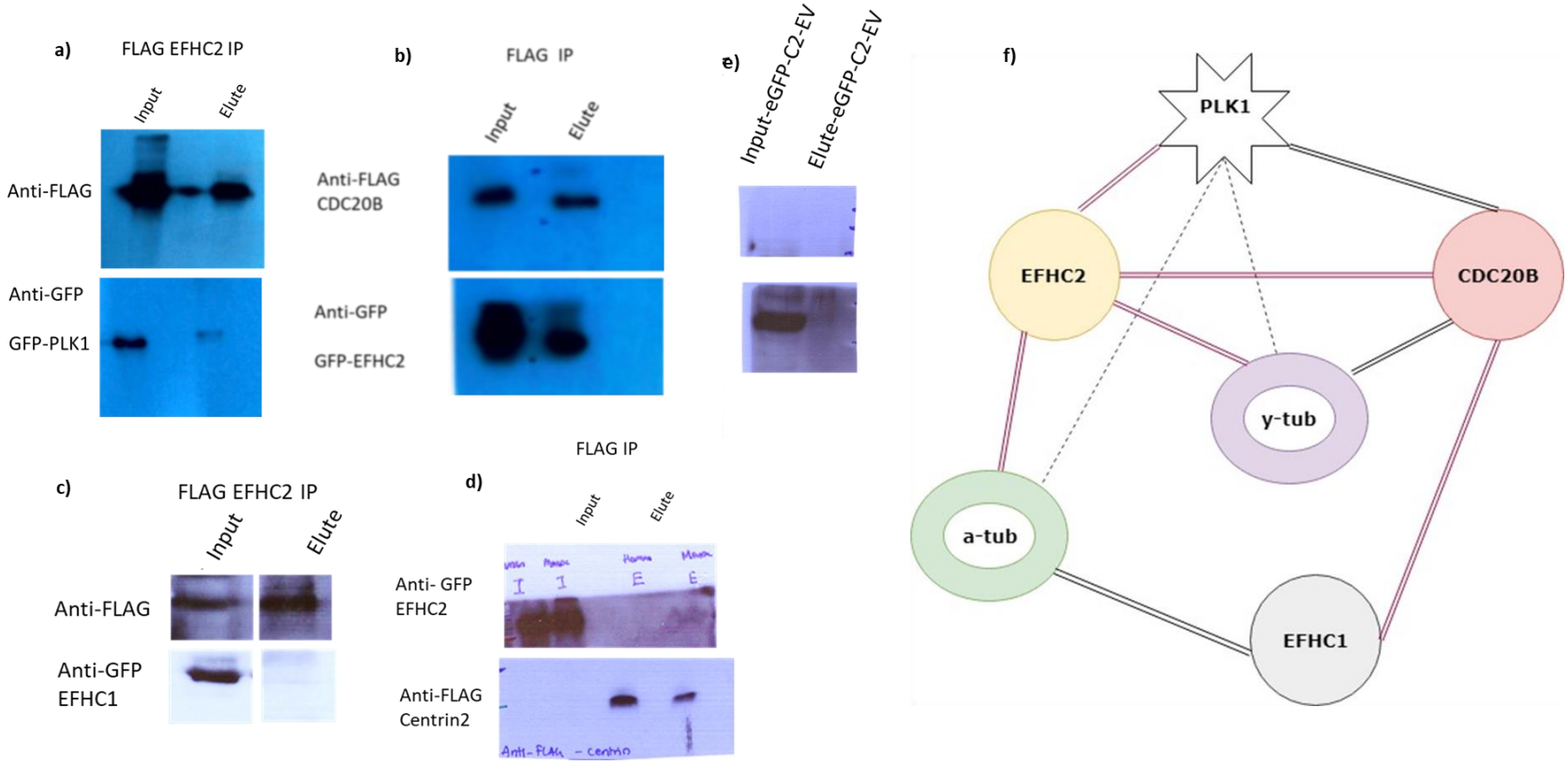


Figure 7: Co-immunoprecipitation of EFHC2 and cell cycle proteins. **(a)** Flag-EFHC2 and PLK1. **(b)** GFP-EFHC2 and CDC20B. **(c)** Flag-EFHC2 and EFHC1. **(d)** GFP-EFHC2 and Centrin2. **(e)** GFP-vector control pull down and **(f)** illustrative representation of the interaction identified.

Centrin2 is encoded by the gene *CENT2*. Centrin2 is a member of the small calmodulin-like family of proteins with EF-hands and binds to calcium. It is a ubiquitously expressed protein that localizes to the centrosome during interphase with important roles in centriole duplication. Knockdown of Centrin2 leads to mitotic spindle abnormalities (Salisbury et al., 2002). The subcellular localization and effect on Centrin2 mutations were like our observations for EFHC2. We thus tested for the interaction of EFHC2 and Centrin2. On immunoprecipitation of EFHC2, we did not detect Centrin2 in the elute fraction. This could be a consequence of lack of physical interaction, or a transient interaction that was not captured in IP (Figure 7d).

2.3.4 DM10 domain-mediated localization of EFHC2

We find the localization of EFHC2 to mitotic spindle from prometaphase through to cytokinesis. Using IP experiments, we find that EFHC2 is a microtubule associated proteins. To identify elements in the protein that facilitate the localization and function of the protein, we generated several synthetic mutants. GFP–tagged truncations and subclones were generated. The expression of the mutant proteins was validated through Western analysis (Figure 8a and 8b). In immunofluorescence assays the following localization was observed. During interphase, the constructs localized largely to the cell cytoplasm and in a sub-population of cells, bright punctate cytoplasmic distribution was observed (Figure 8a and 8b). In cells expressing moderate to low levels of GFP fusion proteins, spindle staining superimposing the signals of alpha (α)- tubulin. Differences in localization was observed during mitotic stages – GFP-EFHC2-Gly8Ter (G8X) and GFP-EFHC2-Ser70Ter (S70X) did not localize to spindle poles or mid body (Figure 8a and 8b).

Similar pattern of localization was observed for the subclone consisting of the C-terminal EF-hand.

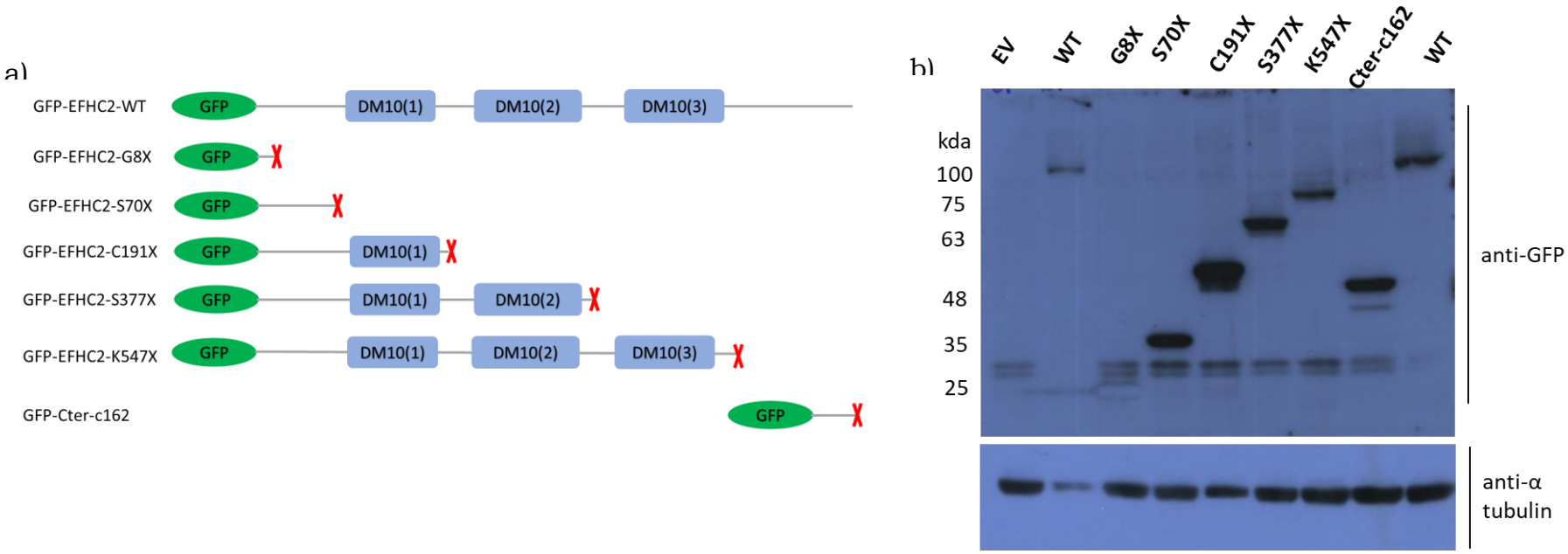


Figure 8: Generation and confirmation of EFHC2 domain truncation mutations. **(a)** A schematic of the domain truncations used in the study and **(b)** confirmation of domain truncation mutant proteins using Western analysis

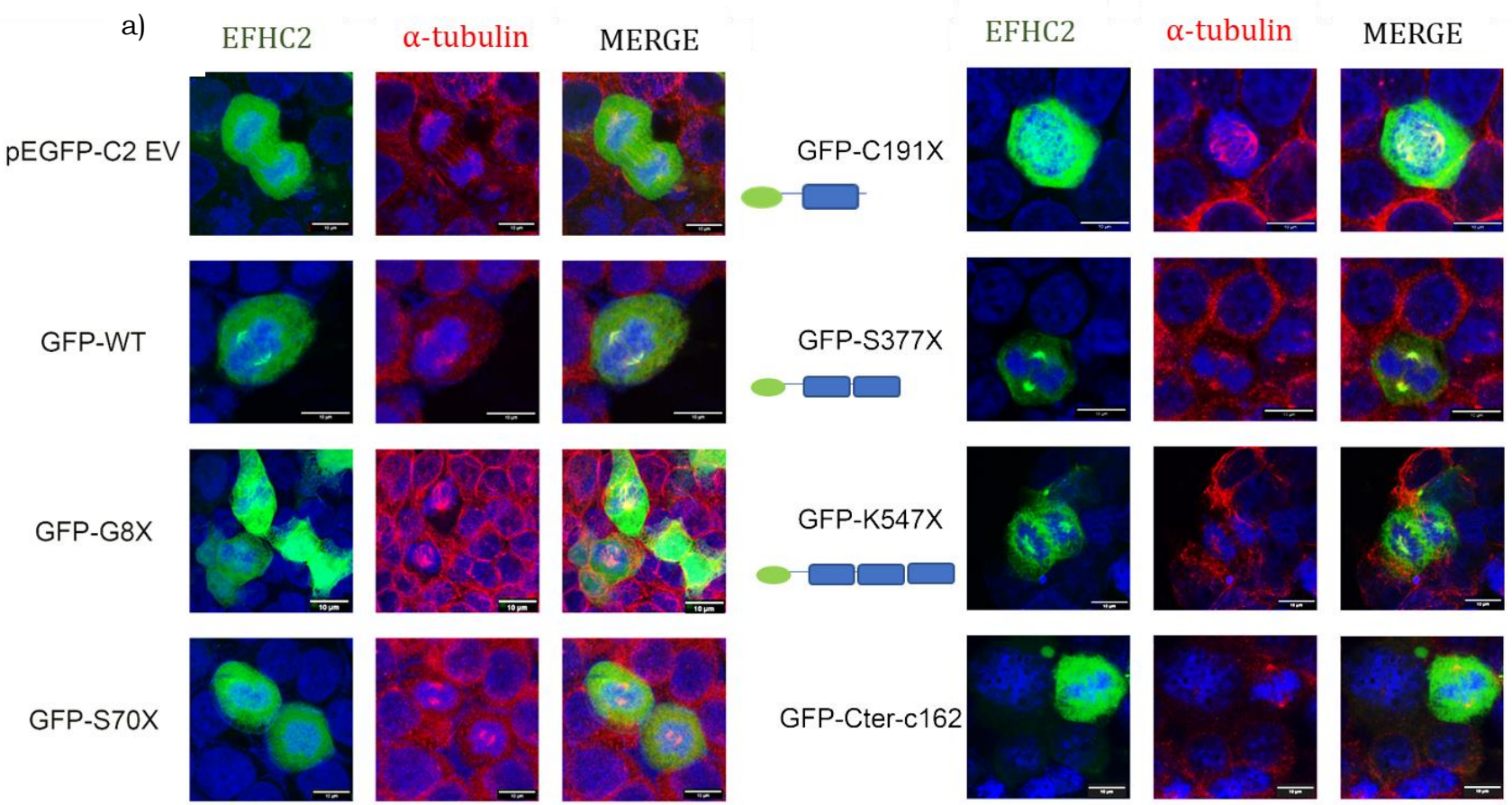


Figure 9; (a) EFHC2 domain-truncation alleles and their sub-cellular localization. HEK293T cells expressing GFP-EFHC2 wildtype and mutant constructs co-stained with α-tubulin (red).

b)

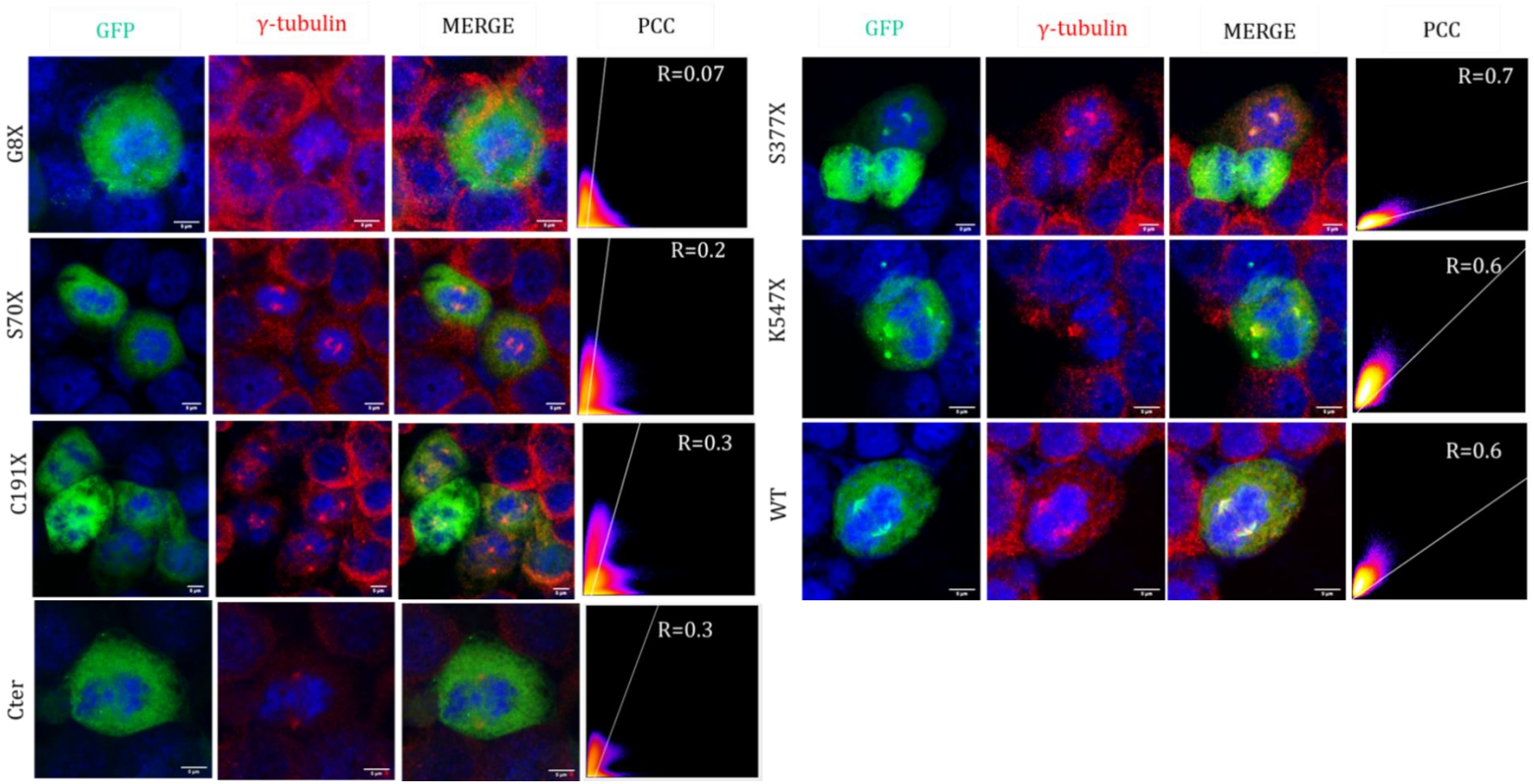
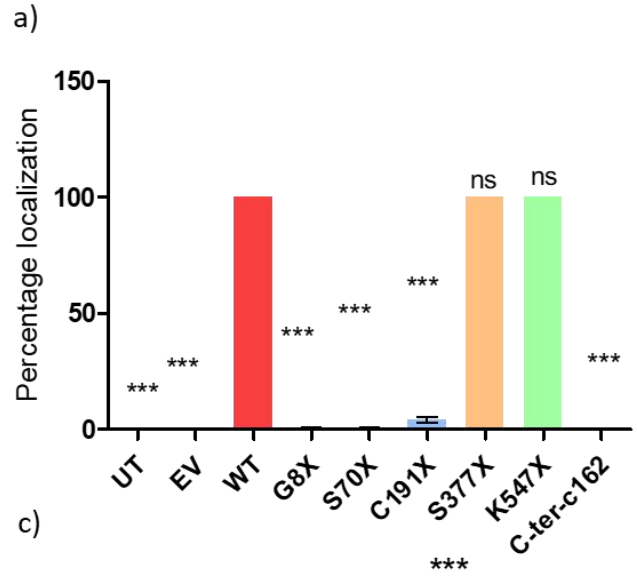
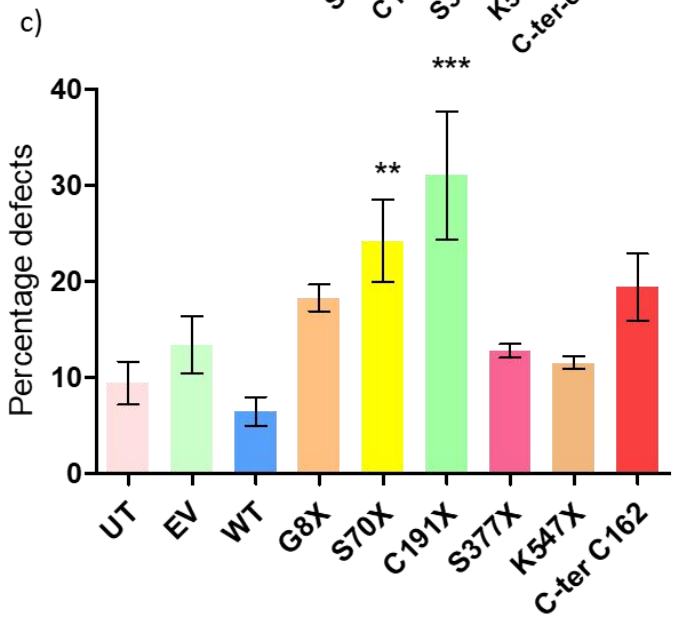


Figure 9: (b) EFHC2 domain-truncation alleles and their sub-cellular localization. HEK293T cells expressing GFP-EFHC2 wildtype and mutant constructs co-stained with γ-tubulin (red). The degree of col-localization between EFHC2 and γ-tubulin is represented as Pearson correlation coefficient (PCC), $0 < R < 1$; A higher R-score indicates better co-localization between EFHC2 and γ-tubulin.

The protein localizations of domain truncation alleles with one or more DM10 domains were as follows - in cells with the truncation after the first DM10 domain i.e., GFP-EFHC2-Cys191Ter (C191X), the protein was found in the cytoplasm. However, in a subpopulation ~10% of cells, we observed the protein localizing to the spindle apparatus (Figure 10a). In the constructs with truncation after the second DM10 domain and third DM10 domains – GFP-EFHC2-Ser377ter (S377X) and GFP-EFHC2-Lys547ter (K547X) respectively, the protein was found at the spindle poles and midbody during cell division (Figure 10a).

Consequently, we found that constructs that failed to specifically localize to the mitotic spindle caused an increased proportion of cells with cell division defects. The defects observed were lagging anaphase chromosomes, multipolar and monopolar spindles, abnormal positioning of the spindle poles, asymmetric divisions, and cytokinesis defects leading to large multinucleated cells (Figure 10b). In comparison to the wildtype protein, expression of GFP-EFHC2-Ser70Ter and GFP-EFHC2-Cys191Ter led to three-fold increase in cells with mitotic defects. Similar trend was observed in cells with GFP-EFHC2-Gly8Ter, GFP-EFHC2-Cter-C162 (Figure 10c). The mitotic defects were less pronounced in cells expressing GFP-EFHC2-Ser377Ter and GFP-EFHC2-Lys547Ter suggesting that the second DM10 domain and third DM10 domains are important for proper localization and cell division related functions of the EFHC2.


a)
Percentage localization
150
100
50
0
***
***
***
***
***
ns
ns
***
UT
EV
WT
G8X
S70X
C191X
S377X
K547X
C-ter-c162


c)
Percentage defects
40
30
20
10
0
***
**
UT
EV
WT
G8X
S70X
C191X
S377X
K547X
C-ter C162

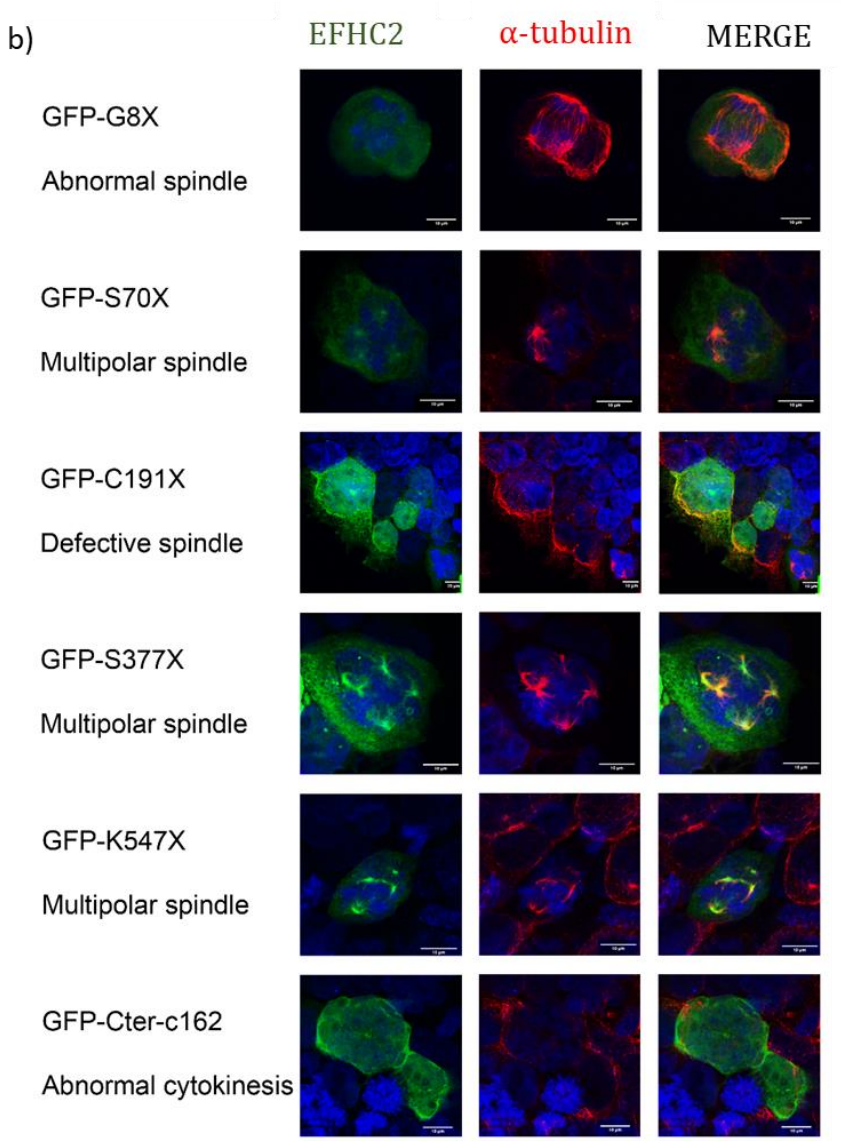

b)
EFHC2
α-tubulin
MERGE
GFP-G8X
Abnormal spindle
GFP-S70X
Multipolar spindle
GFP-C191X
Defective spindle
GFP-S377X
Multipolar spindle
GFP-K547X
Multipolar spindle
GFP-Cter-c162
Abnormal cytokinesis

Figure 10: Localization and cell-division defects of the EFHC2 truncation alleles. (a) Quantitation of constructs localizing to centrosome, spindle apparatus and the midbody (N=4 and n=800-1000 cells/construct; one-way ANOVA, P < 0.05; Dunnett's multiple comparison test). **(b)** Representative images of the mitotic defects observed in HEK293T cells expressing EFHC2 wildtype and mutant alleles and **(c)** Quantitation of the cell division defects on expression of EFHC2 mutant alleles (N=4 and n= 600 cells/construct; one-way ANOVA, P < 0.05; Dunnett's multiple comparison test)

In immuno-pulldown assays, we tested if the absence of the second and third DM10 domains affected the interaction of EFHC2 with α-tubulin. In agreement with our immuno-cytochemistry findings, the alleles that failed to localize specifically to the mitotic spindle or midbody, did not associate with alpha- tubulin (Figure 11).

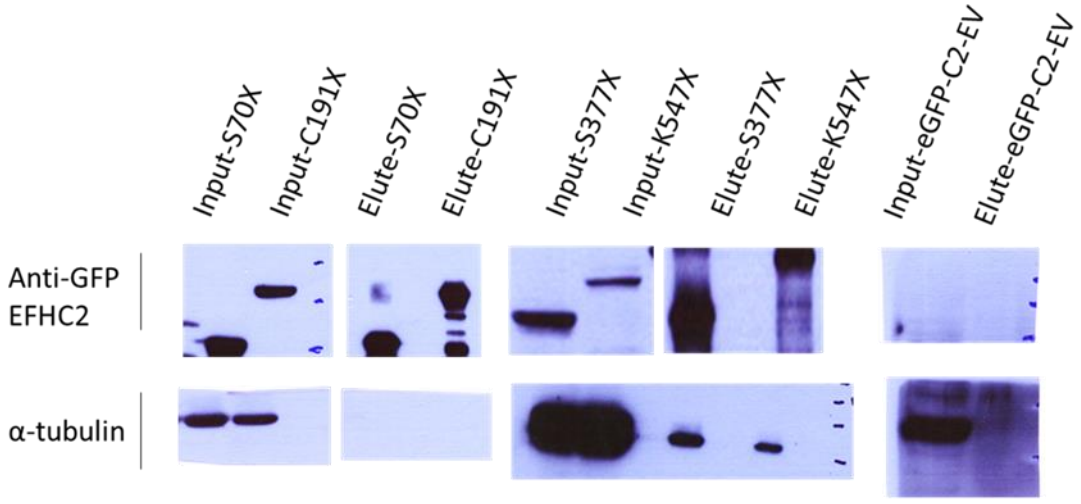
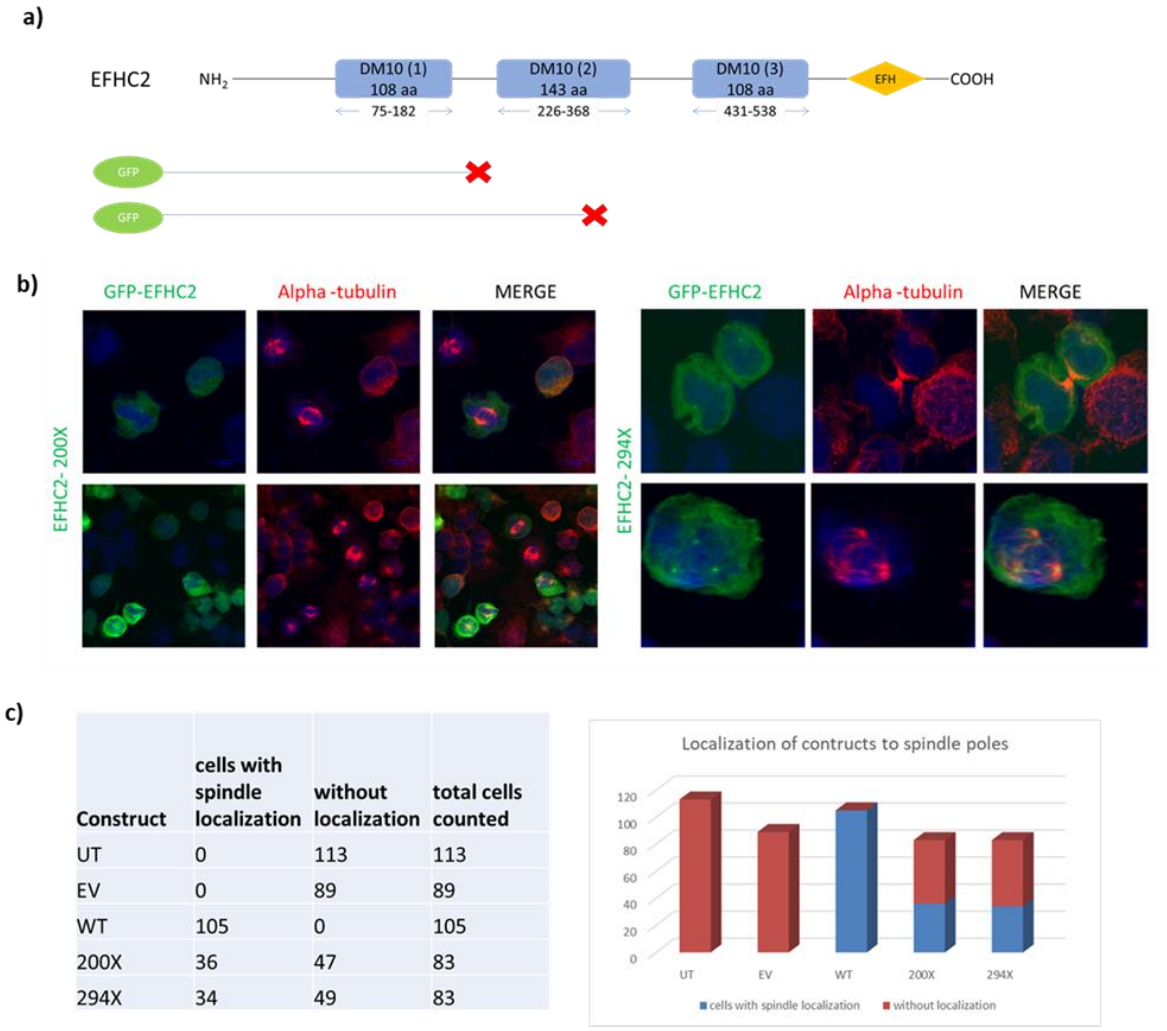


Figure 11: Co-immunoprecipitation of EFHC2 domain truncation constructs with α-tubulin. The presence of protein bands of comparable sizes in the input and elute fractions validate expression of mutant proteins and their immunoprecipitation. The interaction of EFHC2 with α-tubulin was detected only in wildtype and constructs S377X and K547X

We wondered if the spindle and microtubule association of EFHC2 was facilitated by specific motifs of the protein. On bioinformatic analysis using the Microtubule-Associated Protein Analyzer (Zhou et al., 2015) with high stringency cut off, and 90% specificity, two peptide stretches between amino acids 283-292 and 285-292 were predicted to facilitate microtubule association. We tested the prediction using immunocytochemistry by examining the localization of GFP-EFHC2-Arg200ter and GFP-EFHC2-Pro294ter in HEK cells. In around ~75% of cells, the truncated proteins did not localize to centrosomes, mitotic spindles or midbody. However, in a fraction of cells we noticed the truncated EFHC2 associated with the spindles and midbody. Thus, we were unable to define the amino acid motifs necessary for microtubule association of EFHC2 (Figure 12)



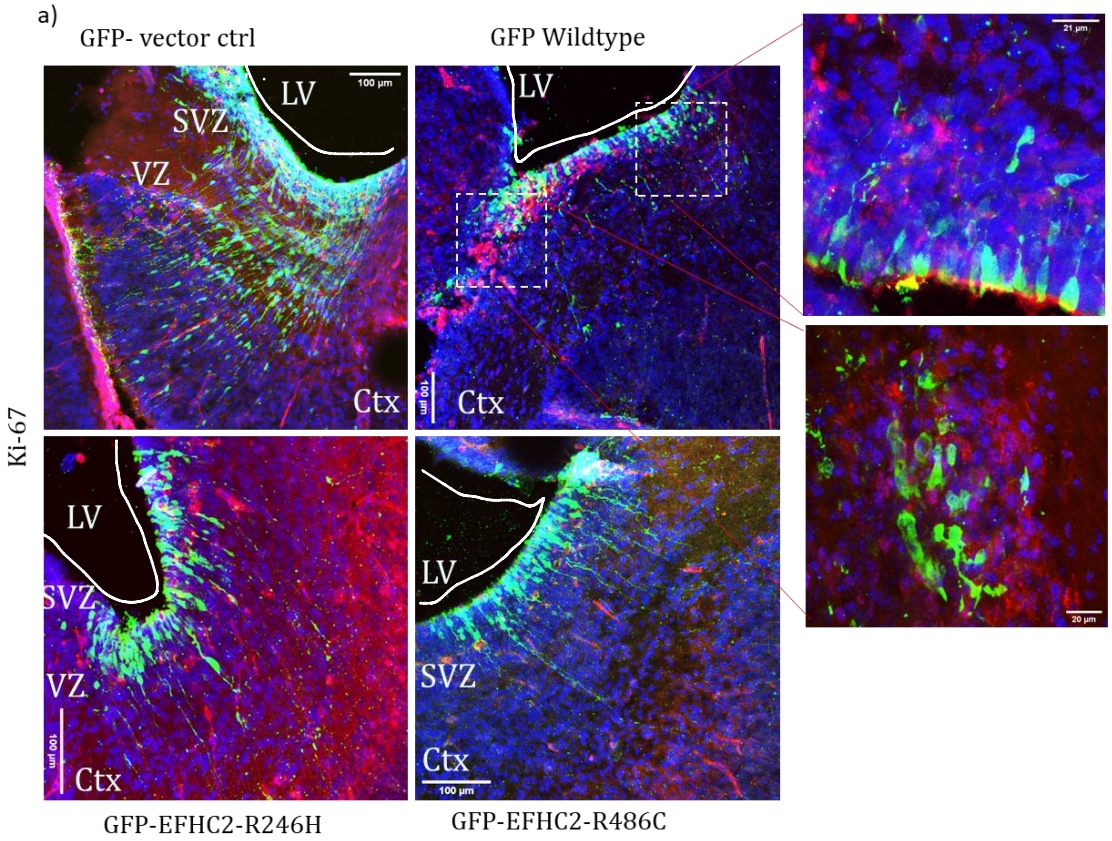
Construct	cells with spindle localization	without localization	total cells counted
UT	0	113	113
EV	0	89	89
WT	105	0	105
200X	36	47	83
294X	34	49	83

Figure 12: **(a)** Representation of the position of truncations in EFHC2, **(b)** Representative images of GFP-EFHC2-200X and EFHC2-294X stained with alpha tubulin, and **(c)** Quantitation of length no. of cells in with protein localizing to centrosome, spindle or midbody.

2.3.5 Expression of EFHC2 during cerebral cortex development

EFHC2 wildtype and patient variants p.Arg246His and p.Arg486Cys were ectopically expressed in the developing mouse brain at e14.5-15.5. pEGFP-C2-empty vector was used as the control. GFP expression was seen in a vast majority of cells spanning the ventricular zone, intermediate zone as well as the cortical plate. EFHC2 expressing cells were mostly

concentrated in the ventricular and subventricular zones. In radial glial cells with distinct foot process, cell body resting in the ventricular zone and filamentous projections in the intermediate zone and the cortical plate, EFHC2 was distributed evenly (Figure 12b). In most of the transfected neuronal and other cell types, EFHC2 stained the cell body, cytoplasm, projections. In a subpopulation of cells, EFHC2 was also observed in the nucleus.



Construct	Avg length(um)
Vector ctrl	279.3
Wildtype	253.6
EFHC2-R486C	202.8
EFHC2-R246H	274.7

Figure 12: *in utero* electroporation of EFHC2 mutations. (a) Immunohistochemical analysis of mouse brain sections ectopically expressing GFP. GFP-tagged EFHC2 wildtype and mutants – p.Arg246His and p. Arg486Cys. Site of injection is marked as LV – lateral ventricle. Other regions are SVZ- sub-ventricular zone, VZ-ventricular zone and Ctx-cerebral cortex and **(b)** Quantitation of length of the projections of transfected cells across constructs. No. of cells analyzed- 100-200/construct.

The GFP-tagged mutant proteins expressed well in the developing brain. The distribution of the mutant protein was comparable to wildtype. The protein was seen in cell body, cytoplasm, filaments of radial glial cells and neurons. The length of projections of radial glial cells, identified by their morphology was found to vary marginally between wildtype and mutant constructs. EFHC2 expressed along the length of the projection of cells often extending from the ventricular zone to the cortical plate. The length of the projections was ~250µm. In the embryonic brains electro-transfected with the p.Arg246His the average length of the projection was higher than wildtype and was ~274µm. The length of projection in cells expressing p.Arg486Cys was ~202µm. The differences observed across wildtype and mutant constructs were not statistically significant.

2.4 Discussion

In this chapter, the results of a genetic study undertaking the sequencing of *EFHC2*, an X-linked paralog gene of *EFHC1*, in causation of juvenile myoclonic epilepsy (JME) are reported. Evidence of EFHC2 as a spindle and microtubule-associated protein gathered through immunocytochemistry, and immuo-pulldown assays are presented.

EFHC2 variants are associated with a range of neurological phenotypes such as social cognitive competence in women with Turner's syndrome (Weiss et al., 2007), in harm avoidance (Blaya et al., 2009), and social cognitive abilities in males (Startin et al., 2015). p.Arg135Gln identified in this study (Table 1) has been identified in a woman with mental retardation in the EuroMRX cohort (de Brouwer et al., 2007). Variants at p.Glu634Gly (Ganapathy et al., 2019) and p.Arg486Ser (Helsmoortel et al., 2015) are associated with autism spectrum disorder and intellectual disability. These data suggest a broad range of brain-related phenotypes linked to EFHC2. In a clinical case of Noorie

disease (ND) – a rare X-linked disorder, contiguous microdeletion at the Xp11.3 involving the *NDP, MAOA, MAOB*, and *EFHC2* genes was reported. In addition to the classical ND clinical features, myoclonic epilepsy was observed in the patient, and the authors suggested that the partial deletion in the *EFHC2* gene may be responsible for seizures (Rodriguez-Revenga et al., 2007). In a study of 96 Turkish idiopathic generalized epilepsy (IGE) patients and ethically matched controls were genotyped, no relation between S430Y polymorphism in *EFHC2* and IGE, was found (Berrin et al., 2015). Recently, in another patient with Noorie disease, a novel microdeletion of 494.6 kb on Xp11.3 was reported involving *NDP, EFHC2* and the first exon of *MAOB*. In this case, the proband did not have seizures, even though the deleted region contained the full-length EFHC2 gene (Bei Jia et al., 2017). In the context of these reports, genetic evaluation of association of variants in *EFHC2* to brain development and disease across diverse populations is important.

In this regard, the *EFHC2* gene was sequenced in epileptic individuals from Southern India. This led to the identification of seven potentially pathogenic variants at evolutionarily conserved residues, and these mutations gave rise to mitotic defects (Raju, 2014). On sequencing the gene in additional 72 patients, I found a potentially pathogenic variant at p.Arg165Glu in five JME patients. Additionally, in seven JME patients, an SNP at p.Ser430Tyr and in one individual p.Glu506Lys was identified. Taken together, the seven mutations identified in EFHC2 alter conserved amino acid residues, and six are predicted to be disease-causing. None of the rare, missense variants identified in the study are reported for GGE in the Epi25 Collaborative database (Epi25 Collaborative. Electronic address & Epi, 2019). This underscores the need

for replicative studies to examine the genetic burden of *EFHC2* for epilepsy.

EFHC2 gene encodes a 749 amino acid protein with three DM10 domains. The structure and biological function of the protein are not known so far. Immuno-fluorescence studies done in HEK293T, U-87MG, and SH-SY5Y cells show the localization of EFHC2 to centrosomes, mitotic spindles, and midbody (Figure 4).

Although the variants identified in JME patients did not affect the localization of the protein, they caused mitotic defects in chromosome segregation, spindle formation, positioning, and cytokinesis (Raju, 2014).

Centrosomes are membrane-less organelles involved in microtubule (MT) nucleation and maintenance during interphase, thus regulating cell shape, polarity, and motility. During mitosis, they increase MT nucleation capacity and form the poles of the mitotic spindle apparatus. Further, they also act as templates for cilium growth. Typically, a mammalian centrosome contains hundreds of proteins involved in centriolar structure in the pericentriolar material (PCM). A set of these proteins are permanently associated with the centrosome, whereas others appear during different stages of the cell cycle (Alieva & Uzbekov, 2008). Proteins such as γ-tubulin, the γ-tubulin complex proteins, and Centrin, a highly conserved subgroup of Ca2+-binding proteins, are permanently associated with the centrosome structure. In our experiments, we find that EFHC2 is present at the centrosomes during interphase, is observed at the spindle pole, and continues to localize to the mitotic spindles through metaphase and anaphase, and at telophase, the protein is found at the midbody (Figure 2 and 3). In literature, EFHC2 is shown to be expressed in cilia (van Dam TJP,

2019). The occurrence of EFHC2 at centrosome, mitotic spindle, midbody and cilia suggests that EFHC2 is a likely permanent resident of the centrosome.

The centrosome is a barrel-shaped organelle composed of centrioles that have 9-fold radial symmetry and nine outer doublet microtubules (DMTs). These DMTs further have ribbons of three to four hyper-stable protofilaments (PFs). EFHC1 and EFHC2 are proteins associated with the protofilaments. Orthologues of EFHC1 and EFHC2 are present across species such as *Chlamydomonas*, Protostomia (*C. elegans* and *Drosophila*), echinoderms represented by *S.purpuratus* and in vertebrates such as *D.rerio*, *M.musculus,* and *H.sapiens*. In sea urchin sperm flagella, they are found associated with the lumen of the DMTs (Linck et al., 2014). In the Xenopus embryos, EFHC1 and EFHC2 are present at the axonemes of multiciliated cells (F. Tu et al., 2018). They are important for the assembly of the A-tubule of the doublet microtubules (Nicastro et al., 2011), which are parts of axonemes. Mutations in centrosomal components are associated with several neurological diseases, such as lissencephaly and microcephaly, psychiatric disorders, such as schizophrenia, or syndromes with neurological symptoms, such as Bardet–Biedl, and Joubert syndrome (Kuijpers & Hoogenraad, 2011). Although the molecular mechanisms underlying these diseases are being researched, an existing hypothesis is that centrosomes aid in spindle positioning in neural progenitors. They are also thought to bring about to a balance between the expansion of progenitors and generation of neurons. Also, some of the divisions with abnormal centrosomes lead to aneuploidy and cell death. It is equally possible that the microtubule-based structure of the cilia is altered, or the structure of the cilia is intact, while its sensory function might be affected.

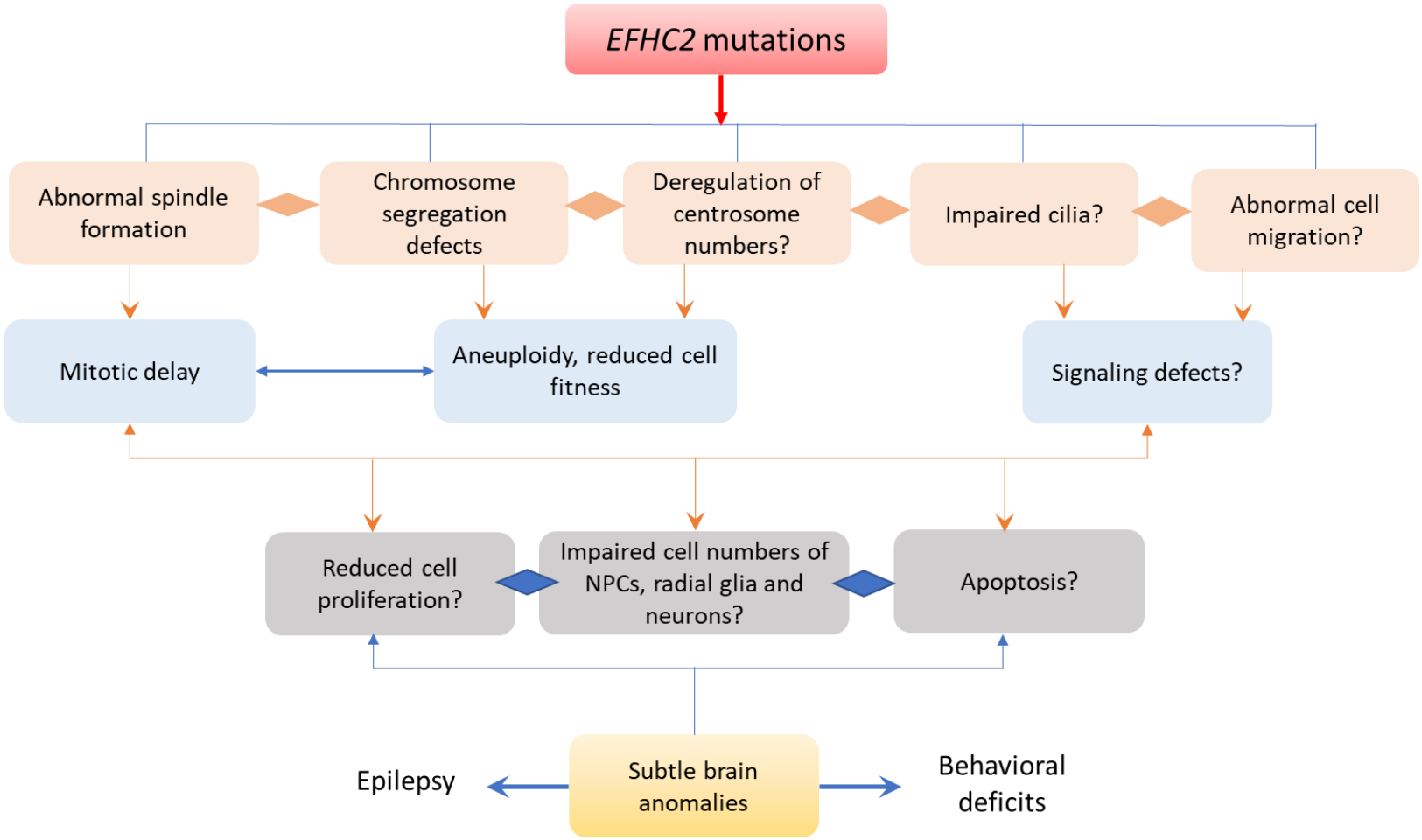


Figure 13: Putative cellular mechanisms and effects of EFHC2 mutations

It is interesting to note that for JME, two ciliary genes are known to be associated with the disease – *EFHC1* and *CILK1*. Mutations in these genes affect the cell division associated and ciliary function of the proteins. Interestingly, EFHC2 shares homology and is the only known paralog of EFHC1. EFHC1 and EFHC2 share 31% identity and 46% similarity. The domain organization is conserved, and among the respective DM10 domains of these two proteins, they have a similarity of 45-60% (Figure 13).

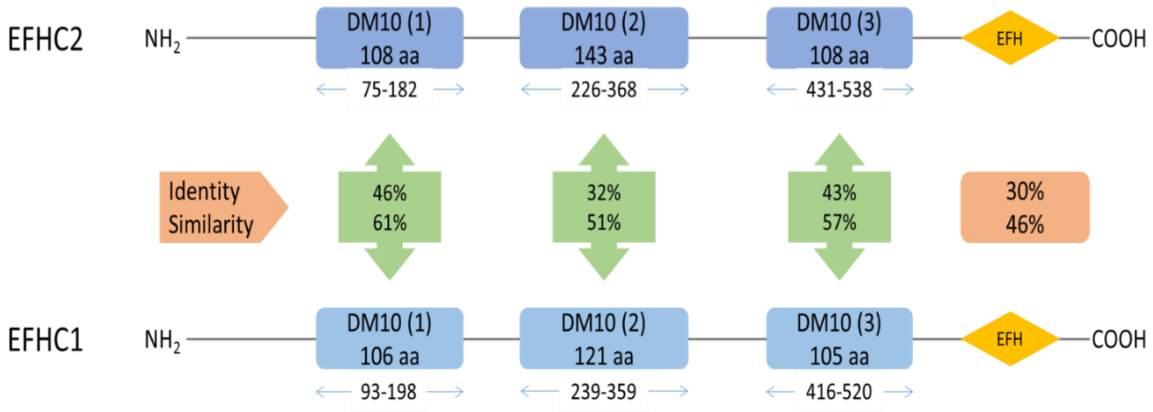


Figure 13: Homology between EFHC2 and EFHC1. The three DM10 domains are marked in blue and EF-hand domains in yellow.

EFHC2 sequence contains three uncharacterized DM10 domains. Bioinformatic analysis of EFHC2 using linear motif and domain predictions using the ELM relational database (http://elm.eu.org/index.html) (M. Kumar et al., 2020), we observe that the N-terminus has disordered regions and three low complexity regions at amino acids 379-393, 546-559 and at 665-686. It has ordered or globular regions between amino acid residues 194-284, 295-372, and 414-749. The protein also contains two potential regions with serine/threonine residue phosphorylated by the PLK1 kinase between 371-377 amino acids [ENFTSVS] and 400-460 [EEDSLRN]. The protein

is predicted to have several residues found in substrates for MAP kinases, which often regulate the stability and offer specificity to protein localization. Interestingly, the protein at the C-terminus, at 583-591 amino acids, has an RxxL-based motif that binds to the CDH1 and CDC20 components of APC/C, thereby targeting the protein for destruction in a cell cycle-dependent manner.

In line with these predictions, in immuno-pulldown assays, EFHC2 co-eluted with several cell division-related proteins such as PLK1 and CDC20B. EFHC2 was found to interact with α-tubulin and co-stained the spindle apparatus, thus proving that EFHC2 is a spindle-associated, microtubule interacting protein. γ-tubulin is a centrosomal protein that is also present at microtubule nucleation sites. Immunofluorescence analysis revealed that EFHC2 and γ-tubulin are localized within the centrosome throughout the cell cycle and co-immunoprecipitates with γ-tubulin, indicating that EFHC2 is part of the γ-tubulin complex. Mutations in EFHC2 lead to alteration of number and position of the mitotic spindle, further influencing the cell cycle. The cell cycle defects due to mutations in *EFHC2* are partially rescued by wildtype *EFHC1* (Raju, 2014).

Given the similarity in sequence, localization, possible redundancy between EFHC1 and EFHC2, and the association of their orthologs with protofilaments in the axoneme, we were curious to see if these proteins are associated physically. To our surprise, in co-immunoprecipitation experiments, we did not detect their co-elution. It is possible that their interaction is dynamic and requires a more sensitive detection method to capture the interaction. We also noticed several differences between EFHC1 and EFHC2. It is reported that EFHC1 interacts with α-tubulin but does not interact with γ-tubulin (de Nijs et al., 2009). In contrast,

we report here the interaction of EFHC2 with both α- and γ-tubulins. EFHC1 contains a microtubule-binding motif in the first 45 amino acids of the protein that drives the protein localization to the mitotic spindle. However, in EFHC2, we find that the amino acid stretch between 300-538 consisting of the second and third DM10 domains is important for the protein localization to the mitotic spindle (Figure 9 and 10) and the association with α-tubulin (Figure 11).

Experiments testing the genetic interaction of human EFHC1 and EFHC2 are not reported to date. *Tetrahymena* and *Chlamydomonas* have two RIB72 proteins, RIB72A and RIB72B. Both these proteins have three DM10 domains. In *Tetrahymena,* RIB72A and RIB72B localized to the cilia and basal bodies. *Tetrahymena,* knockout strains of RIB72A, RIB72B, and both RIB72A/B showed reduced swimming speeds, ciliary beat frequencies, and abnormal highly curved cilia with structural defects in the ciliary A-tubule (Stoddard et al., 2018). It was also found that RIB72A assembles in the cilia and basal body independent of RIB72B (Stoddard et al., 2018). This data is in line with our observations of a lack of interaction of EFHC1 and EFHC2, suggesting their involvement and regulation of similar cellular phenotypes independently. Other genetic models in cells, flies, and mice would be crucial to delineate further the interaction between EFHC1 and EFHC2 and their influence on brain development and disease.

EFHC2 is a spindle – and microtubule – associated protein. EFHC2 influences mitotic spindle pole formation and orientation. During brain development balancing the number of progenitor and differentiated cells is crucial. These cell fate decisions often depend on the orientation of cell division axis and inheritance of midbody which regulate portioning the fate determinants into the daughter cells. It is well known that the

cleavage planes determine symmetric divisions (for the expansion of neural stem cell pool) and asymmetric divisions (for neurogenesis) and influence corticogenesis (Taverna et al., 2014). The mode of cell division is also influenced by duration of cell cycle. EFHC2 mutations cause abnormal spindle numbers and positioning in HEK293T cells. Cells with abnormal centrosome numbers and spindles take longer or fail to form bipolar spindles. Similar mitotic delays and spindle misorientation have been reported associated with *EFHC1, CILK1* mutations associated with JME; *WDR62, ASPM, CDKRAP2* and *STIL* associated with microcephaly (Chavali et al., 2014). It would be interesting to examine if EFHC2 is expressed in neural stem cells (NSCs) and whether the mutations exert any effect on the spindle architecture and precise regulation of duration of cell cycle during neurogenesis (Figure 12). EFHC2 is associated with cilia. In humans, it is expressed in respiratory ciliated cells (X. Sun, 2009), endometrial ciliated cells (https://www.proteinatlas.org/ENSG00000183690-EFHC2/single+cell+type). It remains to be tested if in humans, EFHC2 is present in either or both motile and primary cilia. EFHC2 interacts with CDC20B (Figure 7). Mutations in *EFHC2, CDC20B* cause JME. Recently, CDC20B was identified in deuterosomes during centriole amplification, and at the basal bodies and cilia in fully differentiated mammalian multiciliated cells (Revinski et al., 2018). EFHC2 and CDC20B have common interacting partners γ-tubulin and CCDC57 (https://www.ebi.ac.uk/intact/search?query=EFHC2). CCDC57 is a regulator of centriole duplication, ciliogenesis, and mitosis (Gurkaslar et al., 2020). It is tempting to thus speculate that these cilia associated proteins may play crucial roles in centrosome and cilia biogenesis and function.

In conclusion, in this chapter, we report the contribution of brain expressed *EFHC2* and its variants to an epilepsy phenotype, *in vitro* characterization of missense and domain truncation alleles of EFHC2, as well as evidence to the association of EFHC2 with microtubules. EFHC2 affects cell division, but the mechanisms of mitotic defects due to mutant proteins remain unknown. Studies on the contribution of EFHC2 alleles towards the spectrum of brain-related phenotypes such as intellectual disability and epilepsy and identification of EFHC2-interactome would be important to relate cellular functions of EFHC2 to brain-related phenotypes.

Chapter 3

Characterization of *Efhc2* and generation of *Efhc2* mutant mice

Summary:

Murine *Efhc2* is a gene on the X-chromosome encoding EFHC2, a 750 amino acids protein consisting of three DM10 domains and a predicted C-terminal EF-hand. The expression, cellular localization, and functions of this protein are unknown. In this chapter, the expression and cellular characterization of EFHC2 are reported. Here, the strategies used to generate *Efhc2* mutant mice lines are described, and the outcomes are discussed.

EFHC2 is expressed ubiquitously in the various mouse tissues such as the brain, heart, lung, liver, kidney, etc. On the histological analyses of EFHC2 expression in the mouse brain, EFHC2 was detected in the cerebral cortex, hippocampus, thalamus, hypothalamus, ependymal cells, and cerebellum. Expression of EFHC2 was observed in neurons and not in glial cells. In neurons, EFHC2 localized to the cell body and the axons.

In HEK293T cells, EFHC2 localizes in the nucleus, cytoplasm, and centrosomes. The protein was associated with the mitotic spindle and midbody during cell division. This association was validated by co-immunoprecipitation of EFHC2 with tubulins – α (alpha) and -γ (gamma).

For *in vivo* characterization of *Efhc2*, the knockout-first-reporter tagged *Efhc2*$^{(tm1a)}$ conditional mice lines were established. The knockout strategy targeted the deletion of exon 3 of *Efhc2*.

Attempts at generating the EFHC2 null allele resulted in the generation of two mutant mouse lines - the exon 3 deletion allele and the En2-insertion allele. The resulting mutant proteins were examined in HEK293T cells. The mutant proteins mislocalize and lead to mitotic defects.

3.1 Introduction

EFHC2 is an X-linked gene encompassing about 207.3kb of DNA. It encodes a 2554bp transcript and a 750 amino acids long protein. The protein is predicted to contain three DM10 domains and a C-terminal EF-hand domain (Figure 1a). The human and mouse EFHC2 are both X-linked and are well conserved with an overall similarity of 85% (Figure 1b; for pair-wise sequence alignment, see Appendix -1). Large-scale transcriptomic analyses reported on the EMBL-EBI expression atlas show the *Efhc2* transcript to be ubiquitous. The transcripts are detected in the human and mouse testis, adrenal gland, brain, lung, colon, heart, etc. (Papatheodorou et al., 2020). RNA-Seq (Tahara et al., 2019) and microarray analyses (Ougland et al., 2016) have reported *Efhc2* expression in mouse embryonic stem cells suggesting the potential roles of EFHC2 during development (Figure 2).

Mutations in the human ortholog of EFHC2 are proposed to be associated with neurological disorders and behavioral traits such as - juvenile myoclonic epilepsy (Gu et al., 2005), autism spectrum disorder (Ganapathy et al., 2019), social cognition (Startin et al., 2015), fear recognition and harm avoidance (Blaya et al., 2009). However, the cellular and mechanistic details of EFHC2's involvement in these attributes remain unexplored. To examine the possible roles of EFHC2 for these phenotypes, we have initiated the characterization of EFHC2 in murine and cell culture models.

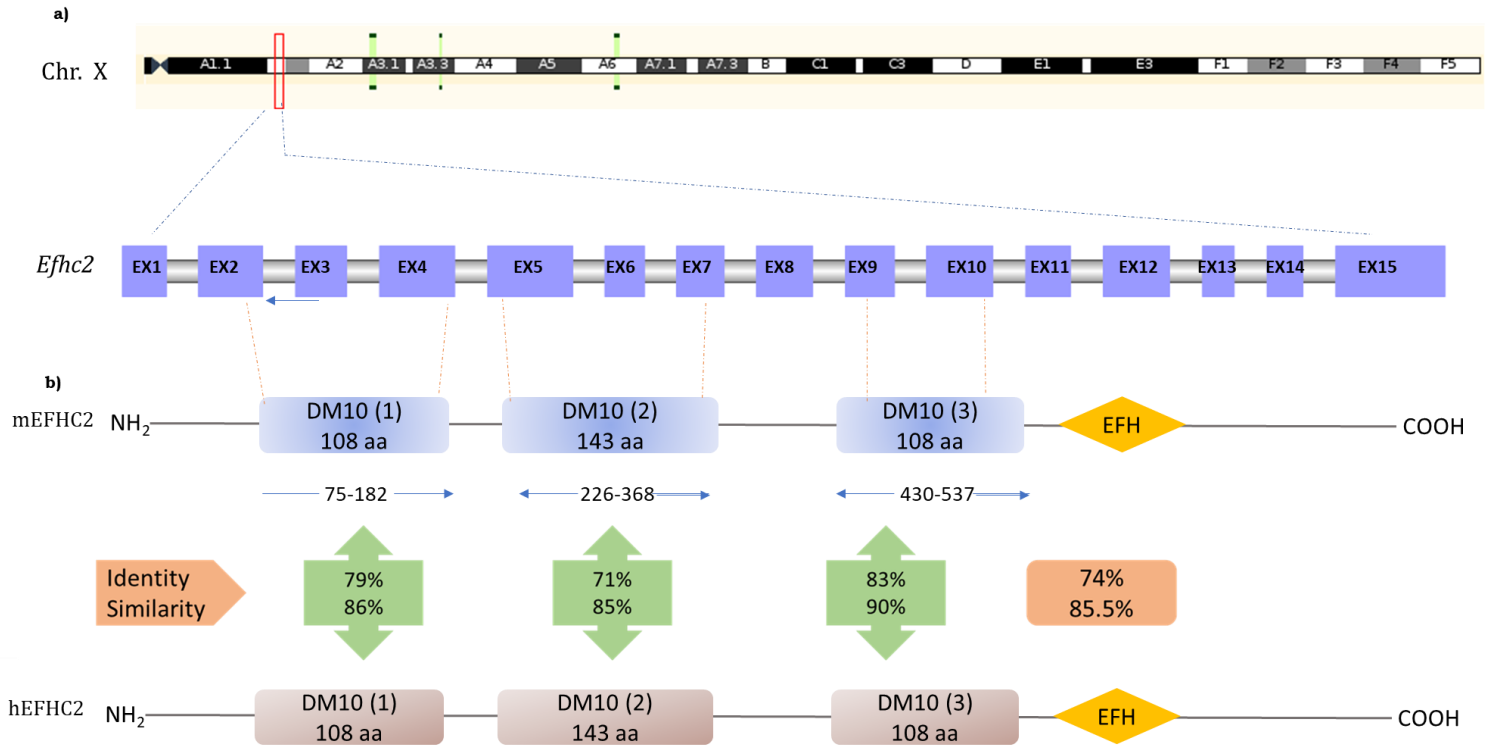


Figure 1: (a) *Efhc2* is located on the X-chromosome and the 2554bp transcript consists of 15 exons. The protein coded for in 750 amino acids in length with three DM10 domains and a probable EF-hand motif at the C-terminus. **(b)** The human and mouse EFHC2 proteins have similar domain architecture with a high degree of conservation between the domains and across the length of the protein

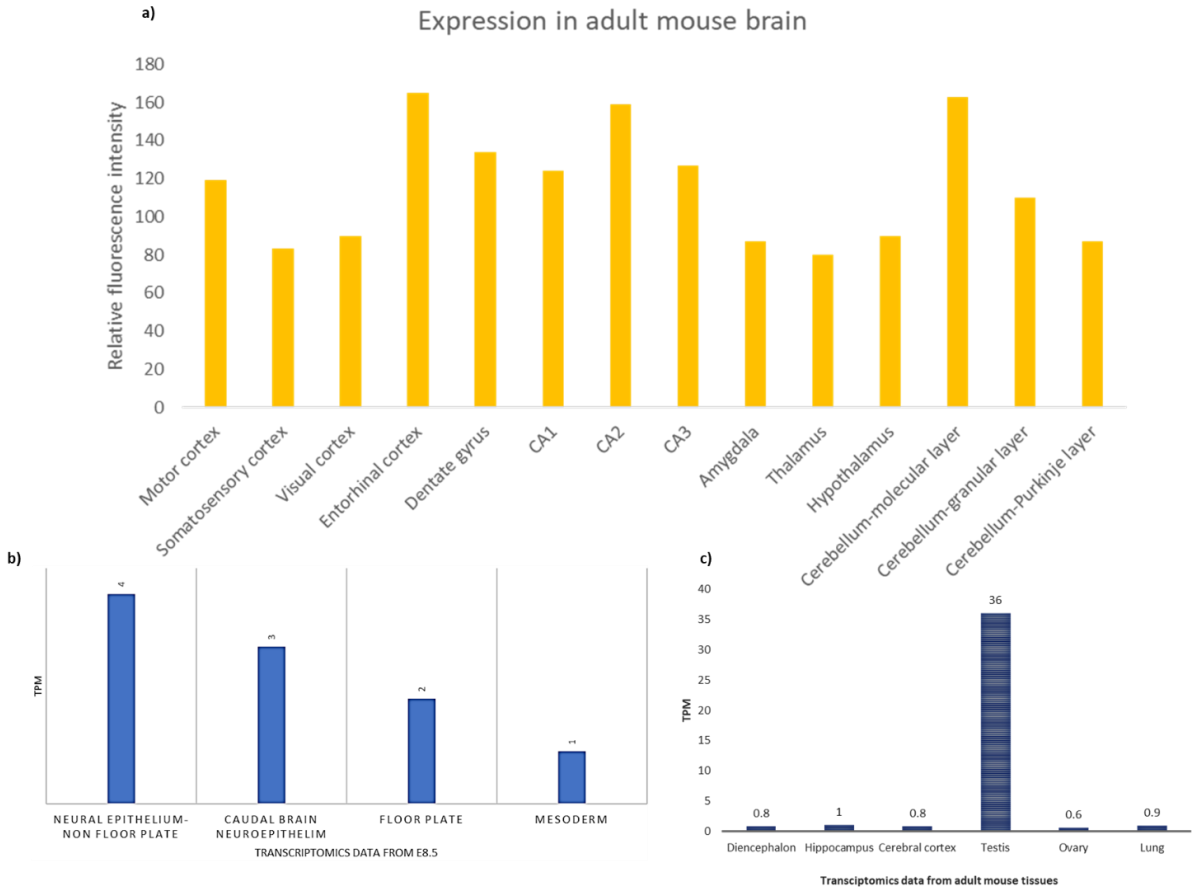


Figure 2: (a) EFHC2 is expressed across various mouse brain regions – expression reported as relative fluorescence intensities in the human protein atlas are plotted here. Transcriptomics data for mouse *Efhc2* at **(b)** embryonic day 8.5 and **(c)** in the adult.

Several aspects relating to mouse EFHC2 remain unknown. This study aims to understand the expression- and localization- of mouse EFHC2. Further, study in detail the sub-cellular localization and function of EFHC2. Considering the association of EFHC2 to several neurological traits, to facilitate the examination of the association of EFHC2 to the reported phenotypes, we aimed to generate EFHC2 mutant mouse lines.

3.2 Materials and methods

3.2.1 Animal breeding and maintenance

C57BL/6NCr, C57BL/6FLPE, B6.β-actin Cre, C57BL/6NCr-*Efhc2*$^{flox/flox}$, C57BL/6NCr-*Efhc2*$^{En2\text{-}ins}$, and C57BL/6NCr-*Efhc2*$^{\Delta/\Delta}$ mice strains were housed in groups under a 12-hour light/dark cycle with access to food and water ad libitum. Matings were set up either in the ratio of 1:1 or 1:2 (male: female) to maintain transgenic lines. All the progenies were weaned post 21 days, earmarked, and tail clipped for genotyping. All experimental procedures followed the guidelines set by the Institutional Animal Ethics Committee (IAEC) of the JNCASR and in accordance with the CPCSEA (Committee for the Purpose of Control and Supervision of Experiments on animals) guidelines.

3.2.2 Generation of *Efhc2* floxed mice

The homozygous floxed *Efhc2* (*Efhc2*$^{flox/flox}$) mice were generated in a collaboration with Dr. Shyam Sharan, National Cancer Institute (NCI), National Institute of Health (NIH), USA. The strategy was as follows: *Efhc2* locus Xp11.3 was targeted using the gene targeting vector introduced into mouse embryonic stem (ES) cells to generate floxed Efhc2, the tm1a allele. The vector design consists of 5.9Kb of the 5'-homology arm mapping to intron 2 of *Efhc2* and 3' homology arm of 4Kb mapping to intron 3. The critical region consisting of exon 3 is flanked by FRT and loxP recognition sites for flippase and Cre recombinase enzymes. The vector used has multiple genetic elements consisting of an Engrailed 2 (En2) splice acceptor followed by an IRES: LacZ trapping cassette and an SV40 polyadenylation signal. A neomycin cassette under the control of the promoter of the human β-actin gene (ACTB) flanked by two FRT recombination sites at the 5' and 3' ends of the cassette is present. LoxP sites are engineered to facilitate recombination and generation of a

conditional allele. It also has a promoterless reporter gene LacZ, regulated by the endogenous promoter of EFHC2 (Figure 3). The gene trap vector was introduced into ES cells and screened for vector integration. In the first step, ES cells were screened for neomycin resistance conferred by integrating the neomycin resistance cassette present in the vector (Figure 3). In the second step, random integrations were selected using the Diptheria toxin fragment A (DT-A) present downstream of the homology region of the construct (Figure 1, Appendix 2). The ES cells with the desired integration were injected into the blastocoel, and the chimeric embryos were transferred into the uterus of recipient pseudopregnant foster females to generate *Efhc2*$^{flox/WT}$ female or *Efhc2*$^{flox/Y}$ male mice. The heterozygous *Efhc2*$^{flox/WT}$ mice were mated to obtain homozygous *Efhc2* $^{flox/flox}$ mice. The boundaries of homology arms, reporters, and the DNA between recombinase sites were confirmed using PCR amplification and sequencing in the mice so generated.

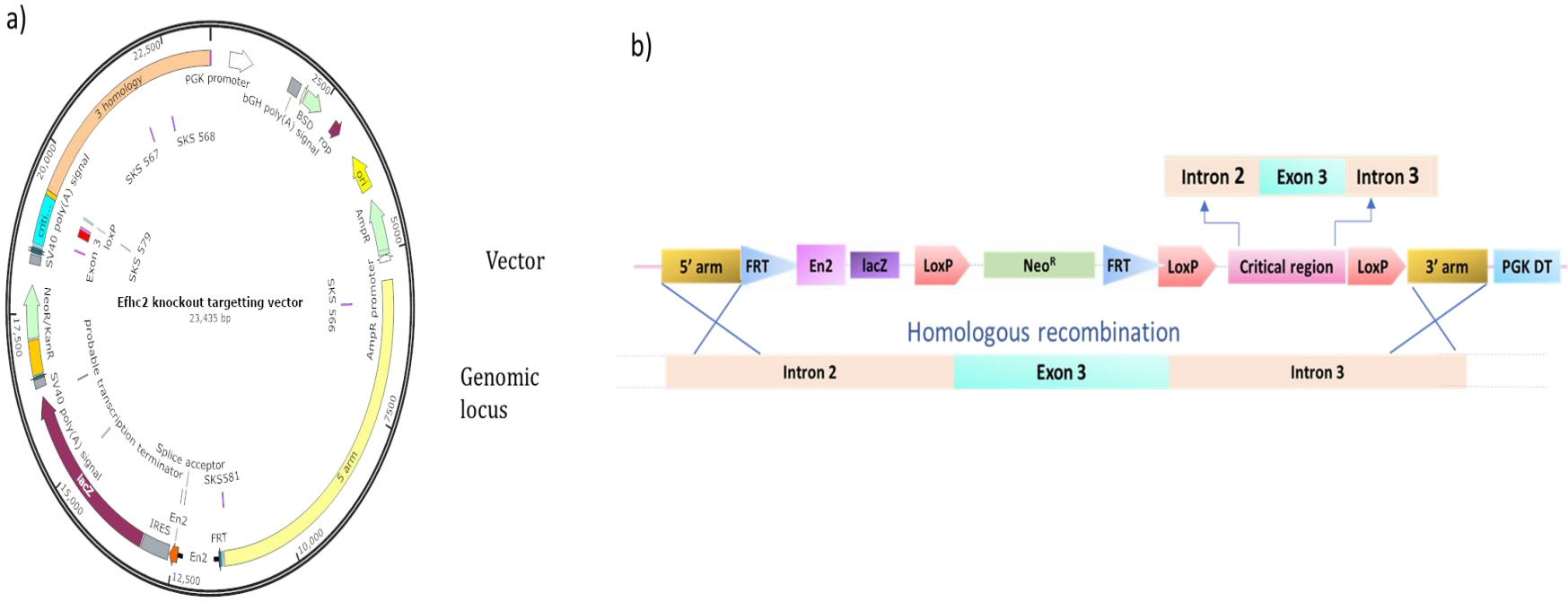


Figure 3: (a) *Efhc2* knockout targeting plasmid used for the generation of *Efhc2* floxed$^{(tm1a)}$ mice. **(b)** The plasmid was generated by cloning critical regions between the recombinase sites, flanked by homology arms from *Efhc2* introns

3.2.3 Generation of *Efhc2* loxed mice

Efhc2$^{flox/flox}$ mice contain loxP sites flanking exon 3 of the *Efhc2* gene. On Cre-mediated recombination, the deletion of the critical exons leads to a null allele. The *Efhc2*$^{flox/flox}$ mice were mated to transgenic mice expressing Cre-recombinase under the β-actin promoter. The resultant mice have a constitutive deletion of exon 3 and are loxed. The heterozygous and hemizygous loxed mice were interbred to generate homozygous loxed mouse lines (Figure 4).

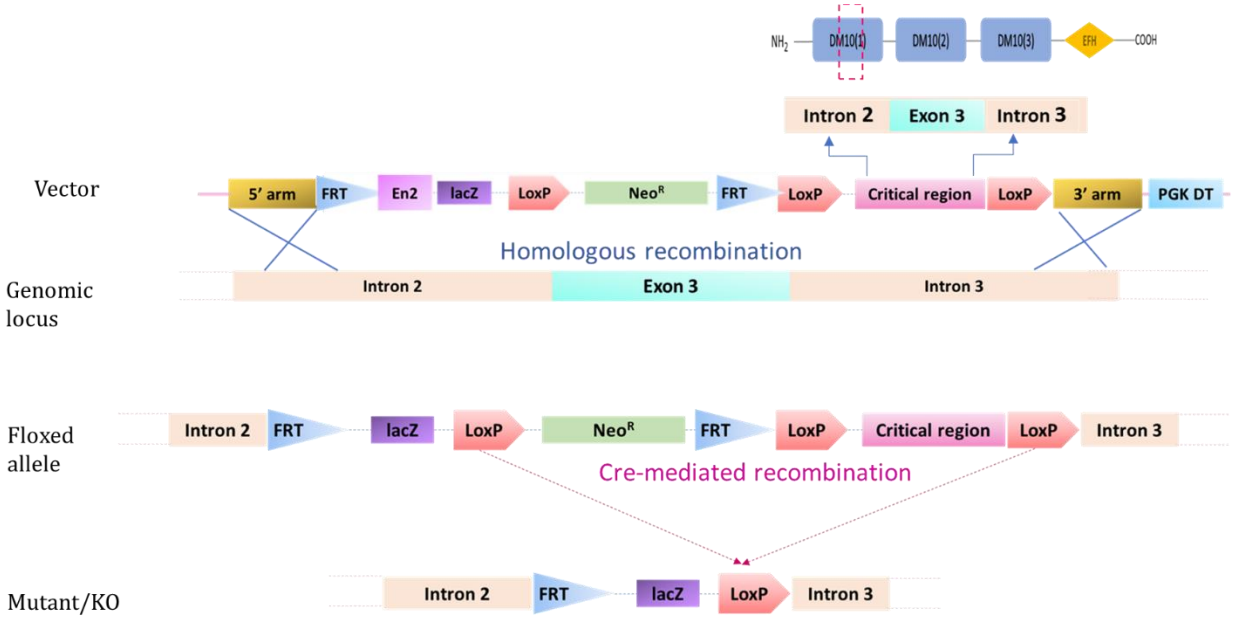


Figure 4: Simplified illustration of *Efhc2* knockout targeting vector. On vector integration at the genomic locus, *Efhc2*$^{flox/flox}$ (*Efhc2*^{tm1a}) allele is generated. Further, on Cre-mediated recombination, the critical region consisting of exon 3 is deleted and *Efhc2* is loxed and *Efhc2*$^{(tm1b)}$ mice are generated.

3.2.4 Generation of *Efhc2* -flipped and -loxed mice

The homozygous floxed *Efhc2* (*Efhc2*$^{flox/flox}$) were mated to Flippase expressing transgenic mice to generate the *Efhc2* conditional allele. The

DNA sequence between the two FRT sites, consisting of En2 splice acceptor, LacZ reporter, and the Neomycin resistance gene, are removed on flippase-mediated recombination. The resultant conditional allele has the critical exon flanked by two loxP sites. We set up crosses between C57BL/6FLPE and C57BL/6NCr-*Efhc2*$^{flox/flox,}$ and the progeny obtained were labeled Tg.FLP; *Efhc2*$^{Flip/WT}$. The litters were genotyped using PCR, and FRT-mediated recombination was confirmed. Tg.FLP; *Efhc2*$^{Flip/WT}$ mice were intercrossed to generate Tg.FLP; *Efhc2*$^{Flip/Flip}$ or Tg.FLP; *Efhc2*$^{Flip/Y}$ (Flipped mice).

To generate the null allele, exon 3 of *Efhc2* was targeted for deletion. In Tg.FLP; *Efhc2*$^{Flip/Flip}$ mice, the critical exon is flanked by two loxP sites. We set up crosses between the B6.β-actin Cre and Tg.FLP; *Efhc2*$^{Flip/Flip}$ mice to obtain flipped and loxed alleles - Tg.FLP;β-actin Cre; *Efhc2*$^{Δ/Y}$ and Tg.FLP;β-actin Cre; *Efhc2*$^{Δ/WT}$ mice. These mice were further intercrossed, and *Efhc2*$^{Δ/Y}$ and *Efhc2*$^{Δ/Δ}$ alleles with constitutive deletion of the exon3 of *Efhc2* were obtained (Figure 5).

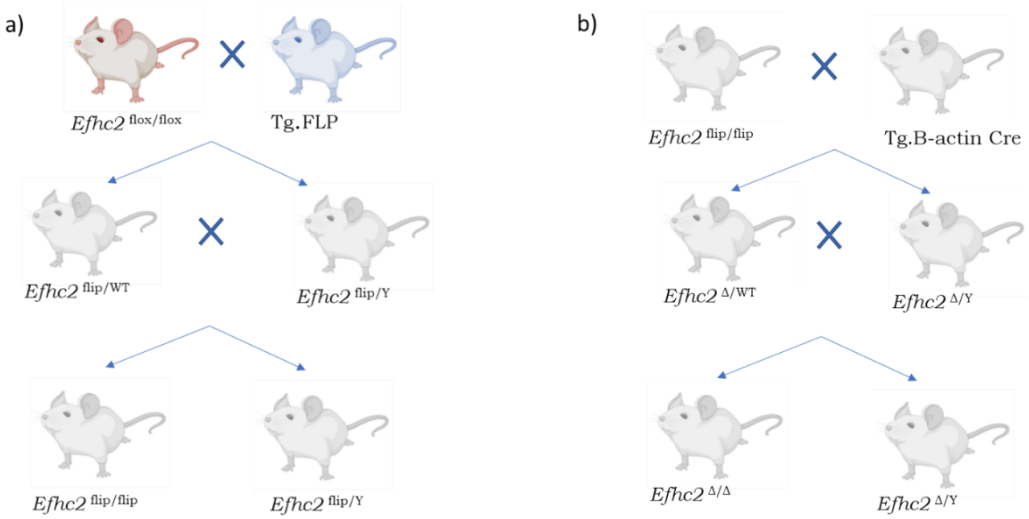


Figure 5: (a) The breeding strategy used to excise the reporter cassette using FRT-mediated recombination to generate the flipped Efhc2$^{(tm1c)}$ allele. **(b)** Breeding strategy used to generate loss of function allele with the excision of exon 3 and the generation of the flipped and loxed *Efhc2*$^{(tm1d)}$ allele.

3.2.5 Genomic DNA isolation and genotyping

1cm tail-clippings were chopped into 4-5 pieces and resuspended in 180µl of 50mM NaOH. The samples were placed on a thermo shaker at 90°C, 1200 RPM, for 15 minutes. Post incubation, 20µl of 1M Tris at pH 8.0 was added to the samples and vortexed thoroughly. The samples were centrifuged at 13000 RPM for 15 minutes, and the supernatant containing the genomic DNA was collected. This protocol is a modification of the HotSHOT protocol (Truett et al., 2000). 1µl of the supernatant was used for genotyping. The primer pairs used are detailed in Table 1 in Appendix section of this chapter. The PCR products were separated on 0.8% or 2% agarose gel. The products from genotyping PCR were confirmed to be specific by sequencing analyses.

3.2.6 RNA isolation

Genotype confirmed mice were sacrificed following animal ethics committee guidelines. DEPC-treated equipment and RNase-free surfaces were used during all dissection procedures to prevent the degradation of RNA. The freshly harvested brain tissue was used for RNA extraction, or tissues were frozen in liquid nitrogen and stored at -80°C. RNA was isolated using the TRIzol method described briefly: tissue samples were macerated and incubated with TRIzol reagent for 5 minutes at room temperature for lysis and complete dissociation of the nucleoproteins complex. 0.2 ml of chloroform per 1 ml of TRIzol reagent was added to the lysed cells and incubated for 2–3 minutes at room temperature. The

samples were centrifuged for 15 minutes at 12,000g at 4°C. The mixture separated into a lower red phenol-chloroform, an interphase, and a colorless upper aqueous phase. The aqueous phase containing the RNA was transferred to a new tube by angling the tube at 45° and pipetting the solution out.

0.5 ml of isopropanol per 1 ml of TRIzol reagent was added to the aqueous phase. After incubation for 10 minutes, the sample was centrifuged at 12,000g at 4°C for 10 minutes. RNA was precipitated as a white gel-like pellet at the bottom of the tube. The supernatant was discarded, and the pellet was resuspended in 1 ml of 75% ethanol per 1 ml of TRIzol reagent. The sample was briefly vortexed followed by centrifugation for 5 minutes at 7500g at 4°C. The supernatant was discarded, and the RNA pellet was air-dried for 5–10 minutes. The pellet was resuspended in 50 µl of RNase-free water and incubated in a dry bath set at 55–60°C for 10–15 minutes for dissolution. The quality of the RNA was assessed using agarose gel electrophoresis and concentrations estimated using Nanodrop 1000 (Thermo Fisher Scientific, Waltham, Massachusetts, USA).

3.2.7 cDNA preparation

The SuperScript III First-Strand Synthesis System for RT-PCR was used to synthesize first-strand cDNA from total RNA. 1µg total RNA, 5uM oligo(dT)20, 1mM dNTP mix, and DEPC treated water to makeup volume to 10µl were combined in a 0.2ml PCR tube and incubated at 65^{o}C for 5 minutes followed by transferring to ice. The cDNA synthesis mix comprising 2X RT buffer, 10mM $MgCl_2$, 0.02M DTT, 40 units of RNase OUT enzyme, and 200 units of SuperScript III RT enzyme was prepared, and 10µl of this mix was added to each RNA-primer mixture. The reaction was incubated at 50^{o}C for 50 minutes. The reaction was terminated by incubation at 85^{o}C for 5 minutes and rapid chilling on ice. 1µl RNase H

was added to each tube, and the reaction was incubated at 37°C for 20 minutes. The synthesized cDNA was stored at -20°C

3.2.8 Cloning and site-directed mutagenesis

Total cDNA amplified from the mouse brain was used as the template. Using *Efhc2* gene-specific primers -wild type, floxed, and loxed cDNA constructs were PCR amplified using primers detailed in Table 2, Appendix 2. The products were confirmed by sequencing and cloned into the pEGFP-C2 vector for expression using EcoRI and XhoI sites using primers in Table 4, Appendix 2. The clones were confirmed by restriction digestion, insert release, and sequencing. The cDNA constructs with domain truncations were generated by site-directed mutagenesis. This was carried out using QuikChange® II XL mutagenesis reagents (Stratagene, California, USA). The mutagenic oligonucleotide primer pairs mentioned in Table 3 of Appendix 2, were designed to incorporate point mutations in the cDNA. The sequence of cDNA was confirmed using primers in Table 4 of Appendix 2.

3.2.9 DNA sequencing

PCR amplification was carried out under the following conditions: initial denaturation at 94°C for 5 minutes, followed by 40 cycles of denaturation at 94 °C for 30 seconds, annealing at 55-60°C for 30 seconds and extension at 72°C for 30 seconds, followed by a final extension at 72°C for 10minutes.PCR reactions of 20 µl-volume containing 100ng of genomic DNA, 5pmol of each forward and reverse primers, 200µM of each dNTP, 1.5mM MgCl2, buffer, and 1U of Taq DNA polymerase (New England Bio Labs, Massachusetts, USA or Kapa Biosystems, Massachusetts, USA). Amplified products were checked using agarose gel electrophoresis. Before

sequencing, the PCR products were purified using MultiScreen-PCR96 filter plates.

PCR-amplified products were single strand amplified by cycle sequencing using 1µl of BigDye Terminator v3.1 Cycle Sequencing reaction mix (Applied Biosystems, Massachusetts, USA), 1X sequencing buffer, 0.25µM primer, and 3µl of purified PCR product in a 20µl volume. The following cycling conditions were used: initial denaturation at 95°C for 1 minute, followed by 25 cycles of denaturation at 96°C for 10 seconds, annealing at 50°C for 5 seconds and extension at 60°C for 4 minutes, and a final hold at 4°C. The samples were purified by alcohol precipitation by adding 16 µl of chilled autoclaved deionized water and 64 µl of chilled 95% alcohol to each well of the sequencing plate. The sequencing plate contents were mixed and incubated at room temperature for 30 minutes. This was followed by centrifugation at 2500g for 30 minutes. The precipitated DNA was washed with 150µl of 70% alcohol and centrifugation at 2000g for 10 minutes. The plate was air-dried to remove all residual alcohol. The DNA was denatured at 95°C in 10µl formamide per well. These denatured single-stranded amplified products were Sanger sequenced using an automated DNA sequencer, DNA Analyzer 3730. The sequence traces were aligned to the reference gene sequences obtained from the Genbank database, and the variants were identified using SeqMan 5.01 (DNASTAR, Madison, Wisconsin, USA).

3.2.10 Cell Culture, transfection, and synchronization

HEK293T cells were maintained in DMEM (high glucose, Sigma, D1152) supplemented with 10% heat-inactivated fetal bovine serum (MP Biomedicals, USA, 29167), 2mM L-glutamine (Sigma, G7513), and antibiotics (100U/ml Penicillin and 10mg/ml streptomycin; Sigma, P4333) in a humidified atmosphere of 5% CO2 at 37°C. Cells were grown to 60-

70% confluence in 6-well culture dishes or on poly-L-lysine (Sigma, P2658) coated 18 mm glass coverslips and transiently transfected as follows: An hour before transfection, 10% complete DMEM was replaced with DMEM with 2 mM L-glutamine. Transfections were carried out using Lipofectamine®2000 (Invitrogen, Thermo Scientific, Massachusetts, USA) at plasmid DNA: Lipofectamine ratio of 1:1.5. After 5-6 hours, transfection media was replaced by 10% complete DMEM. Cells were harvested for lysate preparation or fixed for imaging after 24-48 hours of transfection. To synchronize cells, untransfected and transfected HEK293T cells were grown to 60% confluence and incubated in a culture medium containing 2 mM thymidine (Sigma) for 14 hours. After two washes in plain DMEM, cells were cultured for 9 hours in DMEM having 24 μM deoxycytidine (Sigma) to achieve mitotic enrichment.

3.2.11 Western analysis

C57BL/6 mice aged 1 and 5 months were used for the study. Animals were decapitated, whole brains were dissected on ice. The tissue was held in ice-cold 1X PBS on a filter paper on ice; cortex, hippocampus, and cerebellum were separated using sterile scalpels. The severed tissue was weighed and resuspended in RIPA buffer. The tissue was homogenized using a handheld homogenizer in ice. The solution obtained was centrifuged at 13000 RPM for 30 minutes at 4°C. The supernatant collected was used as tissue lysate, snap-frozen in liquid nitrogen, and stored at -80°C.

Total protein concentration in the lysates was estimated using the bicinchoninic acid assay (BCA) reagents (Sigma). For Western analysis, aliquots of 50- 100μg of protein were mixed with SDS gel-loading buffer, boiled for 5 minutes, and then resolved electrophoretically on a 10% or 8%

SDS-polyacrylamide gel (Laemmli, 1970). The proteins on the gel were subsequently electrotransferred (Bio-Rad, USA) to equilibrated nitrocellulose membrane (AmershamProtran, GE or BioTrace™ NT, Pall life sciences) at constant 400 mA for one hour and fifteen minutes. The transfer was confirmed by Ponceau S staining followed by washes in distilled water. Blocking was performed using 3% Skimmed milk powder (Sigma) and 2% Bovine Serum Albumin (BSA, Sigma) in 1X PBS with 0.05% Tween 20, for 12 hours at 4°C. The membrane was probed with a primary antibody for 10 hours at 4°C. The membrane was washed in 1X PBS with 0.05% Tween20 at room temperature for 5 minutes. The blot was further incubated for one hour at room temperature with a secondary antibody in 1X PBS containing 1% BSA, 1% skimmed milk powder, and 0.05% Tween 20. The membrane was washed twice for 10 minutes each at room temperature in 1X PBS and 0.05% Tween 20. The protein bands were detected using an enhanced chemiluminescent substrate for HRP (Pierce-Thermo Fischer Scientific, USA or Clarity™ Western ECL Substrate - Bio-Rad).

The antibodies used in this study are - rabbit polyclonal anti-EFHC2 antibody (Prestige Antibodies, Sigma) at a dilution of 1:1000, anti-gamma tubulin (monoclonal, mouse raised; Sigma, T5326), anti-alpha tubulin (monoclonal, mouse raised; Sigma, T9026), anti-Flag (monoclonal, mouse raised; Sigma, 1804) and anti-GFP (polyclonal, rabbit raised; Thermo Fisher Scientific, A-6455), anti-IgG from mouse (Sigma, I5381), anti-IgG from rabbit (Sigma, I5006), goat anti-rabbit Alexa flour 488 (Thermo Fisher Scientific, A11008), goat anti-mouse Alexa flour 568 (Thermo Fisher Scientific, A11001), goat anti-rabbit IgG – HRP (Genei, Bangalore, India), goat anti-rabbit IgG – HRP (Genei, Bangalore, India).

3.1.12 Immunohistochemistry

Animals were anesthetized with halothane inhalation anesthetic and transcardially perfused with ice-cold 1X PBS and 4% PFA (pH – 7.4). Brains were removed, post-fixed overnight in 4% PFA, and sequentially equilibrated in 15% and 30% sucrose in 1X PBS. Tissues were mounted in an OCT tissue freezing medium (Leica Biosystems, Illinois, USA). Coronal sections of 20-30µm were obtained using cryostat at -20°C (Leica, CM 3050S, Wetzlar, Germany).

Free-floating sections were collected in an anti-freeze medium, and pre-mounted sections were collected on gelatin-coated slides. The sections were thawed to room temperature, washed three times in 1X PBS for 5 minutes each. Antigen retrieval was performed in citrate buffer (pH 6) at 90°C for 15-20 minutes. Blocking was performed for 2 hours at room temperature using 10% normal goat serum (NGS) (MP Biomedicals, Santa Ana, California, USA) in 1X PBS and 0.05% Tween 20 (0.05% PBST). The sections were washed in 0.05% PBST and incubated overnight at 4°C with a primary antibody in 1% blocking solution. The antibodies used in this study were - rabbit polyclonal anti-EFHC2 antibody (Prestige Antibodies, Sigma) at a dilution of 1:400, mouse monoclonal anti-NeuN antibody (MAB377, Merck), anti-parvalbumin (P3088, Sigma), Anti-GFAP. Subsequently, sections were washed three times for 10 minutes each at room temperature in 0.05% PBST. They were incubated for one hour with 1:500 dilution of secondary antibodies conjugated to Alexa 468 or Alexa 568 (Molecular Probes, USA), followed by washes and DAPI staining. Free-floating sections in 1X PBS were mounted on gelatin-coated slides, air-dried, and washed in distilled water. Sections were cover- slipped using 70% glycerol in 1X PBS and observed at 20X and 63X using a confocal microscope.

3.2.13 Co-immunoprecipitation

HEK293T cells were with pEGFP-C2-mEfhc2 wildtype and mutant constructs. 36-48 hours after transfection, lysates were prepared in the following IP lysis buffer – 50mM Tris at pH-7.4, 150mM- NaCl, 1% Triton-X 100, 10mM NaF, 1mM Na3Vo4, PIC ultra-complete (Roche, Sigma). The lysates were clarified at 16000g at 4°C for 30 minutes. Supernatants were collected as lysate for immunoprecipitation (IP). Dynabeads (Protein-G, Thermo Fisher Scientific) were washed three times in PBS, incubated with 1ug of the antibody or IgG isotype control. Antibody binding was done at room temperature for 10 minutes. Antibody-bound beads were incubated with corresponding lysates, incubated on an end-to-end rotor at room temperature for 30 minutes. Supernatants were collected and stored. The beads were washed twice with 0.05% PBST for 2 minutes each. Each wash was collected and stored. For the final wash step, wash buffer was added, and beads were carefully transferred to a fresh tube; the supernatant was collected and eluted with 20ul of elution buffer or 20ul of 2X-SDS loading dye without β-ME or with 6X-loading dye with β-ME. The elute was collected by boiling the beads directly in the dye. 5-10% of the inputs, washes, and elutes were loaded on a 10% SDS PAGE for Western blotting.

3.2.14 X-gal Staining on adult mouse brain sections

Animals were anesthetized with an inhalation anesthetic halothane and were transcardially perfused with ice-cold 1X PBS and 4% PFA (pH 7.4). Brains were removed post-fixed overnight in 4% PFA. For cryoprotection, the samples were equilibrated in 15% sucrose for 12-14 hours at 4°C. They were transferred to 30% sucrose and incubated overnight at 4°C. Tissues were embedded in the OCT tissue freezing medium (Leica Biosystems, Illinois, USA). Coronal sections of 20-30µm were obtained

using cryostat -20°C (Leica, CM 3050S). Cyro-sections were collected on gelatin-coated slide-glass. Slide-glass was washed once with the staining buffer (2 ml of 1 M MgCl2, 1 ml of 10% sodium deoxycholate, 1 ml of 20% NP-40 with 800 ml distilled water). Slide-glass was incubated with 1 mg/ml X-gal powder dissolved in the staining buffer. The staining buffer was supplemented with 5mM potassium ferricyanide and 5mM potassium ferrocyanide. The slides were placed in a moist chamber for about 28-35 hours at room temperature until color development. The stained sample was incubated with 1X- filtered PBS for five minutes per wash. Slides were washed three times. Slides were then dehydrated in increasing concentrations of ethanol - 50%, 75%, 90%, 100%, for two minutes each (Kokubu & Lim, 2014). The slide was mounted in DPX mountant (HiMedia, Bengaluru, India) and imaged at 10X,20X, and 40X.

3.2.15 Statistical analysis

All the statistical comparisons were made using GraphPad Prism5. Statistical tests conducted were one-way ANOVA followed by Dunnett's test for multiple comparisons. Results are shown as mean ± standard error of the mean (SEM). Differences between groups were considered statistically significant for $P \leq 0.05$.

3.3 Results

3.3.1 EFHC2 is expression in the mouse brain

We investigated the expression and localization of EFHC2 in the mouse brain using Western analysis, RT-PCR and immunohistology. EFHC2 has a theoretical molecular weight of 87kDa. In western analysis using mouse tissue lysates, immunoreactivity to EFHC2 was detected at around 75kDa (Figure 6a). Along with the mouse brain, EFHC2 expression was detected in lungs, testis, liver, kidney, spleen, and heart (Figure 6b). In the mouse

brain, Efhc2 transcript was detected both embryonic and post-natal stages at E11.5, P0, P7, P14 and P21 (Figure 6c)

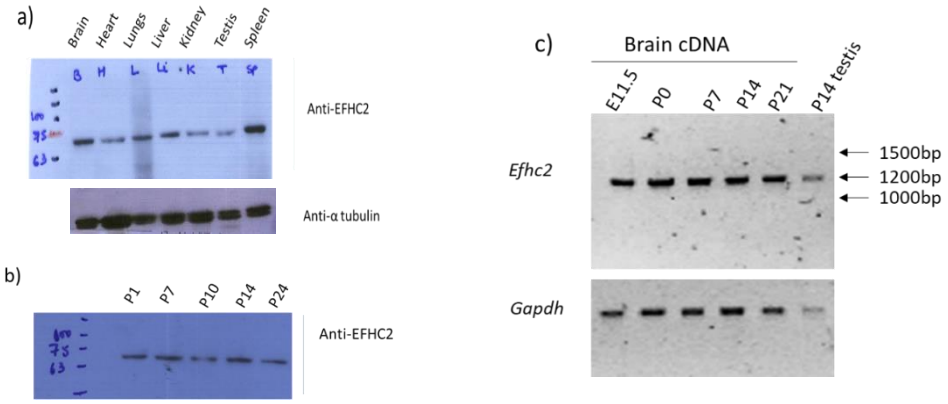


Figure 6: (a) Immuno-reactivity of mouse tissue lysates to anti-EFHC2 antibody. **(b)** Western analysis for expression of EFHC2 in brain lysates of mice across various age groups. **(c)** RT-PCR analysis showing the expression of *Efhc2* in embryonic and post-natal mouse brain tissues. Efhc2 was also detected in the mouse testis.

The expression of EFHC2 in brain regions and specific cell types was analyzed using immuno-histochemistry. In the coronal sections of the adult mouse brain, EFHC2 had a wide-ranging expression. EFHC2 co-stained cells in the cerebral cortex with NeuN, a neuronal marker. Signals were detected across the various layers in the cerebral cortex with abundant expression in the deeper layers (Figure 7a).

Staining for EFHC2 was also observed in the hippocampus, dentate gyrus, hilus, amygdala, hypothalamus, and corpus callosum. The expression was noticed in the cell body, cell cytoplasm, and along the axons. EFHC2 was present in the Purkinje cell layer with projections into the granule cell layer in the cerebellum. The staining in the molecular layer was reduced compared to that of the Purkinje cell layer (Figure 7b). In the brain ventricles, EFHC2 expression was seen in the cells lining the ventricles

and in the choroid plexus, composed of ependymal cells. Also, signals for EFHC2 were detected in the subventricular and ventricular zones.

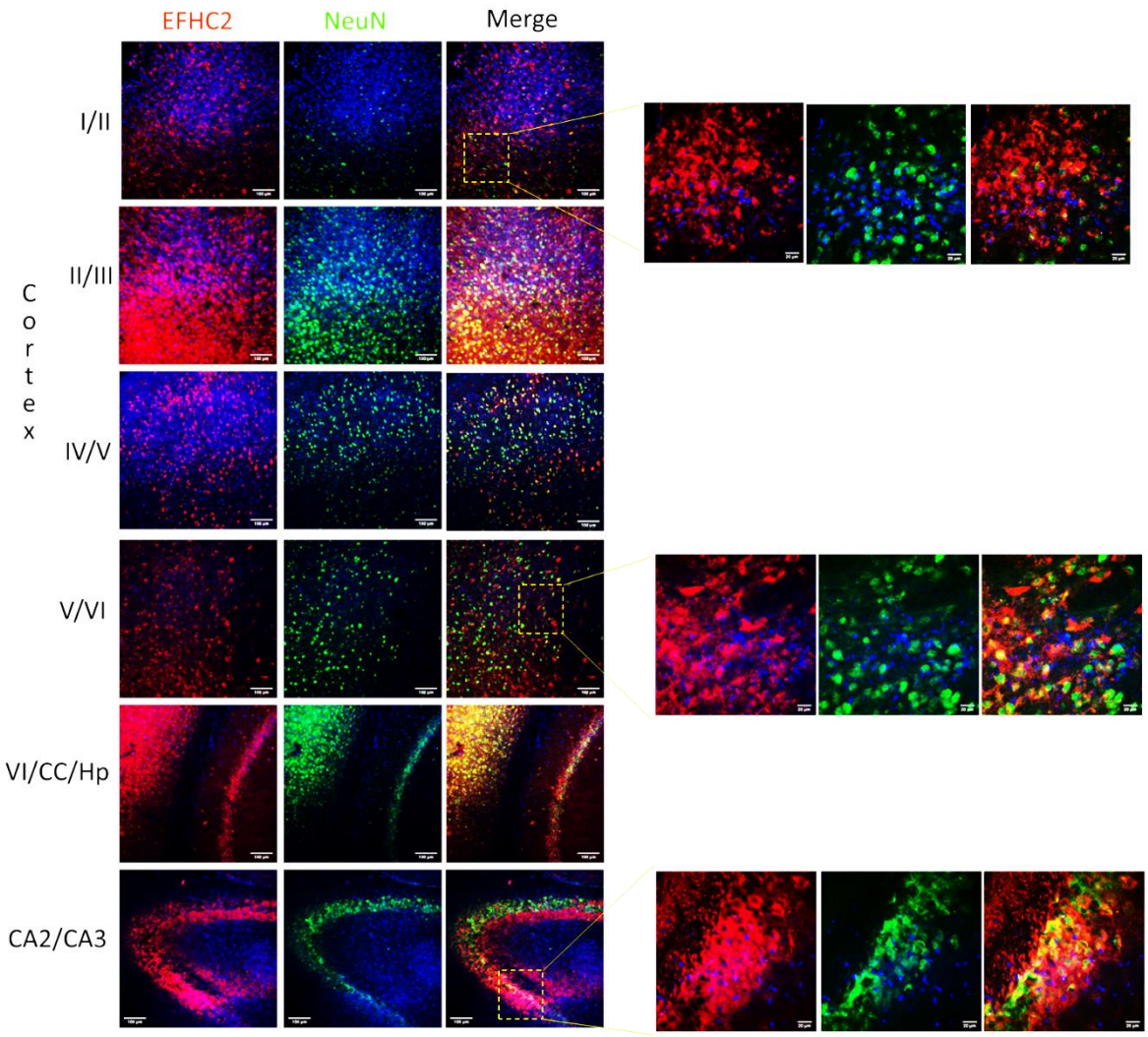


Figure 7: (a) Immuno-histochemical analysis of adult mouse brain coronal sections stained with anti-EFHC2 (red) and anti-NeuN (green) imaged at 20X. The inset are images captured at 63X across cortex and hippocampus

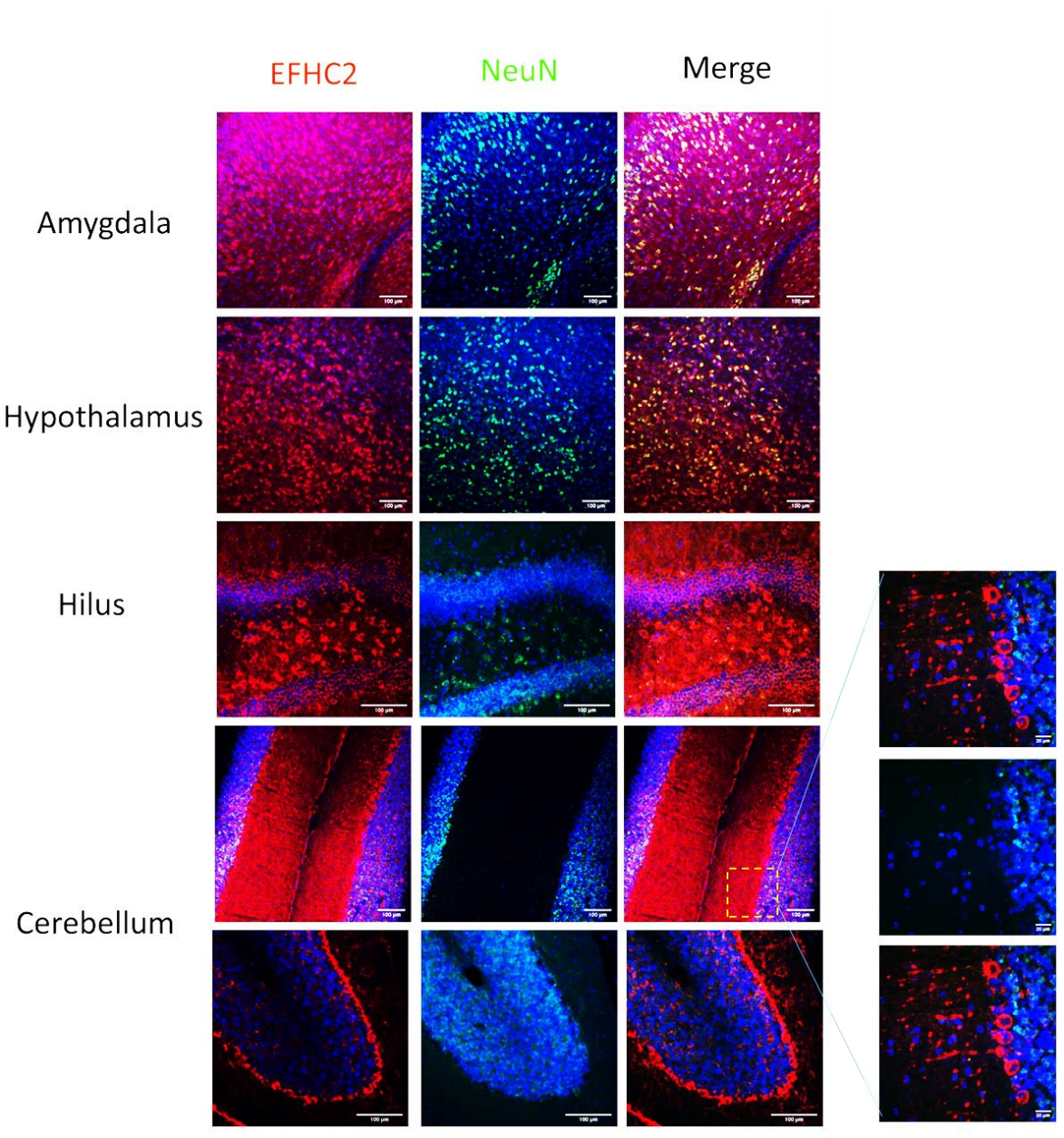


Figure 7: (b) Immuno-histochemical analysis of adult mouse brain coronal sections stained for EFHC2 (red) and NeuN (green) imaged at 20X at hilus, hypothalamus, amygdala, and cerebellum. The inset are images captured at 63X at cerebellum showing the Purkinje cells stained for EFHC2.

Across various mouse brain regions, EFHC2 was also seen in cells devoid of NeuN staining. Considering the possibility of expression in glial cells and astrocytes, tissue sections were co-stained with antibodies against

EFHC2 and GFAP (Figure 8). EFHC2 staining did not overlap with GFAP in either the cell body or the glial cell fibers across the brain regions – cortex, thalamus, striatum, ventricular zone, corpus callosum, hippocampus.

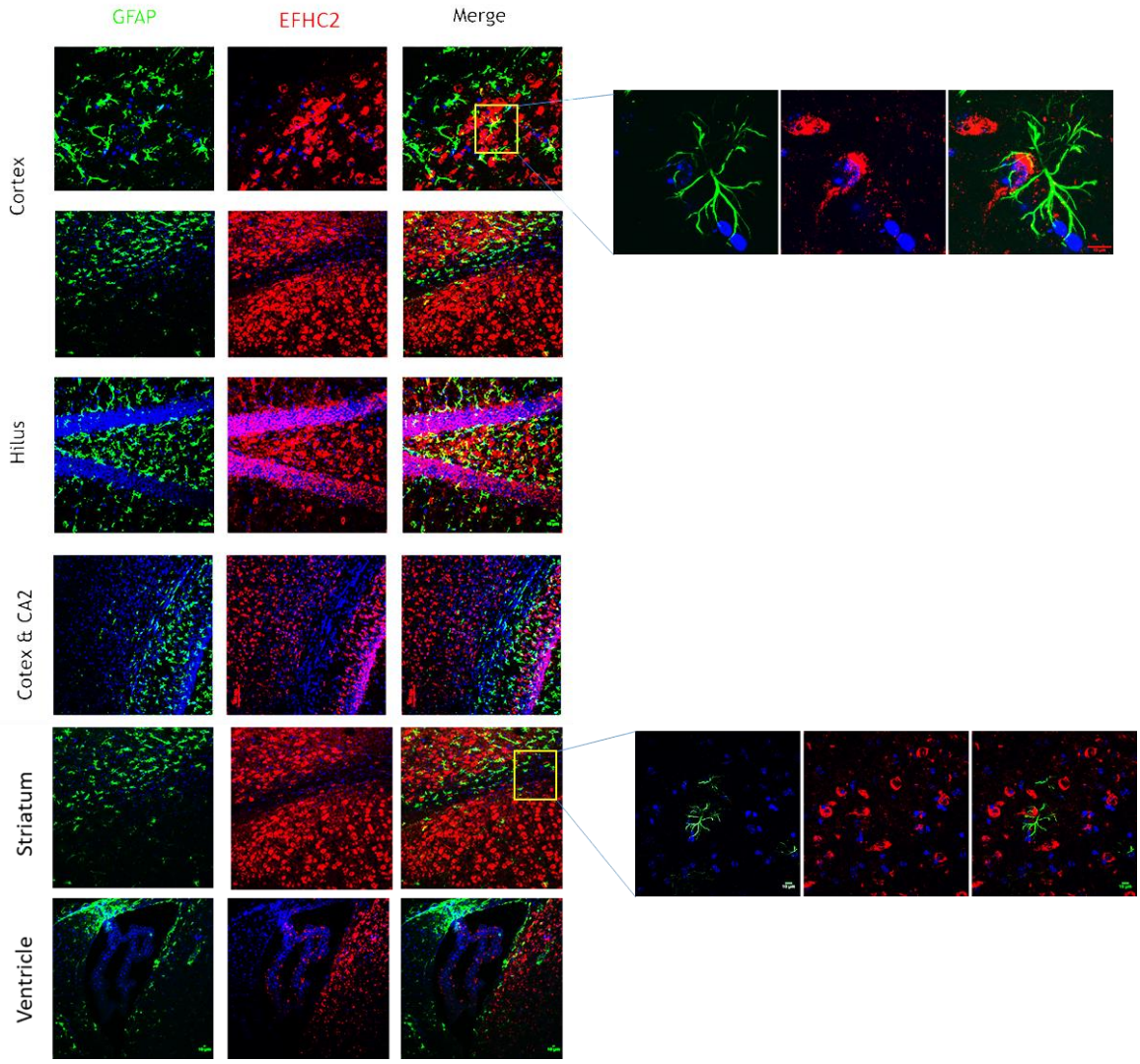


Figure 8: Immuno-histochemical analysis of adult mouse brain coronal sections stained for EFHC2 (red) and GFAP (green) imaged at 20X at cortex, hilus, hypothalamus, striatum, and lateral ventricle. The inset are images captured at 63X at cerebellum showing the glial cells and cells stained for EFHC2.

EFHC2 stained the Purkinje cells in the cerebellum. In the cortex, EFHC2 was observed in cells that morphologically resembled interneurons. So, we asked if EFHC2 is expressed in inhibitory interneurons by co-staining with parvalbumin (PV). Interneurons are GABAergic and vital for optimal excitatory-inhibitory balance in the cortical circuits that drive behavioral responses. In the cerebellum, parvalbumin expression is predominant in the Purkinje cell and molecular layers. Interestingly, EFHC2 expression was observed in the cerebellum with parvalbumin. Also, several cells in the cortex were co-stained for EFHC2 and parvalbumin (Figure 9 and 10).

3.3.2 *Efhc2* cDNA cloning

The *Efhc2* cDNA is 2.2Kb in size. Using adult mouse brain RNA as the template, *Efhc2* cDNA was synthesized by RT-PCR with gene-specific primers binding in exon 1 and exon 15. The cDNA amplified was 2.2Kb in size (Figure 11). The identity of the full length of the PCR amplified cDNA was confirmed by sequencing. The pEGFP-C2 vector and the cDNA insert were digested using EcoRI and XhoI and ligated. The frame of insertion of the cDNA into the plasmid was validated by sequencing.

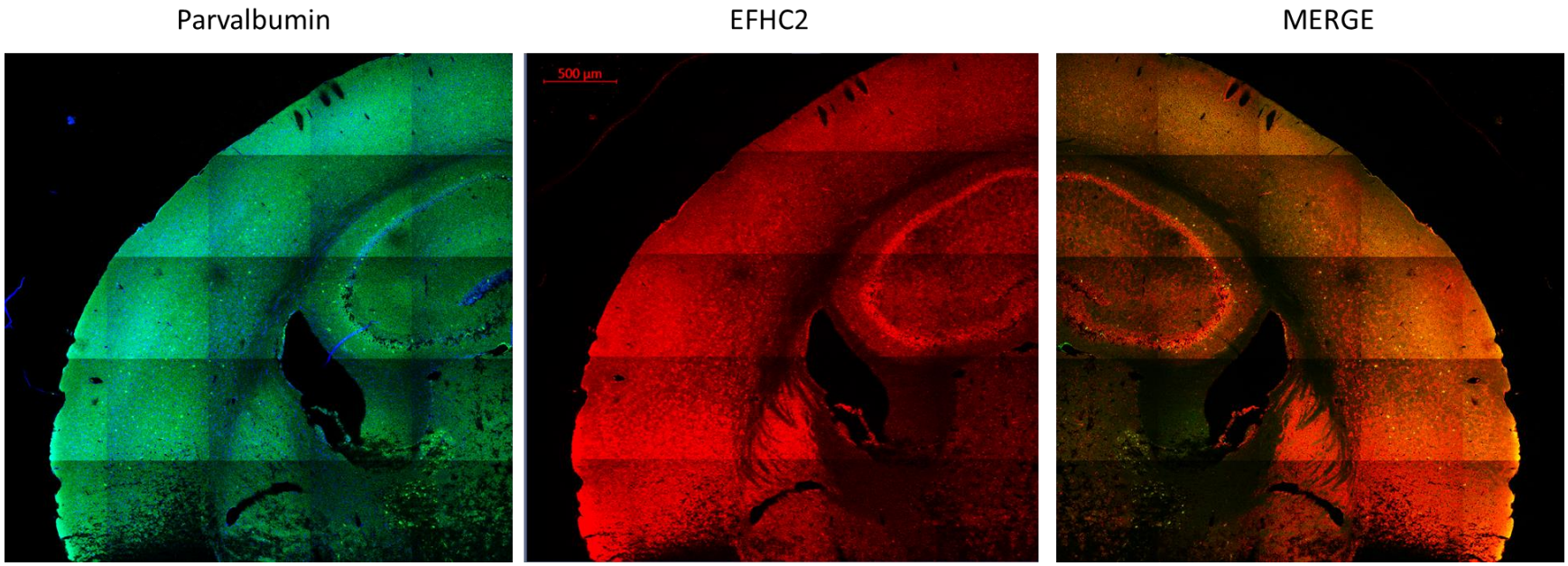


Figure 9: Adult mouse brain coronal sections stained for EFHC2 (red) and parvalbumin (green) – EFHC2 is ubiquitously expressed across regions of the brain – in the cortex, hippocampus, ependymal cells, thalamus are shown here.

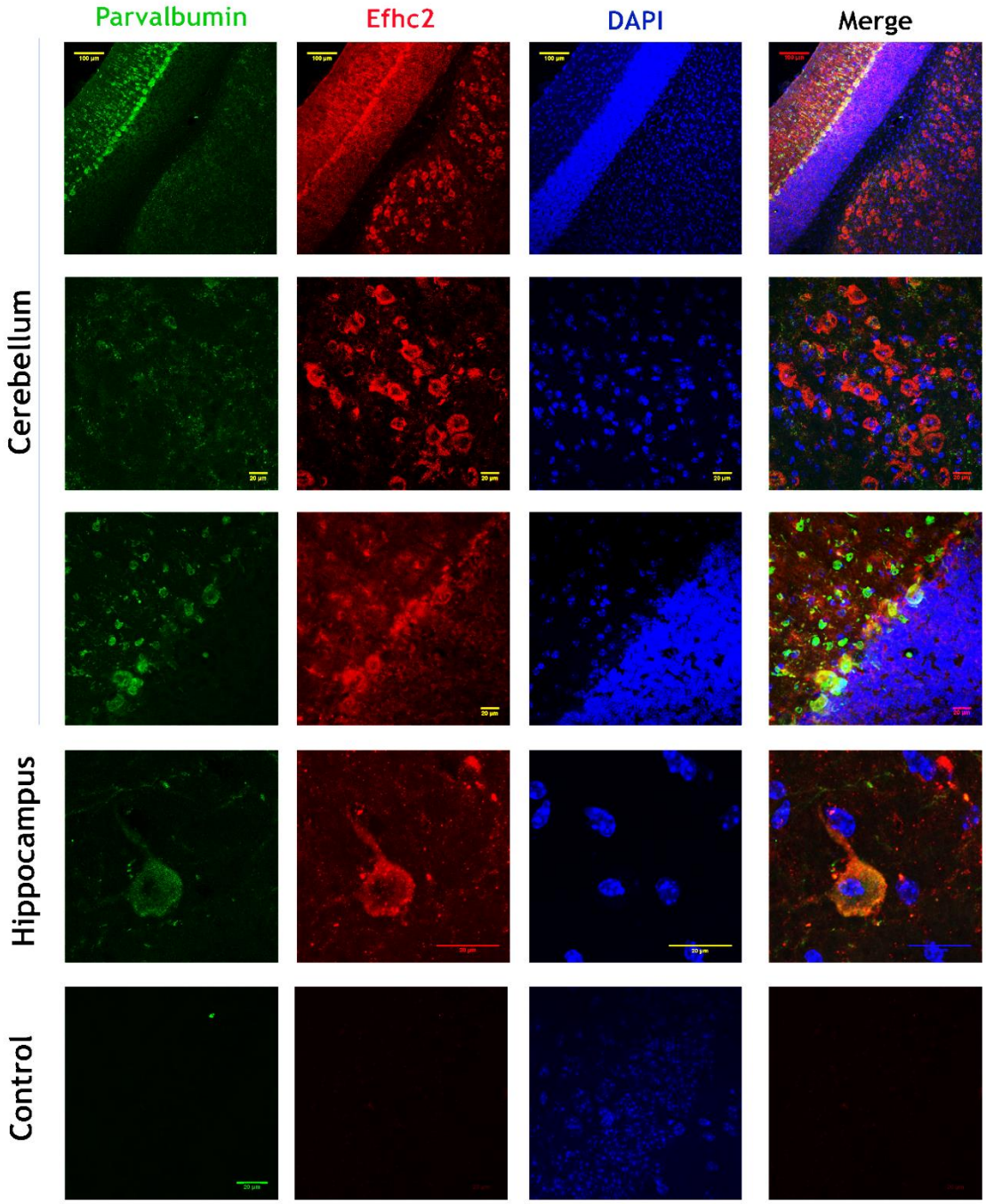


Figure 10: Immuno-histochemical analysis of adult mouse brain coronal sections stained for EFHC2 (red) stained with parvalbumin in the inhibitory interneurons shown in green.

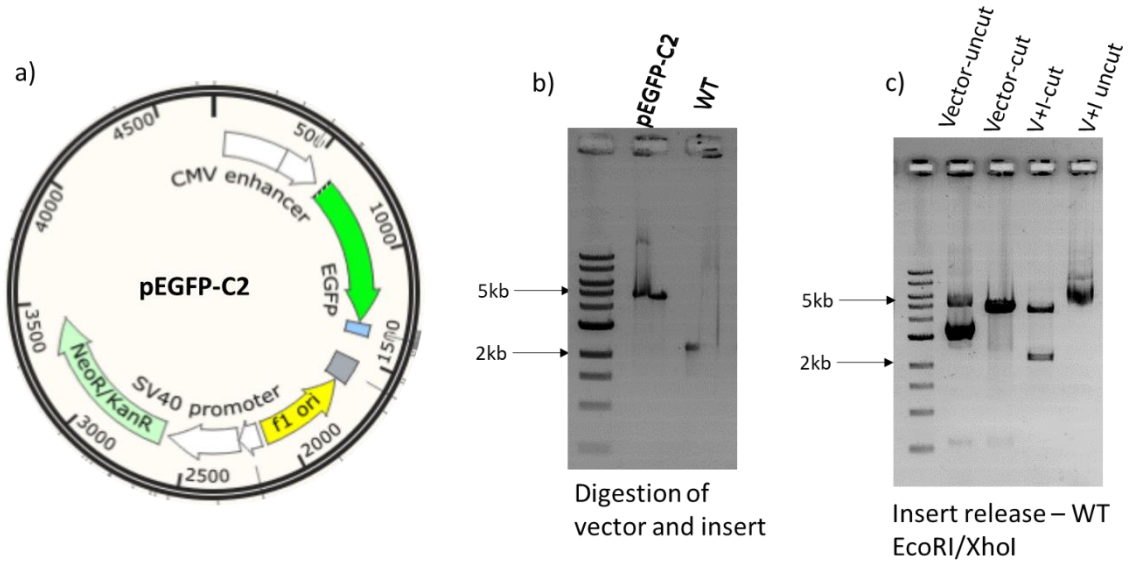


Figure 11: (a) pEGFP-C2 vector was used for cloning *Efhc2* cDNA amplified from adult mouse brain. **(b)** Pre-ligation agarose gel was used to confirm sizes of digested vector and insert. **(c)** Validation of the clones generated by restriction digestion with insert release of a 2.2kb fragment.

3.3.3 Murine EFHC2 is a microtubule-associated protein

Understanding the sub-cellular localization of the protein offers valuable clues about its probable cellular function. We used the GFP-tagged EFHC2 and examined the protein localization in HEK293T cells. The protein was distributed mainly in the cytoplasm and in a few cells, the protein was also present in the nucleus to a lesser extent. In cells undergoing mitosis, the protein co-stained with tubulins -α and -γ at the mitotic spindle (Figure 12) During interphase, we found EFHC2 at the centrosomes with γ-tubulin and through prometaphase, EFHC2 was clustered at the spindle. However, in metaphase, EFHC2 was found at the opposite spindle poles with tubulins -α and -γ. EFHC2 stained the spindle apparatus – the spindle poles and spindle microtubules.

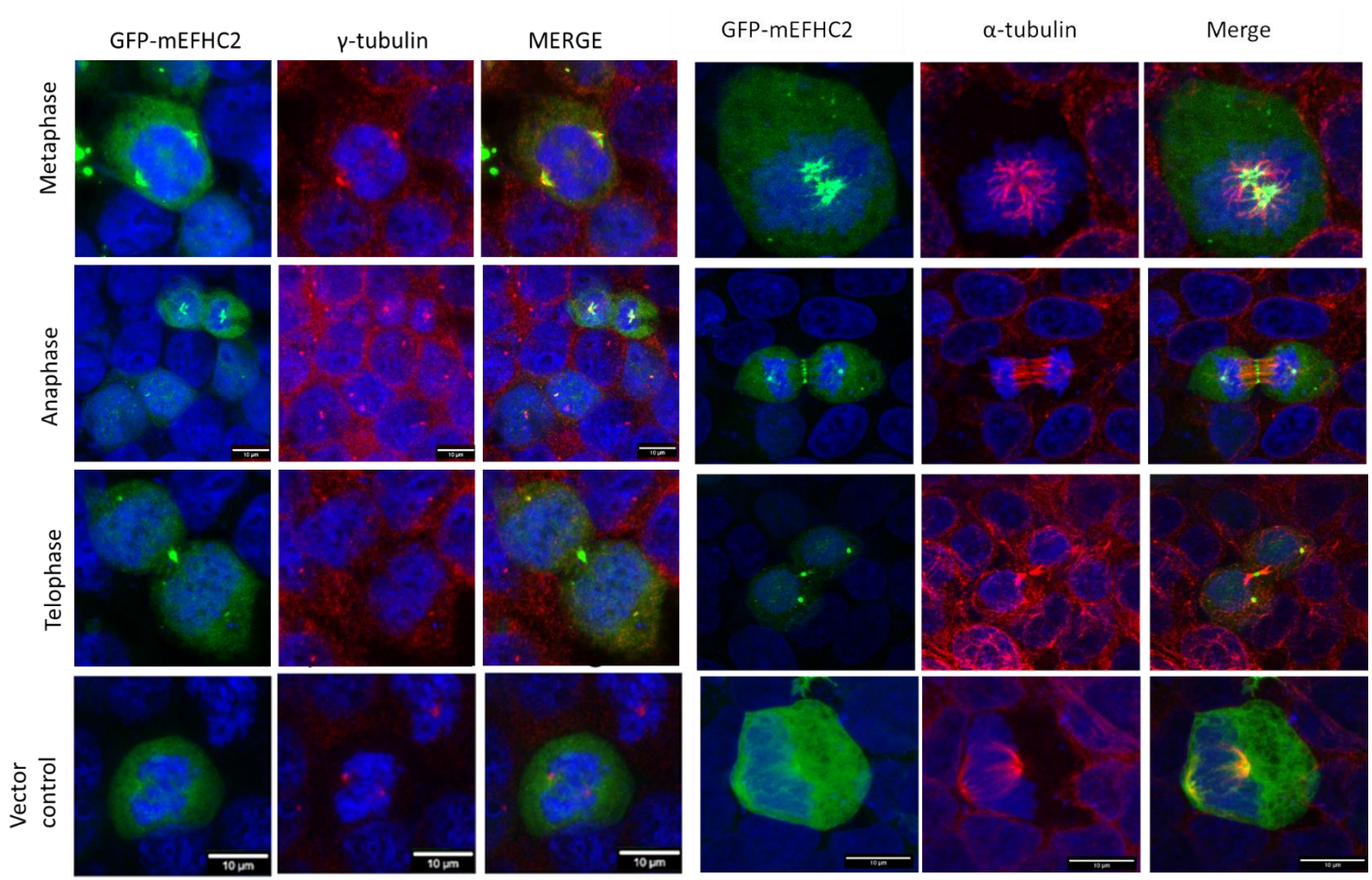


Figure 12: Sub-cellular localization of GFP-tagged EFHC2 in HEK293T cells undergoing cell division. The protein associates to mitotic spindle as shown and co-stains with α- and γ-tubulins.

During anaphase, EFHC2 was retained at the spindle pole. After chromosome segregation, the protein was observed at the centrosomes of the bipolar spindle and at the cleavage furrow, signaling the beginning of cytokinesis. During cytokinesis, the protein was present at the midbody ring.

Organization and rearrangement of the spindle, microtubule cytoskeleton during mitosis controls chromosome segregation, the placement of the contractile ring and the completion of cell cleavage. Consequent to the localization of EFHC2 with the spindles and microtubules throughout cell division, we queried its association with tubulins using co-immunoprecipitation experiments. Using whole-cell lysates of HEK293T cells expressing GFP-EFHC2, IP was conducted with GFP-antibody and in the elute fractions, tubulins -α and -γ were identified. Thus, basis co-localization and IP results, we conclude that EFHC2 is a microtubule-associated protein.

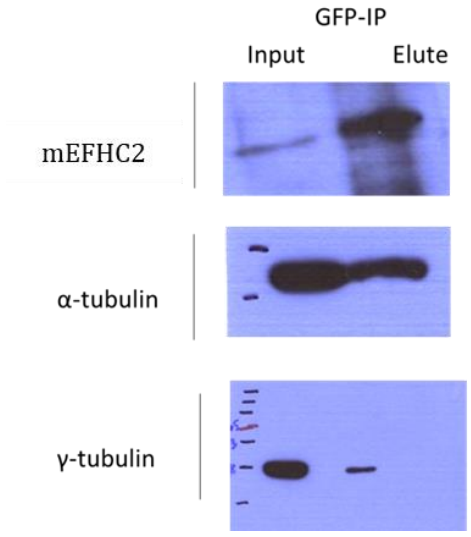


Figure 13: mEFHC2 is a microtubule-associated protein**.** Co-immunoprecipitation of GFP-tagged mouse EFHC2 with tubulins- α and -γ. We detected the tubulin proteins in the elute fractions of EFHC2 immuno-pulldown

3.3.4 DM10 domains in EFHC2 are important for the localization and function

EFHC2 comprises conserved amino acid repeats that form three domains called DM1O domains. The functions of these domains are unknown. The DM10 domains of mouse EFHC2 are similar with respect to the length of the domains and sequence similarity. The first and third domains consist of 108 amino acids, and the second has 143 amino acids. The overall similarity between DM10 domains is around 40%. From the predicted structure of EFHC2, using the Alpha-Fold protein structure database, we find that the DM10 domains have a similar structural organization (Jumper et al., 2021). They have Pleckstrin Homology (PH)-like fold, with seven beta strands, a short 3-4 residue helix after the first strand, and an extended alpha-helical region at the C-terminus. The function of these domains is unknown.

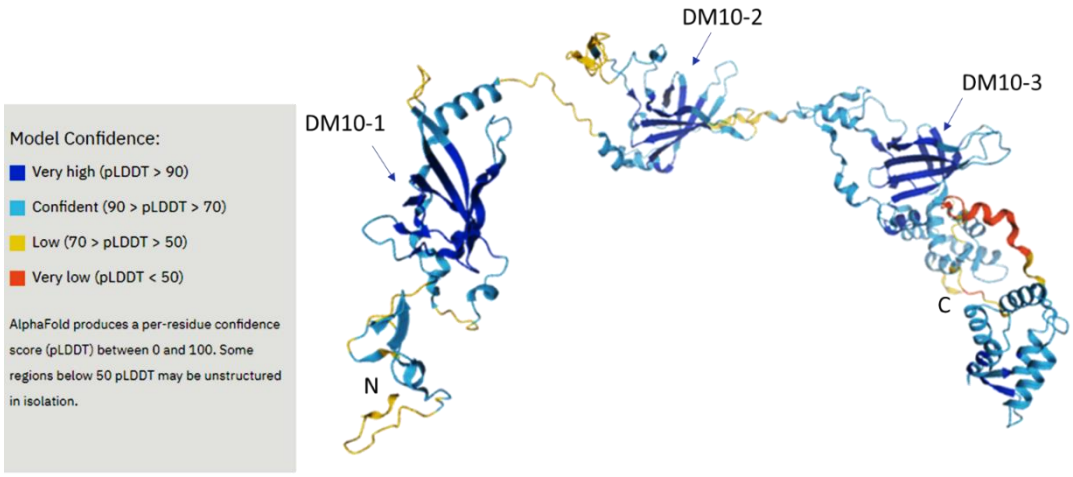


Figure 14: Alpha-fold prediction of three-dimensional structure of EFHC2.

Given the spindle localization and the microtubule association of EFHC2, we wondered if the DM10 domains are critical for localization and function. To test this, we incorporated stop codons after every DM10 domain and generated sub-clones with only the N-terminal and C-terminal portions of the protein with GFP-tag. The localization pattern of these constructs was analyzed and compared to wildtype protein. The mutant p.Phe74Ter and the C-terminal subclone consisting of amino acids 548-750 lack DM10 domains and had a diffused localization in the cell cytoplasm. The mutant proteins were not found at the centrosome, mitotic spindle, or midbody. Consequently, significant mitotic defects were observed.

In cells expressing the constructs with the first DM10 domain (p.Gly189Ter) and the third DM10 domain and the C-terminus (422-750), the protein was broadly distributed in the cell cytoplasm and across the different stages of mitosis, failed to associate with the spindle. As a result, the cells expressing these constructs had significant mitotic defects such as multipolar cells, abnormal spindle, lagging chromosomes, and cytokinesis defects.

The mutants p.Thr374Ter and p.Glu546Ter consist of the first two and all three DM10 domains, respectively. In cells expressing these constructs, we found spindle association of the protein. During cell division, the GFP signal overlapped with α-tubulin. The localization pattern was comparable to the full-length protein. Consequently, fewer cells with mitotic defects were observed.

Among the three DM10 domains, lack of spindle association and localization was observed with constructs having only the first or the third DM10 domain. The presence of all three DM10 domains or the first two

DM10 domains in tandem facilitate the proper localization and function of EFHC2.

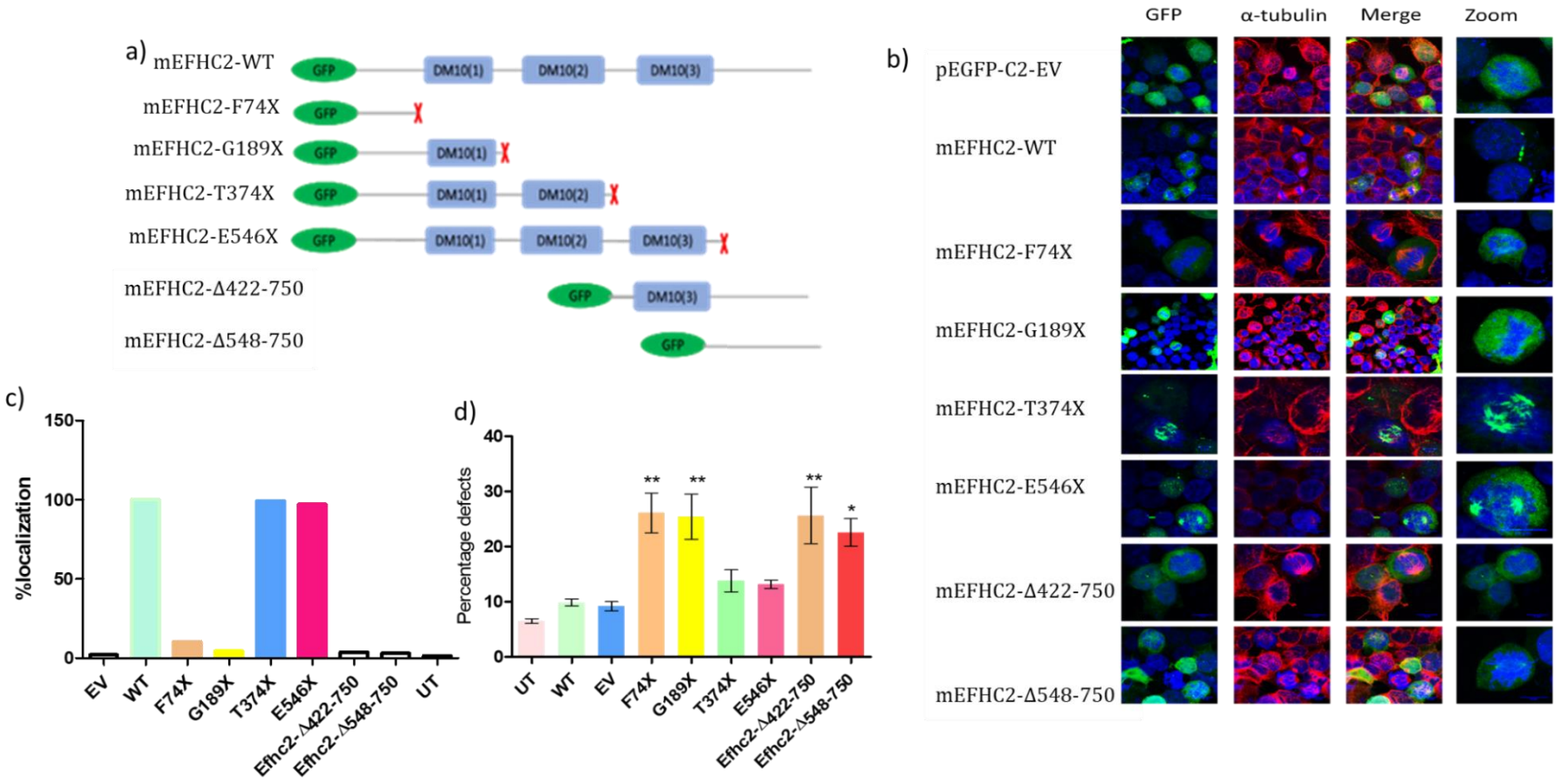


Figure 15: (a) Representation of domain truncations and sub-clones generated. **(b)** Localization of mutant alleles with alpha tubulin during mitotic stages. **(c)** Quantitation of percentage of cells with localization of GFP-tagged constructs to mitotic spindles, centrosomes and midbody. **(d)** Percentage cells with mitotic defects; N=3 and n= around 400 cells/construct; one-way ANOVA, $P < 0.05$; Dunnett's multiple comparison test

3.3.5 Generation and validation of EFHC2 loxed allele

Mice with *Efhc2* exon3 flanked between loxP sites, i.e., *Efhc2*$^{flox/flox}$ mice were received from NIH, USA, through a collaboration with Dr. Shyam. K. Sharan. Progeny from the founder colony of *Efhc2*$^{flox/flox}$ breeders were genotype confirmed. The genotyping PCR resulted in the amplification of a specific band of 327bp in the *Efhc2*$^{flox/flox}$ and *Efhc2*$^{flox/\ Y}$ and 346bp band in *Efhc2*$^{WT/WT}$ (Figure 16b). The identity of these bands was confirmed by Sanger sequencing analysis. The *Efhc2*$^{flox/flox}$ and *Efhc2*$^{flox/Y}$ were viable and fertile with no apparent gross abnormalities.

The *Efhc2*$^{flox/flox}$ and *Efhc2*$^{flox/Y}$ were mated to Tg. β- actin Cre-mice. The Cre-mediated recombination would result in the deletion of exon 3 from the *Efhc2* floxed locus. The resultant litters were confirmed to be heterozygous for the deletion of the critical region – *Efhc2*$^{lox/WT}$ (Figure 16). The heterozygous and hemizygous mutants were intercrossed to expand the mouse colony and generate homozygous mutants of *Efhc2*. The mouse strains were genotype confirmed. The PCR band indicative of loxP mediated recombination was sequence confirmed (Figure 16C). The mutant mice, *Efhc2*$^{lox/lox}$ and *Efhc2*$^{lox/Y}$ were viable, fertile, and with no apparent abnormalities.

The deletion of exon 3 of *Efhc2* was expected to generate several stop codons in exon 4. The introduction of the stop codons would result in an unstable transcript with premature stop codons. This transcript is expected to be degraded through non-sense-mediated decay (NMD). However, if the transcript were to escape NMD, the resultant protein would be 11.6 kDa with 105 amino acids. We performed Western analysis to analyze the outcome of exon 3 deletion.

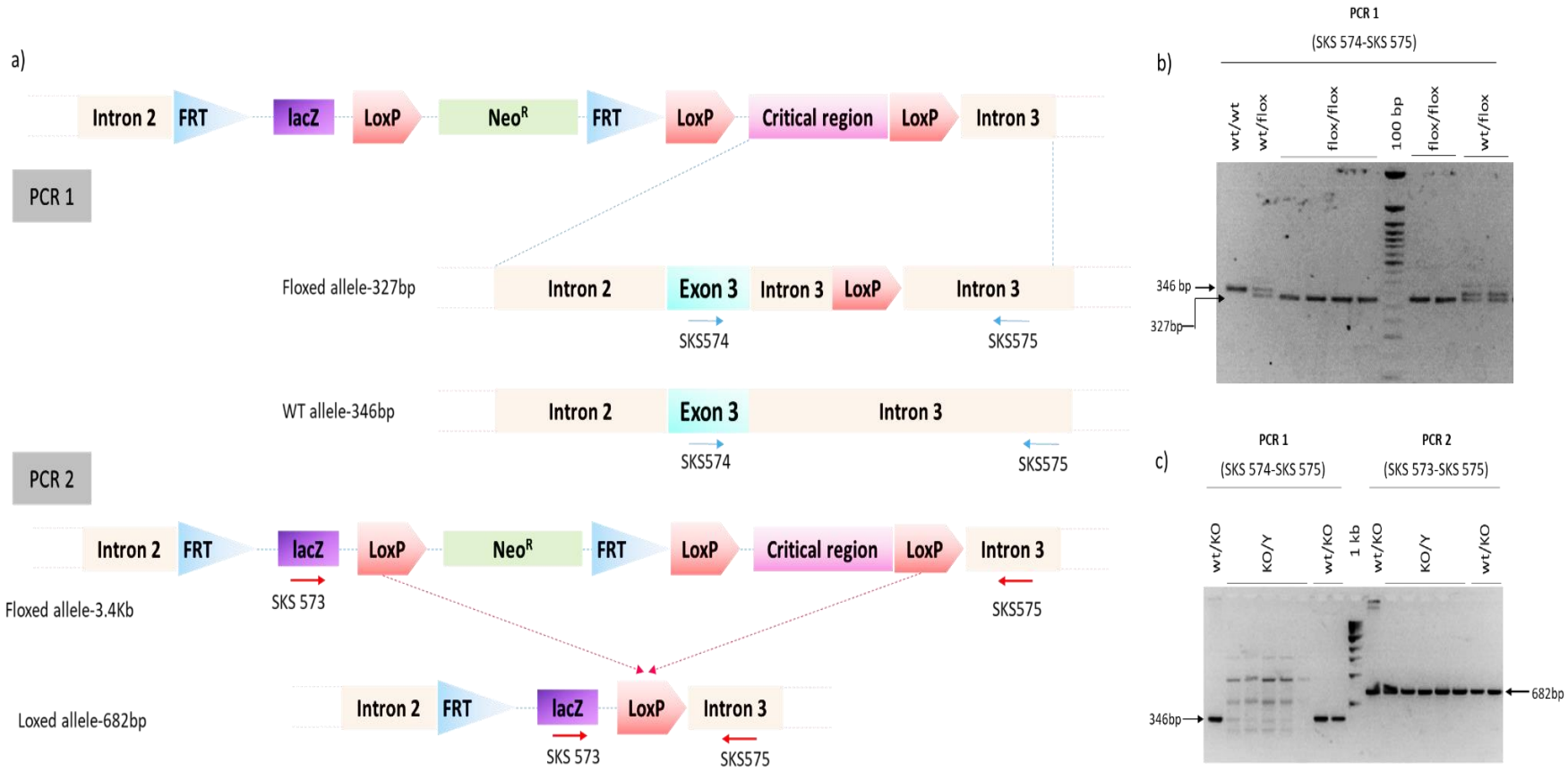


Figure 16: (a) Schematic representation of the strategy with primer binding sites. **(b)** Genotyping for the identification of wildtype and vector integrated floxed allele (tm1a). **(c)** Two-step PCR based genotyping for the identification of wildtype and loxed allele (tm1b)

EFHC2 has a predicted molecular mass of 87kDa. Brain tissues were harvested from *Efhc2*$^{WT/WT}$, *Efhc2*$^{lox/WT,}$ and *Efhc2*$^{lox/Y}$ genotypes. Cortex, hippocampus, and cerebellum were separated, and total protein isolated from these subregions was tested for immunoreactivity against EFHC2 antibodies. In wildtype tissue lysates, EFHC2 was expressed in the cortex, hippocampus, and cerebellum. The expression of EFHC2 in *Efhc2*$^{lox/WT,}$ heterozygotes was comparable to wildtype. To our surprise, in the *Efhc2*$^{lox/Y}$ wherein exon 3 of *Efhc2* was absent from the mouse genome, EFHC2 expression was comparable to wildtype. These results were consistent across independent matings and two different polyclonal antibodies against EFHC2, raised against the N-terminus of the protein (Sigma, HPA034492) and C-terminus (Thermo Fisher Scientific, PA5-31608) (Figure -17a and 17b).

Given the unexpected results in the Western analysis of the mutant alleles, to examine the effect of the putative knockout allele on the function *Efhc2* and ascertain the likelihood of production of a full-length mRNA, the effects on transcription of this presumptive knockout allele were explored.

We extracted RNA from brain tissue of wildtype and loxed mice (Figure 17c) and synthesized cDNA using the gene-specific primer pairs. PCR-amplified cDNA was analyzed using agarose gel electrophoresis and sequencing. cDNA was detected in both the wildtype and mutant brain tissues. However, we noticed a slight reduction in the amplicon size in the *Efhc2*$^{lox/Y}$ by 36 bases (Figure 17d).

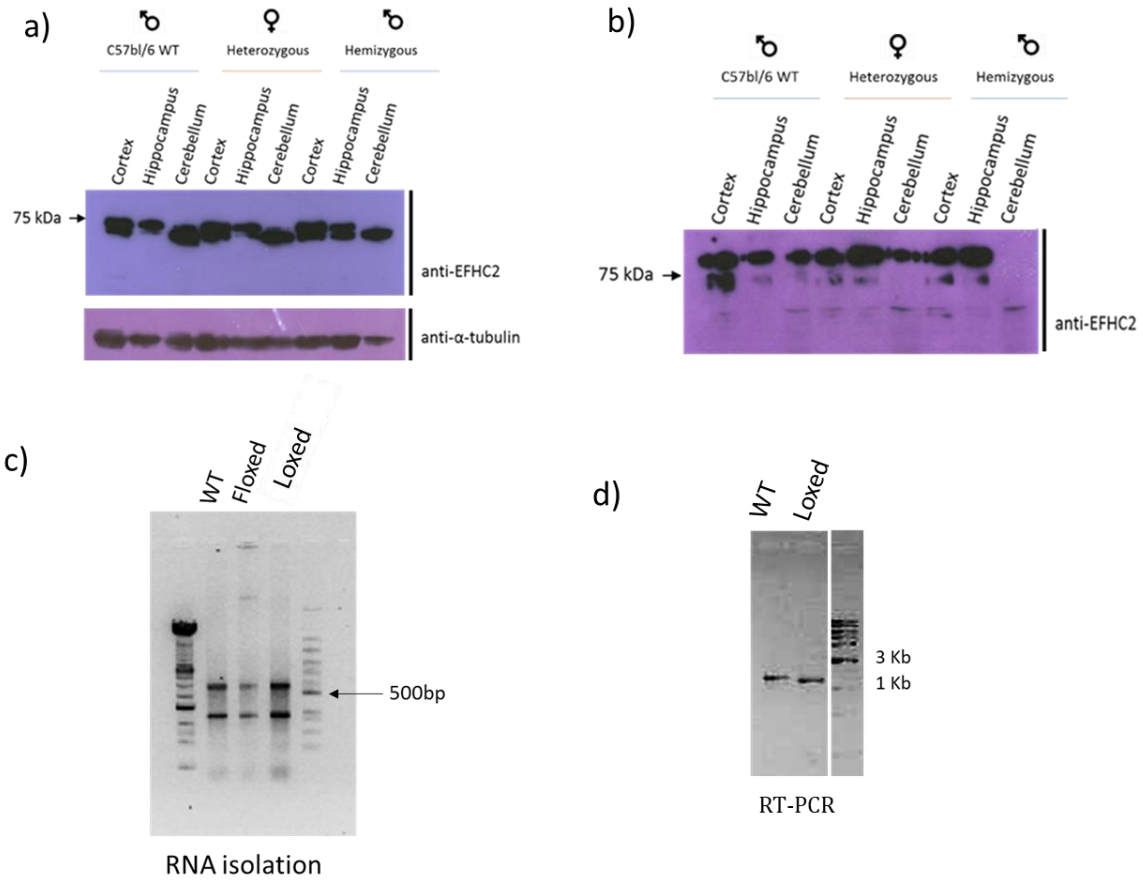


Figure 17: Western analysis of brain lysates from wildtype and loxed mice using two commercial anti-EFHC2 antibodies from – **(a)** Prestige Sigma and **(b)** Thermo scientific. Agarose gel electrophoresis for **(c)** RNA extracted from mouse brain tissues of wildtype and mutants **(d)** RT-PCR amplification of Efhc2 cDNA from mouse brain.

Using multiple primers, we sequenced the entire length of the cDNA from wild type and the loxed mice. For primer sequences, please see Table 4, Appendix 2. In the cDNA from *Efhc2*$^{lox/Y}$, we found that exon 3 (151 bases) was deleted as anticipated. However, a novel transcript with an in-frame insertion of 115 bases in place of exon 3 was detected. Additional sequence alignment studies were conducted to determine the exact origin of the insert 115 bases in length. Results from sequencing determined

precisely that the inserted DNA originates from the gene targeting vector and corresponds to an exon of the *En2*, *Engrailed 2*, originally designed to provide a splice acceptor site. The fusion transcript was generated by activating a cryptic splice donor site located 115 nt downstream in the En2 sequence. This 115 nt insertion offered an additional artificial exon to the *Efhc2* transcript and restored the reading frame resulting in a mutant but near full-length EFHC2 (Figure 18). For clarity, this allele was termed the Efhc2-En2 insertion allele, Efhc2$^{\text{En2-ins}}$ and it represents the outcome of the tm1b allele of *Efhc2*. The pair-wise sequence alignment of EFHC2 wild type and the En2-inserted protein are presented in Appendix 2. Mice harboring the En2-insertion are viable, fertile, and have a life-span comparable to wildtype.

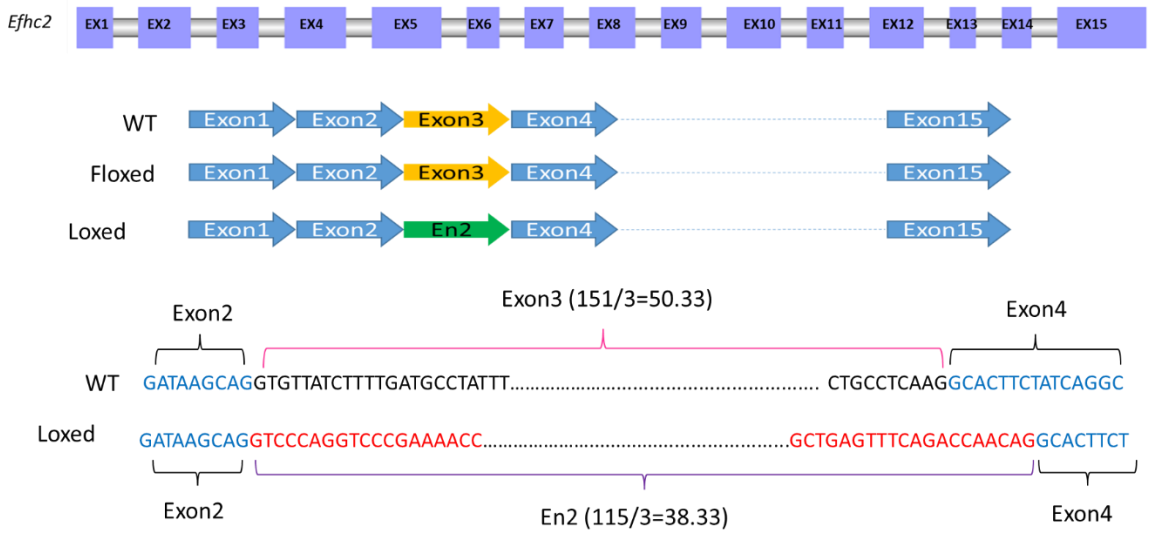


Figure 18: Exon 3 of *Efhc2* is deleted in the *Efhc2* loxed, tm1b allele. During splicing, 115 nt from the Engrailed 2 from the gene-targeting vector is incorporated leading to in-frame insertion.

3.3.6 Cellular characterization of the *Efhc2*-En2-insertion allele

Exon 3 of EFHC2 comprising of 151 bases codes for ~50 amino acids forming a part of the first DM10 domain. We speculated that the mutant protein generated from the Efhc2-exon 3 deletion and En2 insertion alters the protein localization and function. To test this, the Efhc2 cDNA isolated from wildtype and mutant brain tissues were cloned into pEGFP-C2 vector. The identity of the cDNA sequence clones and the frame of insertion were confirmed by Sanger sequencing.

The GFP-tagged wildtype and mutant proteins were overexpressed in HEK293T cells. The localization was studied using immunocytochemistry by co-staining with tubulins. The mutant protein distributed largely in the cytoplasm and did not localize to centrosomes during interphase (Figure 19). In contrast to wildtype protein, the mutant did not localize to the spindle poles or midbody during cell division. The absence of the protein at the cleavage furrow or midbody was a striking observation. The cells expressing mutant protein were frequently observed with an abnormally large nucleus as seen by DAPI staining. We wondered if the lack of localization to spindle-associated structures had any consequence on the protein's function. Using mitotic defects as the readout, cells expressing GFP, GFP-tagged wildtype, and mutant proteins were quantitated for abnormal spindle, multinucleated cells, cells with cytokinetic defects, multipolar spindle, and lagging chromosomes. The mutant protein gave rise to mitotic defects (Figure 20). The moderate mitotic defects could be a consequence of changes in protein folding and structural caused by the differences in the primary amino acid sequence of the mutant EFHC2 protein.

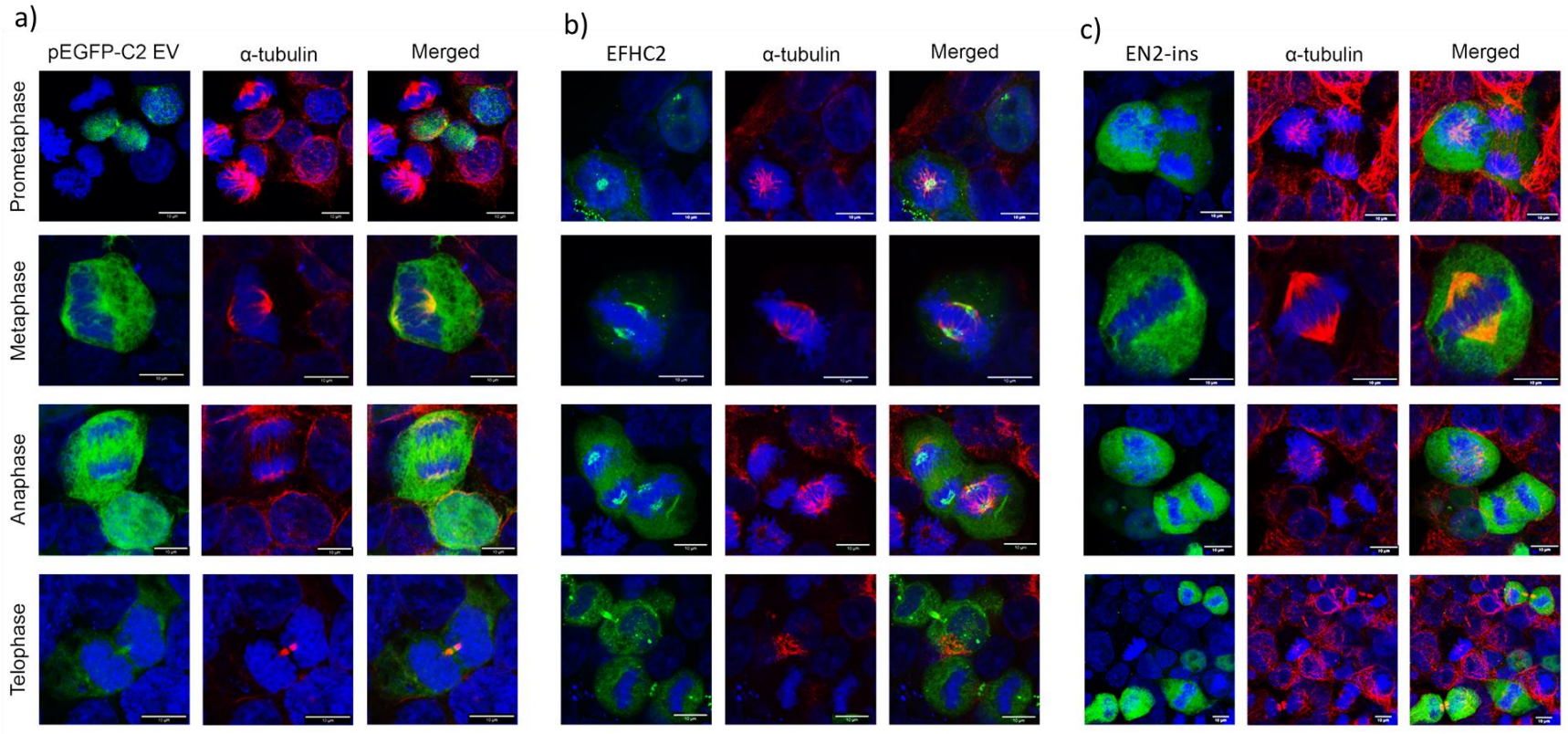


Figure 19: Sub-cellular localization of EFHC2 wildtype and mutants **(a)**Cells expressing vector control. **(b)** GFP-EFHC2 localization in the mitotic stages showing co-localization with alpha-tubulin. **(c)** Sub-cellular localization of En2-insertion mutant showing mislocalized protein.

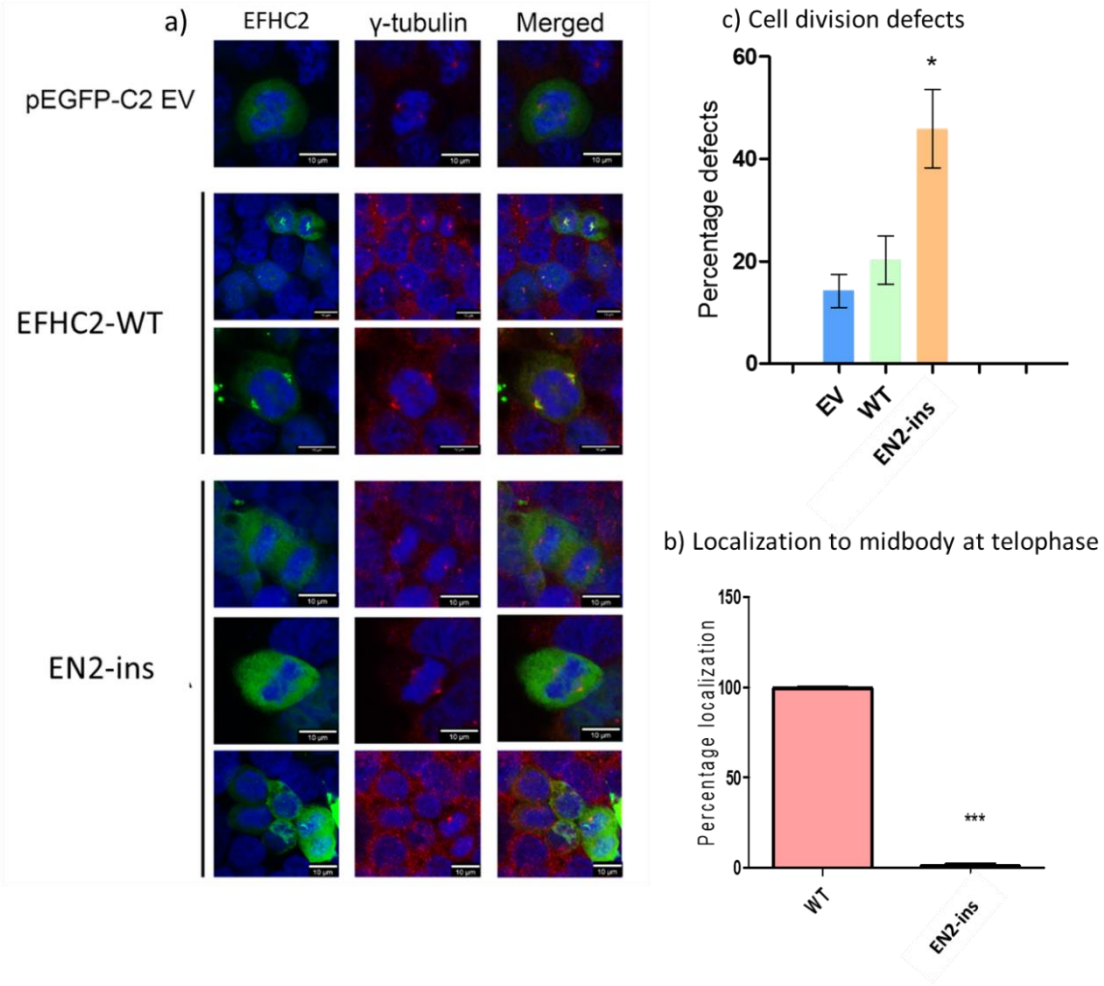


Figure 20: (a) Co-localization of EFHC2 wildtype and mutant with gamma- tubulin. **(b)** Percentage defects on overexpression of wildtype and En2-insertion allele in HEK293T cells (One way ANOVA, $P < 0.05$; Dunnett's multiple comparison test; N=3, n=520-600 cells). **(c)** Percentage localization of wildtype and En2-insertion mutant to midbody during mitosis (unpaired two-tailed T-test; $P<0.0001$; N=3, n=100-150 cells)

3.3.7 Generation and validation of *Efhc2* -flipped and -loxed allele

Our objective was to generate an EFHC2 null allele. Considering that the cryptic splice donor site from the En2 exon induced an unexpected fusion transcript, we converted the knockout first allele to a conditional one to generate the EFHC2 null allele.

The first step was to excise the En2 exon. For this, the *Efhc2*$^{flox/flox}$ mice were mated with mice constitutively expressing the flippase recombinase. On FRT-Flp mediated recombination 6.9kb of the DNA would be deleted and generate the tm1c allele. The deleted segment consisted of the En2-exon, a loxP site, neomycin selection cassette, and the reporter LacZ. The FRT-mediated recombination was confirmed by genotyping PCR resulting in a 1.7Kb band (Figure 21a). The FRT-mediated recombination was confirmed by sequencing the PCR band. In the next step, the flipped *Efhc2* mice were mated to the β-actin driven Cre recombinase-expressing mice to obtain flipped and loxed, tm1d allele. The Cre-mediated recombination would result in the deletion of exon 3 *Efhc2*. The offsprings from the mating were PCR genotyped, and the amplification of a 900bp band confirmed the Cre-mediated recombination (Figure 21a). The PCR band was Sanger sequenced to confirm the deletion of exon 3 in the mutants. The homozygous *Efhc2*$^{\Delta/\Delta}$, hemizygous *Efhc2*$^{\Delta/Y}$, and the heterozygous *Efhc2*$^{\Delta/WT}$ mutants were found to be viable and fertile with no apparent gross morphological defects.

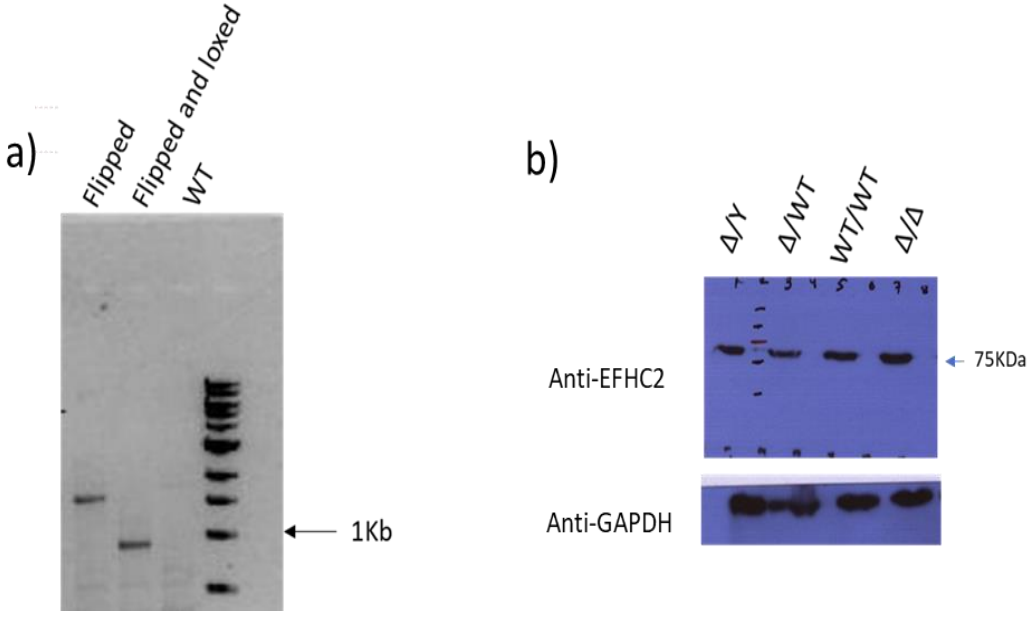


Figure 21: (a) Genotyping PCR bands showing FRT-mediated recombination -1.7kb band in the *Efhc2* (tm1c) alleles and 900bp band in the tm1d allele. (**b)** Western analysis of mouse brain lysates from wildtype, *Efhc2* heterozygous, hemizygous and homozygous potential loss of function alleles.

Western analysis was performed for further validation of the mutants using mouse brain lysates from *Efhc2*$^{WT/WT}$, *Efhc2*$^{\Delta/Y}$, *Efhc2*$^{\Delta/WT,}$ and *Efhc2*$^{\Delta/\Delta}$. In the heterozygous mutants generated, we anticipated two protein bands comparable to wildtype at ~75kDa and 11.6kDa. Only the 11.6kDa protein band was predicted in the homozygous and hemizygous mutants. To our surprise, immunoreactivity was observed at around ~75kDa in all the genotypes tested (Figure 21b).

Given the puzzling results, we validated the deletion of exon 3 of EFHC2 at the genomic DNA by PCR genotyping and Sanger sequencing. Further, the total RNA from wildtype and *Efhc2*$^{\Delta/Y}$ mutant mouse brain was isolated and cDNA was synthesized. Using this cDNA template, we performed RT-PCR using several primer pairs (Table 4, Appendix 2). While we could amplify full-length cDNA consisting of exons 1-15 from the wildtype tissue, we failed to do so in the mutant. However, cDNA fragments were amplified in PCR with forward primer binding in exon 1 and reverse primer in exon 5. We also detected cDNA with forward primers binding in exon 4 and reverse primer in exon 15 (Figure 22). Using Sanger sequencing analysis of the amplicons obtained from the mutant, we confirmed the deletion of exon 3 in the mutant mice. We repeated these experiments in mice obtained from three different matings, and the deletion of exon 3 was found to be consistent.

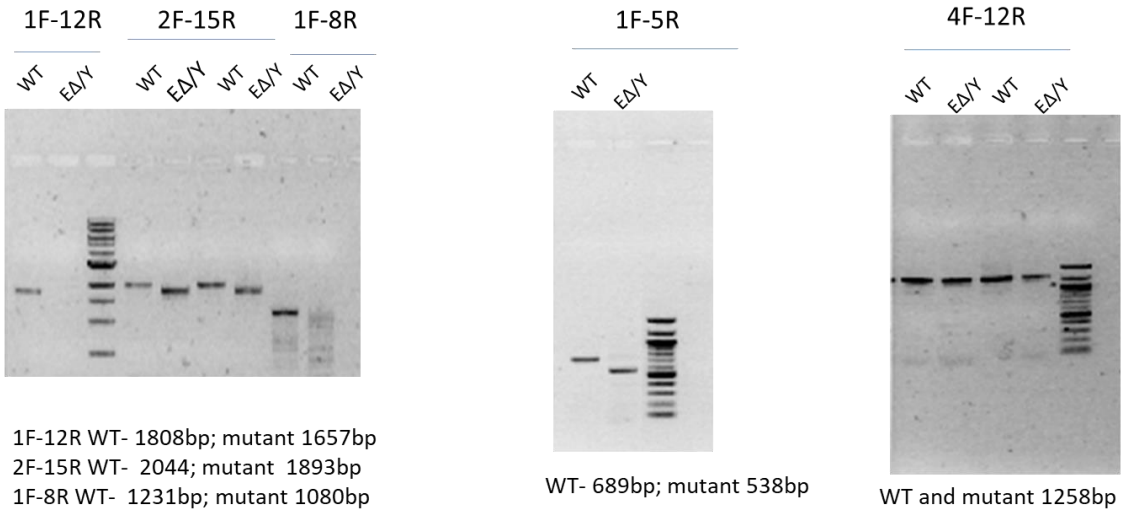


Figure 22: Analysis of *Efhc2* RT-PCR from cDNA of WT and *Efhc2* mutant mice. These PCR fragments were sequenced for exon-walking.

The consequence of the deletion of exon 3 would be that the transcript would be unstable and hence be subjected to non-sense mediated decay. If the transcripts evaded this degradation system, the mutant mRNA predicted to harbor multiple stop codons in exon 4, leading to a peptide of 105 amino acids. This protein would thus be equivalent to the loss of function of EFHC2.

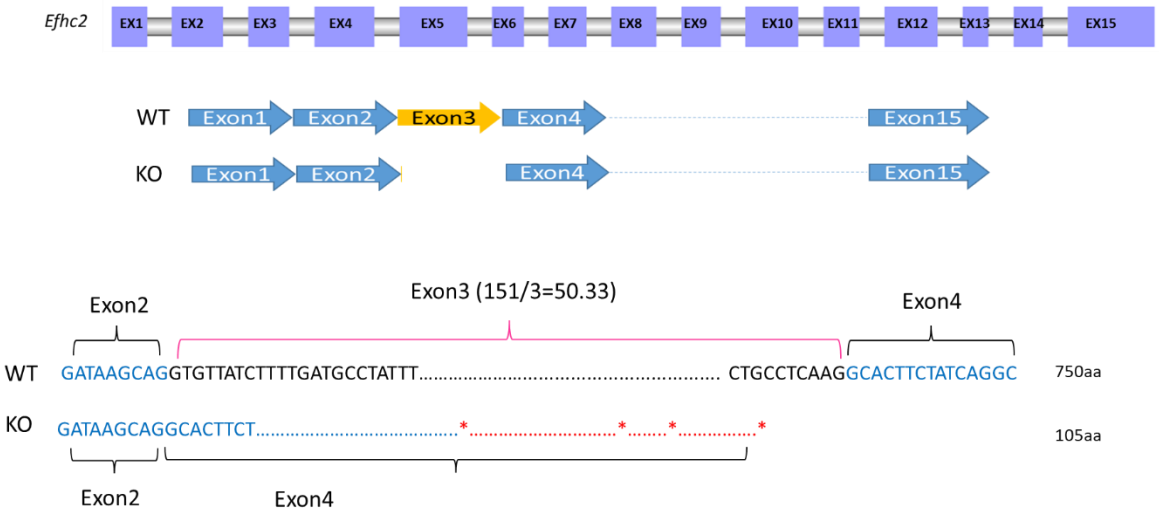


Figure 23: *Efhc2* flipped and loxed, tm1d allele, harbors the deletion of critical region consisting of exon3. This deletion leads to generation of stop codons in exon 4 of the transcript. The resultant mutant protein generated has 105 amino acids.

The deletion of critical exon could be confirmed at the genomic DNA using genotyping PCR and in the brain cDNA using RT-PCR. Western blotting results were in disparity with our findings from PCR, causing the speculation of cross-reactivity of the antibodies. We thus verified our findings of EFHC2 expression in the mouse brain sections using the X-gal staining. For this, we benefited from the gene targeting vector incorporated *Efhc2f*$^{lox/lox}$ mouse lines. Here, the LacZ reporter is driven by the native gene promoter. Thus, the expression of β-galactosidase and the resultant X-gal staining would serve as a reporter for the expression of EFHC2. The coronal mouse brain sections from the *Efhc2* floxed mice were used for X-gal staining. The X-gal staining pattern validated the IHC data for the expression of EFHC2. X-gal staining was seen broadly in the cerebral cortex, hippocampus, dentate gyrus, hilus, thalamus, and cerebellum regions of the mouse brain sections (Figure 24).

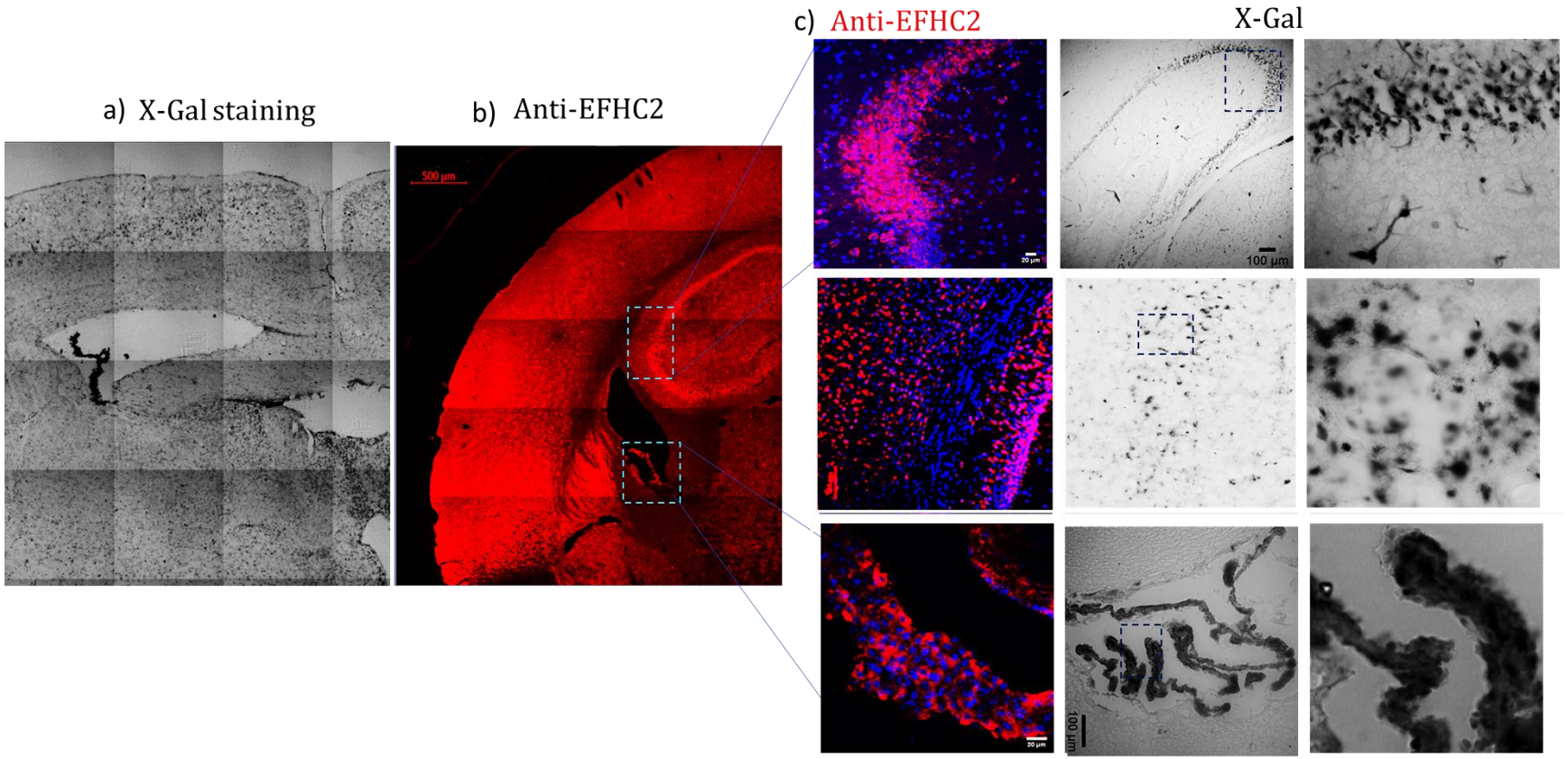


Figure 24: Adult mouse brain section stained for **(a)** LacZ reporter activity **(b)** EFHC2 **(c)** Overlapping regions of LacZ and EFHC2 staining in the hippocampus, cerebral cortex, and choroid plexus.

3.3.9 Cloning and cellular characterization of the *Efhc2* -flipped and -loxed alleles

EFHC2 has three tandem DM10 domains. Of the three, we find that the presence of the second DM10 domains facilities proper localization and function of the protein (Figure 15). In *Efhc2* mutants, *Efhc2*$^{\Delta/Y}$ and *Efhc2*$^{\Delta/\Delta,}$ the mutant protein, if expressed, is 105 amino acids long, lacking all the three DM10 domains and would be non-functional.

To test the effect of the mutant protein, we synthetically generated the mutant cDNA through overlap PCR such that exon 3 would be selectively deleted (Figure 25a). For this, we PCR amplified two amplicons; one consisting of exon 1 and 2, labeled PCR-A, and the other consisting of exon 4 through 15, PCR-B (Figure 25). The products from PCR-A and PCR-B were combined into a linear fragment using overlap PCR using primers binding in exon 1 and 15. The entire length of the cDNA with exon 3 deletion was confirmed by Sanger sequencing. This cDNA was cloned into the pEGFP-C2 vector backbone (Figure 25b). The frame of insertion was confirmed by Sanger sequencing.

The mutant clone generated was expressed in HEK293T cells. The expression of the mutant protein was tested using Western analysis. As predicted, the cDNA generated stop codons on the deletion of exon 3, rendering the production of a protein of 11.6kDa and 105 amino acids (Figure 25c).

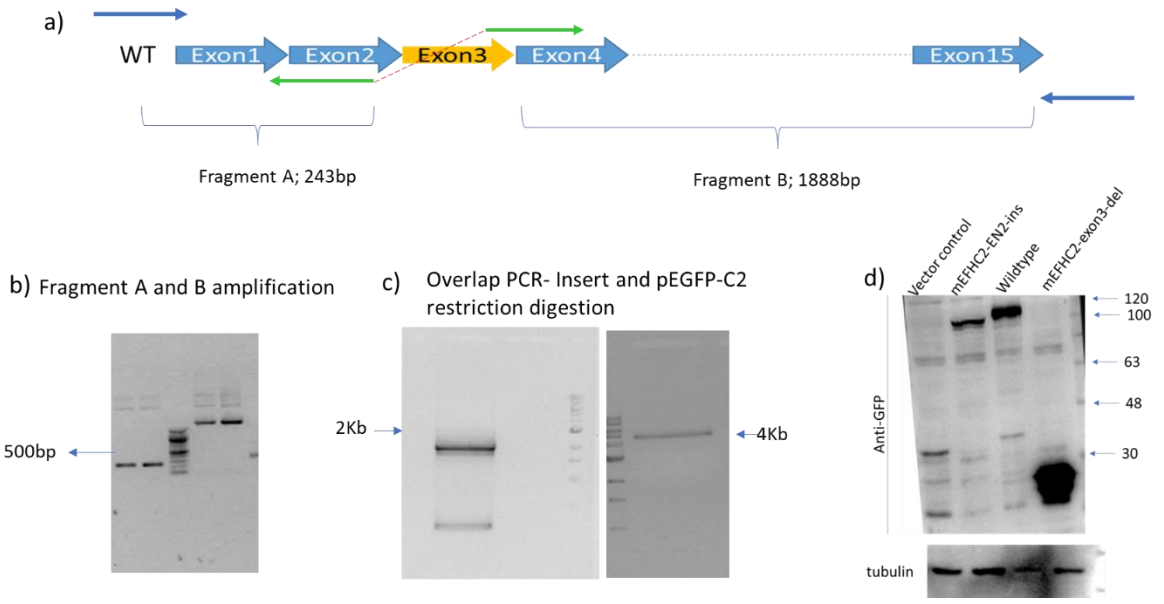


Figure 25: (a) Strategy for overlap PCR cloning with primer binding sites and amplicon sizes. **(b)** PCR amplification of fragments A and B for overlap cloning. **(c)** Insert and vector digestion. **(d)** Western analysis of lysates from HEK293T cells expressing Efhc2 wild type and mutant constructs.

Next, we examined the localization of the mutant protein in comparison to wildtype using HEK293T cells. Cells were stained with α- or γ- tubulins to mark the spindle poles, centrosomes, and midbody. The mutant protein had a diffused and non-specific localization (Figure 26b). The GFP signal could not be detected at the mitotic spindle or midbody in dividing cells expressing the mutant protein. This was in stark contrast with the wildtype protein found at the spindle poles and midbody. We asked if the mutant protein caused mitotic defects due to mislocalization. Cells expressing the mutant protein were quantitated for mitotic defects such as multipolarity, monopolar spindle, abnormal spindles, chromosome segregation defects, and cytokinesis defects. It was found that the mutant protein caused significant mitotic defects (Figure 26).

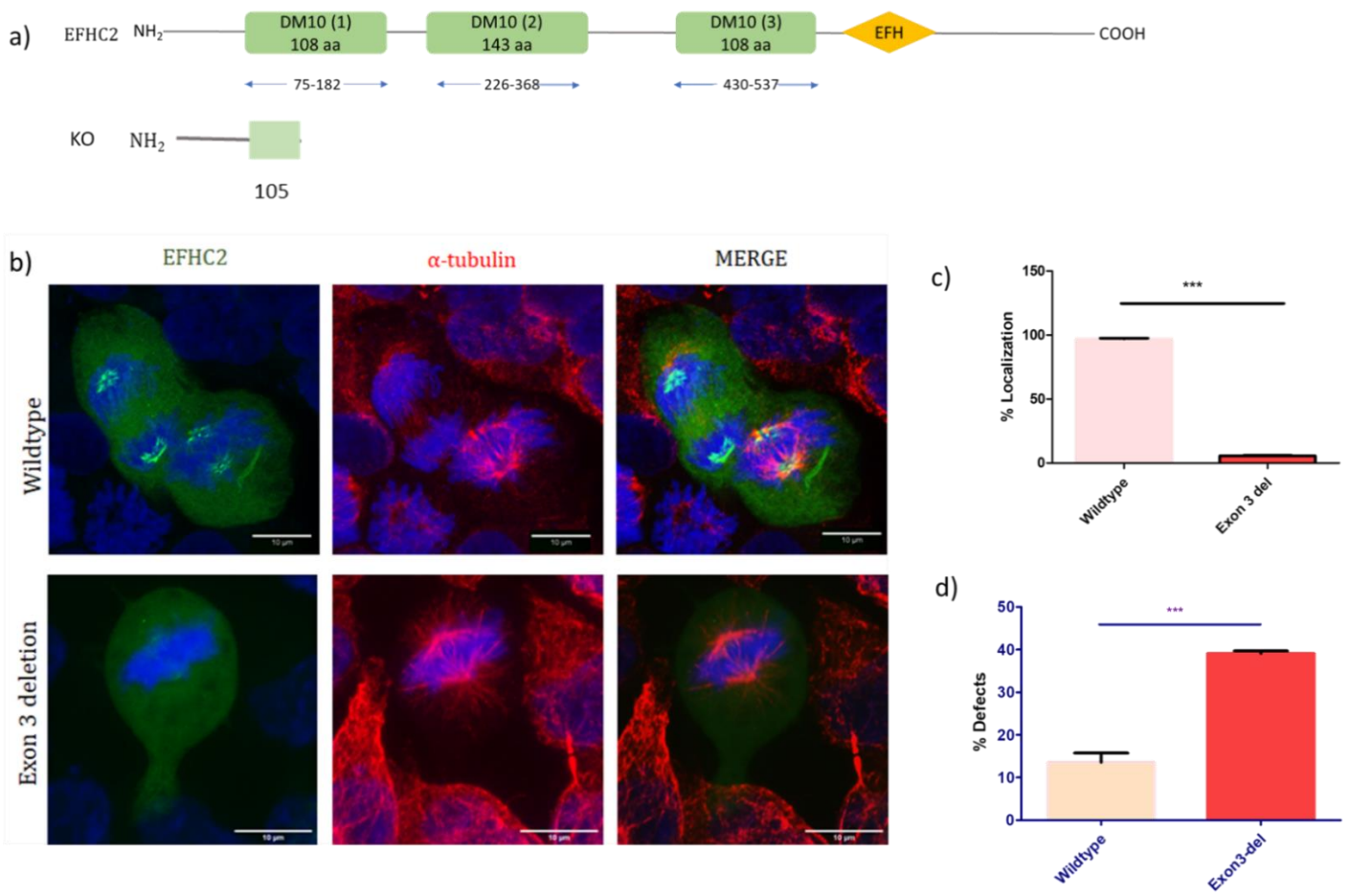


Figure 26: (a) Comparison of EFHC2 wildtype and mutant protein. **(b)** Co-localization of EFHC2 wildtype and exon3 deletion constructs in HEK293T cells. **(c)** Percentage of cells with GFP-tagged wildtype and mutant proteins localizing to centrosomes, spindle pole and midbody. **(d)** Mitotic defects were significantly enhanced in cells expressing the mutant construct; unpaired two-tailed T-test; P<0.0001; N=3, 100 cells per experiment, per construct counted.

3.4 Discussion

EFHC2 variants are associated with a range of neurological phenotypes such as social cognitive competence in women with Turner's syndrome (Weiss et al., 2007), in harm avoidance (Blaya et al., 2009), and social cognitive abilities in males (Startin et al., 2015). Variants at p.Glu634Gly (Ganapathy et al., 2019) and p.Arg486Ser (Helsmoortel et al., 2015) are associated with autism spectrum disorder and intellectual disability. Understanding how genetic mutations affect diverse molecular pathways and lead to behavioral attributes requires an integrative, multistep approach. One of the first steps is to examine the protein in its normal environment, i.e., to study the expression, localization, and cell biological attributes described in Chapter 2. Another approach is to remove the protein from its normal environment and examine its effects. For brain-related phenotypes, this approach is limited by the availability of human brain tissues and ethical concerns of experimentation on human subjects.

Model systems help circumvent these problems. They provide a reproducible and effective means to study brain biology. Genetically altered mice have been very useful in this regard. They allow the examination of a specific gene mutation in multiple individuals with a homogeneous genetic background. The disease onset and progression can be studied in a highly controlled lab environment. Gene ablation can be controlled precisely using mouse recombinases specific to tissue and cell types and time points. Mouse models allow the testing of physiological parameters, behavioral paradigms, and therapeutics in a defined manner.

Variants in EFHC2 are associated with neurological and behavioral attributes. In cell culture, variants identified in JME individuals

EFHC2 were shown to affect the protein's function and cause mitotic defects (Raju, 2014). To further our understanding of the roles of EFHC2 in the brain and assess its contribution to behavioral and seizure phenotypes, we generated Efhc2 mutant mice.

Efhc2 is present on the X-chromosome in both the human and mouse genome. Along with the synteny, the protein architecture of EFHC2 across the two species is well conserved (Figure 1). Earlier work in the lab identified the expression of EFHC2 in the human brain regions cerebral cortex, thalamus, hypothalamus etc. (Raju, 2014). The expression of EFHC2 in the mouse brain was tested using Western analysis, immunohistochemical analysis and RT-PCR. The expression of EFHC2 was seen across various regions of the mouse brain. Our findings were in agreement with the high throughput data reported (Figure 2, Appendix 2).

I examined the expression of EFHC2 in the mouse brain in further detail using immunohistochemical analysis. The protein expression was detected in the cerebral cortex, hippocampus, hilus, dentate gyrus, thalamus, hypothalamus, ependymal cells, and cerebellum. In the cerebral cortex, in layers IV, -V, and -VI, EFHC2 was seen more frequently in the cell body cytoplasm and axons. The ventricles are lined by specialized multiciliate ependymal cells. We found the expression of EFHC2 in the choroid plexus and the ependymal cells. Interestingly, EFHC1 is also expressed in these cell types (Conte et al., 2009; Suzuki et al., 2008).

It has been suggested that EFHC1 and EFHC2 have partially overlapping and redundant functions. Considering the overlapping expression and data from cell culture-based assays showing complementary roles for EFHC1 and EFHC2 it would be appealing to test the effects of double

mutants on viability, fertility, seizure generation and behavior. Another prospect would be to examine if EFHC2 wildtype protein rescues the seizure phenotype in EFHC1 knockout mouse lines using *in utero* electroporation.

Our findings of the expression of EFHC2 in the mouse brain aligned with our objective of generating the mutant EFHC2 lines to examine the underpinnings of the neurological disorders it is implicated in – such as intellectual disability, autism spectrum disorders, and epilepsy.

The cerebral cortex has neurons, glia, and astrocytes organized into six layers to form an ordered and interconnected structure. In the developing mouse brain, corticogenesis and organization occur between embryonic days 10.5 to 18.5 (Sousa et al., 2017). The radial glial cells (RGCs) give rise to most cortical neurons via intermediate progenitor cells (IP). The RGCs divide symmetrically to generate two daughters RGCs or asymmetrically to give rise to one of the following pairs - RGC/IP; RGC/neuron; neuron/IP. This process is influenced by cell cycle regulation in RGCs and IPs (McConnell & Kaznowski, 1991). Through these processes, the cells must maintain their centrosomal numbers, positioning and regulating the mitotic spindle's axis of formation (Xie et al., 2013). The spindle's orientation, the cell cycle duration, and the midbody's inheritance influence the proper formation and function of cells in the neocortex (Marthiens & Basto, 2020). Mutations that lead to misregulation of these processes often result in varied phenotypes as seen in – *Efhc1* causing epilepsy (Grisar et al., 2012), - *Lis1* leading to degeneration and lethality (Yingling et al., 2008), *Lgn* leading to tissue architecture defects (Konno et al., 2008), *Wdr62* with cortical abnormalities and microcephaly (Chen et al., 2014), *Diaph3* with microcephaly and autistic-like behavior, with typical altered motor activity

and social interactions (Lau et al., 2021). The *Efhc2* mutants generated in our study gave rise to mitotic spindle defects *in vitro*. It would be interesting to examine their effects in vivo for phenotypes such as - alterations in cell numbers across the cortical layers, defects in the orientation of mitotic spindle of neural stem cells as has been studied for TPX2 (Vargas-Hurtado et al., 2019).

EFHC2 was detected in the parvalbumin-positive interneurons (PV-Ins), which are a sub-class of inhibitory GABAergic interneurons which drive fast-spiking action potentials at high energetic costs (Inan et al., 2016). EFHC2 expresses in all the cortical layers. We observed co-staining of EFHC2 and PV in a subset of cells in the cerebral cortex layers 2,3,4. Co-staining was also observed in the thalamus and in the cordate putamen. We observed the staining in the cell body and along the axon and dendrites of bipolar interneurons. In the cerebellum, EFHC2 and PV staining was observed in the molecular layer of the cerebellar cortex -in the soma and as dendritic projections from the Purkinje cells extending throughout the molecular layer (Figure 10).

Within the cerebral cortex and thalamus, the PV-positive interneurons are highly interconnected, and these networks are critical for plasticity during development and through adulthood. Their misregulation, positioning, variability in their numbers, and death through apoptosis have been identified as causative in several brain disorders such as autism spectrum disorder, bipolar disorder, and schizophrenia (Ruden et al., 2021). Interestingly, for JME, data from humans and mouse has shown an intricate relation between the GABAergic system and oligodendrocytes, underpinning the pathophysiology (Gilsoul et al., 2019). In mice, the oligodendrocyte progenitor cells and their neighboring interneurons form transient structures synaptic networks. These often regulate the

myelination and tune the conduction speeds (Hamilton et al., 2017). It would be interesting to examine if EFHC2 is involved in this process.

Our findings of the expression of EFHC2 in the mouse brain aligned with our objective of generating the mutant EFHC2 lines to examine the underpinnings of the neurological disorders it is implicated in – such as intellectual disability, autism spectrum disorders, and epilepsy.

The EFHC2 locus was targeted via homologous recombination to generate the floxed allele for *Efhc2*. Using Cre-loxP system, we deleted the exon 3 of *Efhc2*. The excision of exon 3 was predicted to yield a frame shift mutation and a functional null allele. As expected, at the genomic DNA, we confirmed the Cre-loxP mediated deletion of exon 3. Given this, we anticipated that the resultant mutant lines would be a null for EFHC2. However, to our surprise, the mutant mouse expressed a near full length EFHC2. On analyzing the RT-PCR and cDNA sequencing data from the mutant mouse brain a serendipitous, in-frame insertion of the En2 exon in position of exon 3 was found.

The aberrant splicing through the En2 splice acceptor–lacZ reporter cassette although very rare, has been reported earlier. For example, a conditional knockout allele for *Fntb*, encoding the b-subunit of protein farnesyltransferase (FTase), was initially reported to be a loss of functional allele (Mijimolle et al., 2005) basis characterization at the genomic DNA level. However, careful analysis of the transcript from these mutant mice showed the predominant transcript of the mutant allele did not contain a frameshift mutation but rather a protein with a short in-frame deletion (S. H. Yang et al., 2009). The unexpected En2 exon insertion was identified in the generation of mutant mouse for a microcephaly associated gene, *Diaph3*. *Diaph3* codes for a regulator of actin cytoskeleton and aids the transition of cortical progenitors from metaphase to anaphase. However,

in this case, insertion of 115 nucleotides from the cassette created a frameshift mutation with 3 stop codons resulting in absence of Diaph3 protein (Damiani et al., 2016). In the case *Igf2bp2*, coding for insulin like growth factor 2 mRNA binding protein 2, gene trap was used to generate a null allele. However, an in-frame insertion of En2 exon was identified in the *Igf2bp2* mutant mice (Hanisch Jesse J, 2013). These reports along with our study provide a cautionary note and call for careful analysis of the consequences of gene knockout strategy.

We hypothesized that the properties of EFHC2 mutant to be altered in the *Efhc2*-En2-insertion allele. Considering that the wildtype EFHC2 protein localized with the spindle microtubules, we asked if the mutant protein affected the localization and cell division related functions. While the wildtype protein distinctly associated with the spindle poles and the midbody, the mutant protein failed to do so. The effect was pronounced at telophase where in, the mutant protein was not observed at the cytokinetic furrow, with midzone microtubules or at the midbody ring (Figure 20). We followed these leads and quantitated for cells exhibiting mitotic errors such as misaligned chromosomes, multipolar or monopolar spindles, lagging chromosomes and multinucleate cells indicative of abnormal cytokinesis. In comparison to wildtype, the mutant showed a moderate increase in the mitotic defects. We speculate that the deletion of exon 3 and in-frame insertion of the En2-exon causes misfolding of the protein. The misfolded protein fails to properly localize and participate in the regulation of spindle dynamics and cell abscission thus leading to mitotic defects.

Simultaneous deletions and in-frame insertion mutants rarely occur. However, in this case, such an allele was generated serendipitously with the deletion of exon 3 and insertion of En2 exon in the EFHC2 transcript.

In cells, moderate mitotic defects are seen due to this mutant allele. Given that the human mutations identified in individuals with JME display a range of mitotic defects, it is tempting to extrapolate that the *Efhc2*-En2 insertion mutant could manifest a seizure phenotype in the mouse. It would be interesting to examine age dependent spontaneous and PTZ-induced seizures in these mutant mice. Taken together, *Efhc2*$^{\text{En2-ins}}$ mutants are viable, fertile and offer a rare and novel tool to examine the biological roles of EFHC2.

Our aim was to generate EFHC2 null mouse lines. The Cre-loxP mediated recombination resulted in an unexpected in-frame insertion of the En2 exon in the transcript and a near full-length protein. So, we decided to modify our strategy. The gene targeting vector has a splice acceptor (SA) module from the engrailed-2 (En2) gene, followed by an internal ribosome entry site, which directs translation of a fusion reporter protein with β-galactosidase (LacZ), which is dependent on transcription of the endogenous *Efhc2* promoter. A separate transcription unit consisting of a promoter-driven neomycin gene for antibiotic selection is present. These two modules are flanked by FLP sites, permitting flippase-mediated recombination. It has exon 3 of Efhc2 flanked by LoxP sites, allowing conditional ablation of the exon on Cre-mediated recombination. To eliminate the cryptic splicing donor site in the En2 exon, we used Flippase mediated recombination. The resultant flipped allele is a conditional knockout with exon 3 between the loxP sites for Cre mediated excision. Further, on Cre-mediated recombination, exon 3 was deleted. This deletion resulted in generation of premature termination codons (PTC) and the resultant mRNA is expected to be rapidly degraded through non-sense mediated decay (NMD). To our surprise we were able to amplify the mutant cDNA using RT-PCR (Figure). This could be due to ineffective non-sense mediated decay (NMD) of the aberrant transcript. There is growing

evidence to suggest that some NMD-competent PTC-containing mRNAs evade NMD, exist stably in cells or can be recovered from NMD surveillance under specific physiological conditions (W. K. Kim et al., 2017; Trcek et al., 2013).

Given the possibility of the mutant mRNA escaping NMD, we next examined the effect of the possible mutant protein, if any. Efhc2 cDNA lacking exon 3 was generated using overlap PCR and cloned with GFP. The GFP tagged mutant protein was found to be truncated and expressed at 11.6kDa while the wildtype protein was detected around 100kDa. The mutant protein lacking any of the DM10 domain, even if expressed is expected to be severely compromised in its cellular functions. Using cell culture, we tested the effect of the mutated construct on localization and cell division related functions. In stark contrast to the wildtype, the protein failed to localize with centrosomes, spindle poles and midbody, subsequently leading to mitotic defects (Figure 26). These observations are in alignment with the domain truncation alleles wherein, the absence of the second DM10 domain results in mislocalization and mitotic defects.

In cell culture, expression of cDNA cloned from *Efhc2*$^{\Delta/\Delta}$ and *Efhc2*$^{\text{En2-ins}}$ lead to protein mislocalization and mitotic defects. Thus, it would be interesting to compare the mouse brain sections from wildtype and mutants for mitotic errors, spindle misalignment, aneuploidy and apoptosis. EFHC2 is a ciliary protein. The effect of EFHC2 mutant proteins on ciliary length, structure and beat frequencies could be examined. This could offer insights into novel microtubule associated functions of EFHC proteins. Mutations in centrosomal and ciliary proteins also affect the radial and tangential migration of neurons and leads to seizures in mice, as in the case of *Efhc1* (de Nijs et al., 2009; de Nijs et al., 2012) and *Cilk1* (Bailey et al., 2018). It would be important to assess if

Efhc2 mutant mice exhibit impaired neuronal migration in embryonic stages.

Another approach would be to examine Efhc2 mutant mice for seizure phenotypes -spontaneous seizures and -susceptibility to seizure induction on administration of Pentylenetetrazol (PTZ). This would validate the *Efhc2* mouse model for the JME phenotype and offer a new genetic tool to understand the pathophysiology of JME. In this regard, I have undertaken preliminary experiments and titrated the dosage of intraperitoneal injection of PTZ and recorded the phenotypes presented in Table 6, Appendix 2. These preliminary experiments could provide a basis to the phenotypic evaluation of *Efhc2* mutant mice.

Mutations in EFHC1 and EFHC2 cause JME. In cell culture assays, over-expression of EFHC1 was found to partially rescue the mitotic defects caused by mutant EFHC2 and *vice versa* (Raju, 2014). The mutant Efhc2 mice generated here, offer a model system to test the possible redundancy of *Efhc1* and *Efhc2 in vivo*. This could be achieved by siRNA mediated silencing of *Efhc1* in the developing *Efhc2* mutants using *in utero* electroporation.

In summary, I report here the generation and characterization of two EFHC2 mutant mouse lines -$Efhc2^{\Delta/\Delta}$ and $Efhc2^{En2\text{-}ins}$ which offer novel tools to study cellular roles and to examine the behavioral outcomes of *Efhc2* mutations.

Chapter 4

Identification and characterization of *CILK1* variants in JME

Summary

In a previous genetic study, ciliogenesis associated kinase-1 *(CILK1)* variants were causative of juvenile myoclonic epilepsy in 7% of 310 families. *CILK1*, also known as intestinal cell kinase *(ICK)*, encodes an evolutionarily conserved serine/threonine kinase. The protein is critical for the development of nervous, endocrine, and skeletal systems. Pathogenic variants identified in individuals with JME are reported to cause mitotic defects by disrupting neuroblast migration and apoptosis. The contribution of *CILK1* variants to JME in several ethnically distinct populations remains unexplored. In this study, I have examined the protein-coding regions of *CILK1* among 380 individuals with JME. This led to the identification of 23 variants consisting of eleven exonic, nine intronic, and three UTR changes. Of these three are rare, missense changes occurring at conserved amino acid residues. These results expand the repertoire of variants in *CILK1* associated with JME. Cellular localization of the missense variants showed varying extents of localization and cell division defects. Initial *in silico* analysis using homology modeling suggests that mutation p.Phe79Val occurring in the kinase domain will likely disrupt the Van der Waals interaction with ATP. Collectively, this study indicates the contribution of *CILK1* as a causal gene for JME.

Two JME genes, *EFHC1,* and *CILK1,* are situated 0.58Mb apart on the 6th chromosome. Mutations in both genes affect similar cellular functions such as mitosis, cell migration, and apoptosis. Our findings detailed in Chapter 2 show that *EFHC2,* the paralog of *EFHC1,* is a potential JME gene, and mutants lead to mitotic defects. Considering the overlap of cellular effects of mutants of *EFHC1, CILK1,* and *EFHC2*, I examined their co-localization and *in-vitro* interaction using co-immunoprecipitation.

4.1 Introduction

CILK1 at the 6p12.1 is a 60kb gene coding for a 1.8kb transcript consisting of 14 exons and 632 amino acid protein (Figure 1). CILK1 is a serine/threonine kinase with a dual phosphorylation site and shares sequence similarity with mitogen-activating protein (MAP) kinases (Togawa et al., 2000). While it is ubiquitously expressed in the brain, intestine, adrenal gland, testis, its expression is high in the intestinal crypt epithelium. Intracellularly, it localizes to the nucleus, cytoplasm, centrosomes, and cilia. CILK1 and its homologs are components of several cellular processes such as cell -proliferation and -cycle progression, ciliogenesis, hedgehog signaling, inter-flagellar transport (IFT), and autophagy (Fu et al., 2019) (Figure 2).

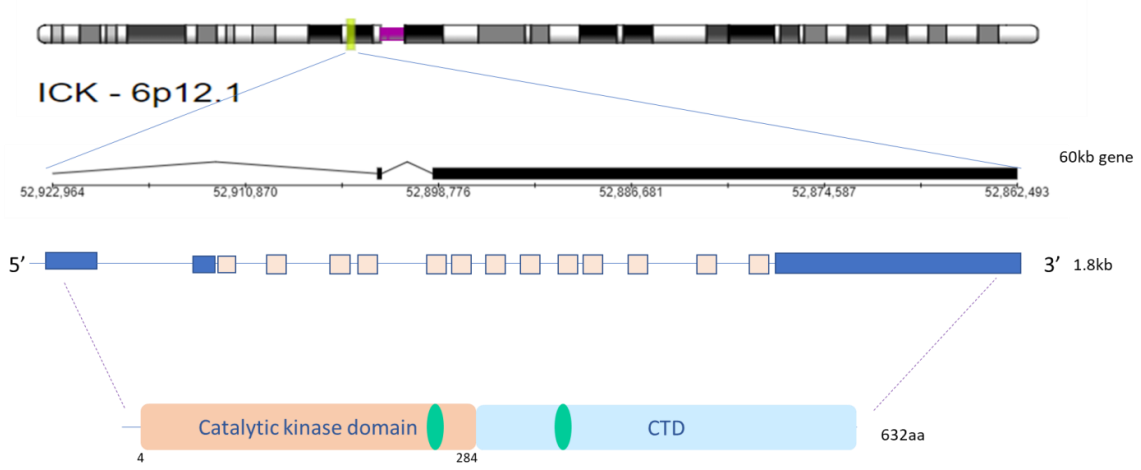


Figure 1: Chromosomal location of *CILK1* gene, transcript structure and the protein domain architecture. Green ellipses represent the predicted nuclear localization signals (NLS).

CILK1 mutation, p.Arg272Gln was first reported using autozygosity mapping and sequencing of members of an Amish family. Six infants of the family were identified with a multisystem, neonatal lethal condition called Endocrine-Cerebro-Osteodysplasia (ECO) syndrome. The mutation p.Arg272Gln in the kinase domain at the conserved residue led to mislocalization and lowered kinase activity (Lahiry et al., 2009). *CILK1* mutations p.Gly120Cys and p.Arg272Gln were identified as rare, pathogenic, and causative of the ECO syndrome in a Turkish family (Oud et al., 2016). Using mIMCD3 cell lines, the mutants were shown to localize to the ciliary tip instead of the ciliary base, as seen with wildtype CILK1. Additionally, reduced ciliary numbers and ciliogenesis were reported in the ECO patient-derived skin fibroblasts harboring p.Arg272Gln (Oud et al., 2016). Mutation at p.Glu80Lys in CILK1 was identified in another patient with ECO syndrome. This mutation in the

catalytic kinase domain also abrogated the kinase activity, inhibited ciliogenesis, and altered ciliary length and distribution (Paige Taylor et al., 2016).

Homozygous mutations in the catalytic domains of CILK1 lead to neonatal lethality with clinical phenotypes correlating to ciliopathies. Recently heterozygous variants in *CILK1* were reported to be causative of juvenile myoclonic epilepsy (Figure 3). The study identified a variant, p.Lys305Thr and additional twenty-one variants in 310 JME patients examined (Bailey et al., 2018). Of these variants, four potentially pathogenic changes at p.Lys220Glu, p.Lys305Thr, p.Asp615Thr, and p.Arg632X were examined by *ex vivo* electroporation in the mouse brain at E14.5, which resulted in mitotic defects, delayed cell-cycle exit, and impaired radial neuroblast migration while promoting apoptosis (Bailey et al., 2018).

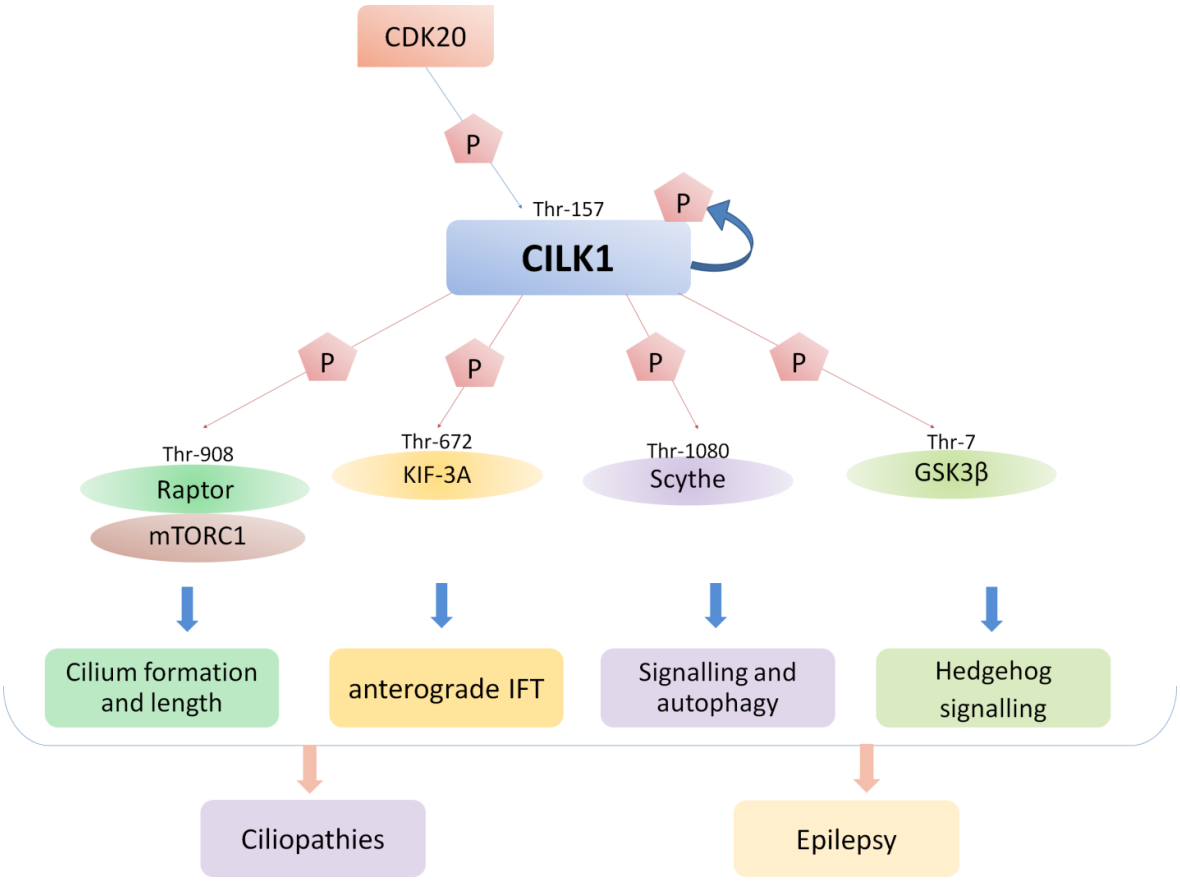


Figure 2: Cellular pathways and regulation of CILK1. CILK1 is regulated by autophosphorylation and through its phosphorylation by CDK20. CILK1 phosphorylates mTOR, KIF-3A, Scythe and GSK3β influences ciliogenesis, autophagy and ciliary hedgehog (Hh) signaling.

In a study involving single-gene burden analysis in 1149 individuals with genetic generalized epilepsy, of which 357 patients had juvenile myoclonic epilepsy, no rare pathogenic variants were identified in *CILK1*. The authors also speculated that the contribution of *CILK1* variants to JME could be a population-specific observation. Genetic studies of this nature

often seek examination of risk factors across ethnically and geographically distinct and diverse populations. This study examines evidence of enrichment of CILK variants in JME patients of Indian origin.

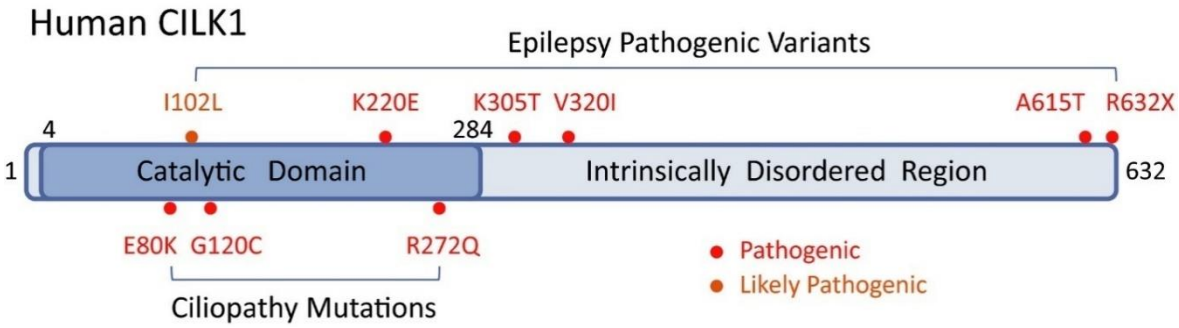


Figure 3: The pathogenic variants identified in human ciliopathies and epilepsy. Three pathogenic mutations p.Glu80Lys, p.Gly120Cys, and p.Arg272Gln) are associated with ciliopathies. Six pathogenic or likely pathogenic variants located in both the catalytic (p.Ile102Leu and p.Lys220Glu) and the non-catalytic (p.Lys305Thr, p.Val320Ilu, p.Ala615Thr, and p.Arg632Ter) domains are associated with epilepsy. Figure from - (Fu et al., 2019).

4.2 Materials and methods

4.2.1 Subjects

A set of 380 JME patients were ascertained from the National Institute of Mental Health and Neurosciences (NIMHANS), Bengaluru. Inclusion criteria for JME were: occurrence of frequent, early morning myoclonic jerks without alteration of consciousness, often accompanied with generalized tonic-clonic seizures (GTCS) starting with bilateral myoclonic jerks, fast 4–6 Hz generalized spike-and-wave and polyspike-and-wave discharges on interictal EEG, and age-of-onset between 8-25 years. Patients manifesting partial seizures of any form, atonic/astatic and tonic jerks, stimulus-induced jerks only, and those with family histories of

progressive myoclonic epilepsy (PME) were not included in the study. These criteria were chosen as per the current consensus on JME diagnosis and management (Kasteleijn-Nolst Trenite et al., 2013). Genomic DNA was extracted from the peripheral blood samples of the participants using the phenol-chloroform method. This study had the approval of Institutional Ethics Committees of the NIMHANS and JNCASR. All patients provided written informed consent.

4.2.2 Mutational analysis of *CILK1*

Sequences of coding regions consisting of 13 exons of *CILK1*, exon-intron boundaries obtained from the Human Genome Data Viewer database at National Center for Biotechnology Information (NCBI), National Institute of Health (NIH), USA and Ensembl Genome Browser, Ensembl Release 75, were examined. For information about the primer pairs used for PCR amplification, please see Table 1 in Appendix 4. Amplification was carried out under the following conditions: initial denaturation at 94 °C for 5 minutes, followed by 40 cycles of denaturation at 94 °C for 30 seconds, annealing at 55-60 °C for 30 seconds and extension at 72 °C for 30 seconds, followed by a final extension at 72 °C for 10 minutes. PCR reactions of 20 µl-volume containing 100 ng of genomic DNA, 5 pmol of each forward and reverse primers, 200 µM of each dNTP, 1.5 mM MgCl2, buffer and 1U of Taq DNA polymerase (New England Bio Labs, Massachusetts, USA or Kapa Biosystems, Massachusetts, USA). Amplified products were checked using agarose gel electrophoresis. Before sequencing, the PCR products were purified using MultiScreen-PCR96 filter plates. PCR-amplified products were single strand amplified by cycle sequencing using 1µl of BigDye Terminator v3.1 Cycle Sequencing reaction mix (Applied Biosystems, Massachusetts, USA), 1X sequencing buffer, 0.25µM primer, and 3µl of purified PCR product in a 20µl volume.

The following cycling conditions were used: initial denaturation at 95°C for 1 minute, followed by 25 cycles of denaturation at 96°C for 10 seconds, annealing at 50°C for 5 seconds and extension at 60°C for 4 minutes, and a final hold at 4°C. The samples were purified by alcohol precipitation by the addition of 16µl of chilled autoclaved deionized water and 64 µl of chilled 95% alcohol to each well of the sequencing plate. The sequencing plate contents were mixed and incubated at room temperature for 30 minutes. This was followed by centrifugation at 2500g for 30 minutes. The precipitated DNA was washed with 150 µl of 70% alcohol and centrifugation at 2000g for 10 minutes. The plate was air-dried to remove all residual alcohol. The DNA was denatured at 95°C in 10µl formamide per well. The denatured single-stranded amplified products were Sanger sequenced using an automated DNA sequencer, DNA Analyzer 3730. The sequence traces were aligned to the reference gene sequences obtained from the Genbank database, and the variants were identified using SeqMan 5.01 (DNASTAR, Madison, Wisconsin, USA).

4.2.3 Bioinformatic analysis of *CILK1* variants

Complete nucleotide sequence of *CILK1* gene is available (NCBI accession number, NT_079573.5). The nucleotide sequence corresponding to accession number NT_079573.5 encompasses 60 kb and contains the reverse complement of the *CILK1* coding sequence from nucleotides 53,001,303-53,061,824. The reference sequence of the transcript with the ID NM_014920.5 was used for analysis. Potential pathogenic effects of the variants were examined using SIFT (P. Kumar et al., 2009), PolyPhen2 (Adzhubei et al., 2010), Mutation Taster (Schwarz et al., 2010), and using VEP annotation. Allele frequencies of the variants were checked in the dbSNP, 1000 Genomes (Genomes Project et al., 2012) and exome variant

server (NHLBI Exome Sequencing Project (ESP), Seattle, WA, December 15, 2012) databases.

4.2.4 Structural modeling

The three-dimensional model for CILK1 was generated using the AlphaFold protein structure database (Jumper et al., 2021). CILK1-specific functional elements, as predicted by NCBI Conserved Domain Database (Marchler-Bauer et al., 2013), were identified and mapped onto the structure using PyMoL. The kinase domain with 1-285 amino acid residues was used for ATP docking. The structure of ATP (PDB ID: 3C4W, 1JNK) was docked on to CILK1 wildtype and mutants using homology modeling and coordinate transfer

4.2.5 Plasmids and antibodies

CILK1 cDNA in pCMV6-AC-GFP was obtained from OriGene (Rockville, Maryland, USA). The sequences of the insert and the reading frame were validated by Sanger sequencing using primers detailed in Table 2 of Appendix 4. Site-directed mutagenesis was used to incorporate the missense mutations identified in patients. The cDNA construct with mutations incorporated were generated by site-directed mutagenesis, carried out using QuikChange® II XL mutagenesis reagents (Stratagene, California, USA). The mutagenic oligonucleotide primer pairs detailed in Table 3, Appendix 4, were designed to incorporate point mutations in the cDNA. cDNA of *EFHC1* and *EFHC2* cloned in pEGFP-C2 vector backbone were used.

The antibodies used in this study are- anti-ICK G-8 (monoclonal, mouse raised; SC365244; Santa Cruz Biotechnology, Dallas, USA), anti-gamma tubulin (monoclonal, mouse raised; Sigma, T5326), anti-alpha tubulin (monoclonal, mouse raised; Sigma, T9026), anti-Flag (monoclonal, mouse

raised; Sigma, 1804) and anti-GFP (polyclonal, rabbit raised; Thermo Fisher Scientific, A-6455), anti-IgG from mouse (Sigma, I5381), anti-IgG from rabbit (Sigma, I5006), goat anti-rabbit alexa flour 488 (Thermo Fisher Scientific, A11008), goat anti-mouse alexa flour 568 (Thermo Fisher Scientific, A11001), goat anti-rabbit IgG – HRP (Genei, Bangalore, India), goat anti-rabbit IgG – HRP (Genei, Bangalore, India).

4.2.6 Cell culture and transfections

HEK293T cells were maintained in DMEM (high glucose, Sigma, D1152) supplemented with 10% heat-inactivated fetal bovine serum (MP Biomedicals, USA, 29167), 2mM L-glutamine (Sigma, G7513), and antibiotics (100U/ml Penicillin and 10mg/ml streptomycin; Sigma, P4333) in a humidified atmosphere of 5% CO_2 at 37°C. Cells were grown to 60-70% confluence in 6-well culture dishes or on poly-L-lysine (Sigma, P2658) coated 18 mm glass coverslips and transiently transfected as follows: an hour before transfection, 10% complete DMEM was replaced with DMEM with 2 mM L-glutamine. Transfections were carried out using Lipofectamine®2000 (Invitrogen, Thermo Scientific, Massachusetts, USA) at plasmid DNA: Lipofectamine ratio of 1:1.5. After 5-6 hours, transfection media was replaced by 10% complete DMEM. Cells were harvested for lysate preparation or fixed for imaging after 24-48 hours of transfection.

4.2.7 Immunocytochemistry

Cover slips with HEK293T cells, transfected with plasmid constructs, and untransfected controls were washed three times in 1X PBS (Phosphate Buffer Saline). Cells were fixed in 2% PFA for 15 minutes at room temperature or in ice-cold methanol for 7-15 minutes at -20°C. Cells were permeabilized in 0.1% Triton X-100, followed by blocking for one hour

with 3% BSA (Sigma) in 1X PBS. Cells were incubated with primary antibodies for one hour. Primary antibodies used in these experiments were (i) mouse monoclonal anti-α–tubulin antibody (1:5000, Sigma) and (ii) Mouse monoclonal anti-FLAG antibody (1:2000, Flag-M2, Sigma) (iii) mouse monoclonal anti-γ-tubulin antibody (1:5000, Sigma). Cells were then washed with 1X PBS and incubated for one hour with 1:500 dilution of secondary antibodies conjugated to Alexa 468 or Alexa 568 (Molecular Probes, USA), followed by DAPI staining and mounting. Imaging was done using Zeiss LSM 880 confocal microscope, and images processing using ImageJ (NIH, USA) and Adobe® Photoshop® software.

4.2.8 Western analysis

Whole-cell lysates of HEK 293T cells were prepared in RIPA buffer (150mM NaCl, 10mM Tris pH 7.5, 0.1% SDS, 1% Triton X-100, 1% sodium deoxycholate, 5mM EDTA) with Protease Inhibitor Cocktail (Sigma). Adherent cells in culture dishes were washed twice in 1X PBS, incubated with RIPA buffer for 30 minutes on ice with intermittent scraping. The suspension was collected and homogenized through a 1ml, 26G syringe. This suspension was centrifuged at 13000 RPM at 4°C for 30 minutes. The supernatant was collected as whole cell lysate.

Total protein concentration in the lysates was estimated using the bicinchoninic acid assay (BCA) reagents (Sigma). For Western analysis, aliquots of 50-70 µg of protein were mixed with SDS gel-loading buffer, boiled for 5 minutes, and then resolved electrophoretically on a 10% SDS-polyacrylamide gel. The proteins on the gel were subsequently electrotransferred (Bio-Rad Laboratories, California, USA) to equilibrated nitrocellulose membrane (AmershamProtran, GE or BioTrace™ NT, Pall life sciences) at constant 400mA for one hour and fifteen minutes. The transfer was confirmed by Ponceau S staining followed by washes in

distilled water. Blocking was performed using 3% skimmed milk powder (Sigma) and 2% Bovine Serum Albumin (BSA, Sigma) in 1X PBS with 0.05% Tween 20, for 12 hours at 4°C. The membrane was probed with a primary antibody for 10 hours at 4°C.

The primary antibodies used in this study were anti-GFP antibody, A6455 at 1:3000 in 0.05% PBSTween-20 (Thermo Fisher Scientific, Massachusetts, USA); anti-Flag antibody 1:3000 in 1% BSA in PBS (Flag-M2, Sigma), anti-alpha tubulin antibody, and anti-gamma tubulin at 1:5000 (Sigma). Post incubation with primary antibody, the membrane was washed in 1X PBS with 0.05% Tween-20 at room temperature for 5 minutes. The blot was further incubated for one hour at room temperature with a 1:5000 dilution of horseradish peroxidase (HRP)-conjugated, goat anti-rabbit IgG (Genei, Bangalore, India) or goat anti-mouse IgG (Genei, Bangalore, India) in 1X PBS containing 1% BSA, 1% skimmed milk powder, and 0.05% Tween 20. The membrane was washed twice for 10 minutes each at room temperature in 1X PBS and 0.05% Tween 20. The protein bands were detected using an enhanced chemiluminescent substrate for HRP (Pierce-Thermo Fischer Scientific, Massachusetts, USA or Clarity™ Western ECL Substrate - Bio-Rad laboratories, California, USA).

4.2.9 Co-immunoprecipitation

HEK293T cells were transfected individually or co-transfected with pCMV6-AC-GFP-ICK, p3X-FLAG-EFHC1, or p3X-FLAG-EFHC2. 36-48 hours after transfection, lysates were prepared in the following IP lysis buffer – 50mM Tris at pH-7.4, 150mM- NaCl, 1% Triton-X 100, 10mM NaF, 1mM Na3Vo4, PIC ultra-complete (Roche, Sigma). The lysates were clarified at 16000g at 40 degree C for 30 minutes. Supernatants were collected as lysate for immunoprecipitation (IP). Dynabeads (Protein-G,

Thermo Fisher Scientific) were washed three times in PBS, incubated with 1µg the antibody or IgG isotype control. Antibody binding was done at room temperature for 10 minutes. Antibody-bound beads were incubated with corresponding lysates, incubated on an end-to-end rotor at room temperature for 30 minutes. Supernatants were collected and stored. The beads were washed two times with 0.05% PBST for 2 minutes each. Each wash was collected and stored. For the final wash step, wash buffer was added, beads were carefully transferred to a fresh tube, the supernatant was collected and eluted with 20ul of elution buffer, or 20ul of 2X-SDS loading dye without β-ME or with 6X-loading dye with β-ME. The elute was collected by boiling the beads directly in the dye. 5-10% of the inputs, washes, and elutes were loaded on a 10% SDS PAGE for Western blotting.

4.2.10 Statistical analysis

All the statistical comparisons were made using GraphPad Prism5. Statistical tests performed were one-way ANOVA followed by Dunnett's test for multiple comparisons. Results are shown as mean ± standard error of the mean (SEM). Differences between groups were considered statistically significant for $P \leq 0.05$.

4.3 Results

4.3.1 *CILK1* variants among the JME patients examined

Sequencing analysis of the 13 exons of *CILK* in 380 individuals with JME led to the identification of 23 variants. Fifteen of these occur at minor allele frequency less than 0.005 across databases (Table 1). Five missense changes, six synonymous, nine intronic, and three UTR rare variants were identified. For details, please see Table 4. Among the missense changes, three variants were absent or rare (MAF<0.005) in databases dbSNP153,

gnomADv3, and Asia 100K. The rare missense variants at p.Phe79Val, p.Thr348Met, p.Trp624Arg occur at conserved amino acid residues (Figure 3). The p.Phe79Val occurs in the evolutionarily conserved kinase. p.Thr348Met and p.Trp624Arg occur in the c-terminal domain. Conservation analysis across various species shows that Phe79 and Trp624 are highly conserved, and Thr348 is moderately conserved. The variants occurring at MAF>0.005 are detailed in Table 2.

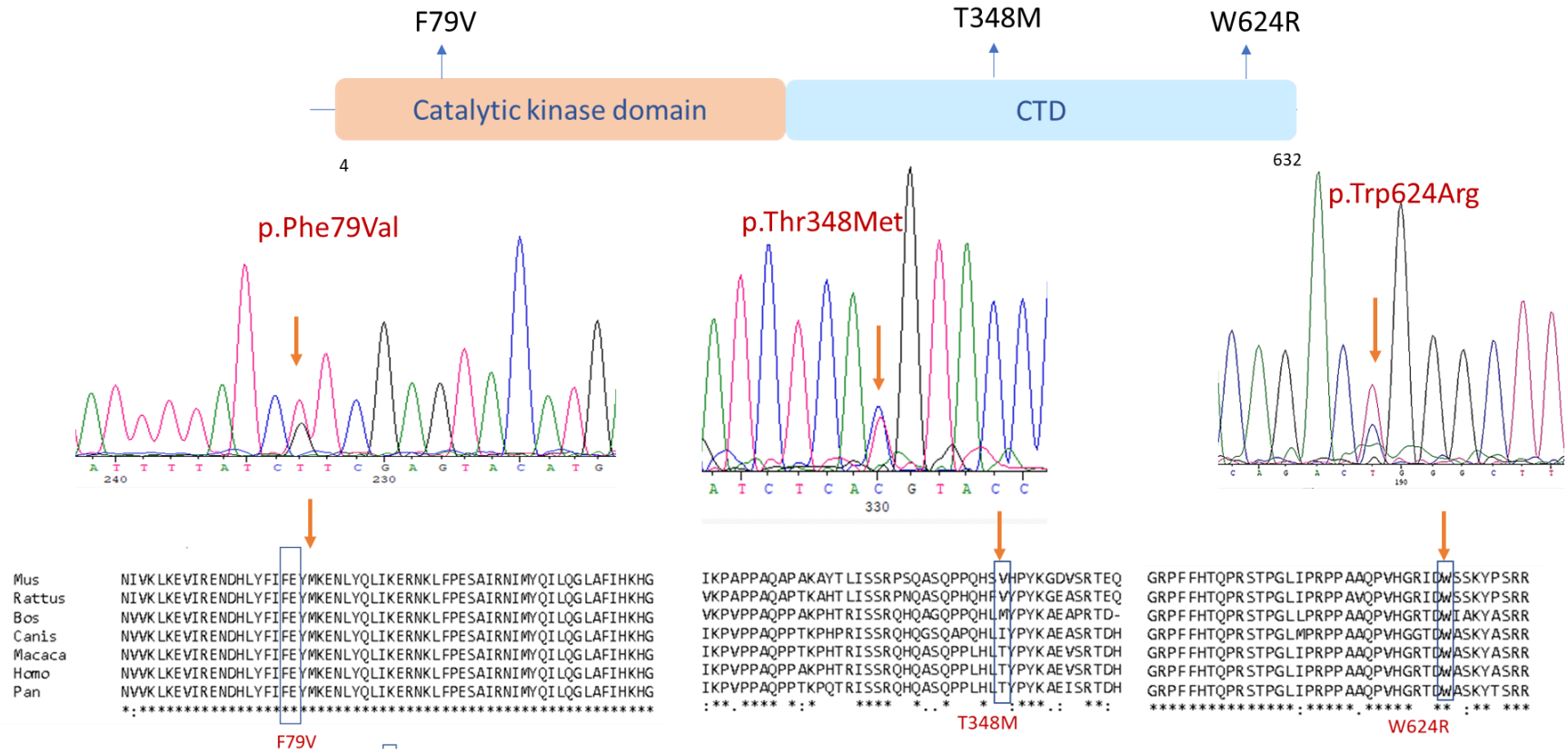


Figure 3: Potentially pathogenic variants in *CILK1* identified in this study – Position, conservation of the amino acid and

Table 1: Rare variants in *CILK1*

cDNA position	Amino Acid change	Occur-rence in patients (n= 380)	MAF in gnomAD, dbSNP 153 or Asia 100K	SNP ID rsID	CADD score	SIFT[a] score	Polyphen[b] 2 Score	Mutation Taster[c]	FATHMM[d]
c.636-A-C	p.Phe79Val	0.001	gnomAD: NA dbSNP: NA Asia 100K: NA	-	27.2	Deleterious	Prob_ Damaging (0.997)	Deleterious	Pathogenic
c.2771-T-C	p.Trp624Arg	0.001	gnomAD: NA dbSNP: NA Asia 100K: NA	-	29.6	Deleterious	Prob_ Damaging (0.99)	Deleterious	Pathogenic
c.1444-C-T	p.Thr348Met	0.001	gnomAD: 0.003012 dbSNP:0.000297 Asia 100K: 0.000862	rs758321988	5.98	Tolerated (0.15)	Benign (0.003)	Benign	Benign
c.452-C-T	p.Ser17Ser	0.001	gnomAD: 0.0004789 dbSNP:0.000173 Asia 100K:NA	rs373305658	5.04	-	-	Benign	Benign
c.1598-A-G	p.Pro399Pro	0.001	gnomAD: NA dbSNP: NA Asia 100K:0.000287	rs575793151	11.6			Benign	Benign

c.1835-G-A	p.Thr478Thr	0.001	gnomAD: 0.002959 dbSNP:0.0006 Asia 100K:0.000575	rs563467548	1.22			Benign	Benign
c.1835-G-C	p.Thr478Thr	0.002	gnomAD: NA dbSNP: NA Asia 100K: NA		1.22			Benign	Benign
c.2460-T-A		0.002	gnomAD: 0.00002942 dbSNP:0.000019 Asia 100K: NA	rs1304308432	7.09			Benign	Benign
c.2460-T-G		0.002	gnomAD: NA dbSNP: NA Asia 100K: NA		7.09			Benign	Benign
c.278+13C>G		0.007	gnomAD: 0.00002415 dbSNP:NA Asia 100K: NA		10.3			Benign	Benign
c.279-455A>G		0.001	gnomAD:0.000015 dbSNP: 0.000007 Asia 100K: NA	rs1562025848	7.06			Benign	Benign
c.156+31T>C		0.001	gnomAD:0.000414 dbSNP: 0.000045 Asia	rs753024438	7.56			Benign	Benign

			100K:0.000287						
c.278+102 T>G		0.002	gnomAD: NA dbSNP: NA Asia 100K: NA		1.14			Benign	Benign
c.1744+14 2G>A		0.001	gnomAD: NA dbSNP: NA Asia 100K: NA	rs548673992	0.318			Benign	Benign
c.157-43C>T		0.001	gnomAD: NA dbSNP: NA Asia 100K: NA		0.5			Benign	Benign

Variants in CILK1 occurring at MAF < 0.005 across databases are detailed here

Abbreviations are as follows: dbSNP = database of single nucleotide polymorphism database; **SIFT =** sorting intolerant from tolerant; **MAF** = minor allele frequency; **CADD** = combined annotation dependent depletion, FATHMM = Functional analysis through hidden Markov models [a]Damaging if score is smaller than 0.05. [c]probably damaging if the score is 0.909-1; possibly damaging 0.447-0.908; benign 0.0-0.446; [c]Probability of causing disease; values close to 1 indicate high confidence of the prediction; [d] SNV is damaging if the score is </=-1.5.

Table 2 – *CILK1* variants with MAF >0.005

cDNA position	Amino Acid change	Occur-rence in patients (n= 380)	MAF in gnomAD, dbSNP 153 or Asia 100K	SNP ID	CADD score	SIFT score	Polyphen2 score	Mutation Taster	Epi25K
c.420-A-G	p.Ile7Val	0.001	gnomAD:0.005917 dbSNP:0.0012 Asia 100K: 0.000862	rs558312131	20.7	Tolerated (0.12)	Benign (0.102)	Benign	1.09E-04
c.1359-G-A	p.Val320Ile	0.001	gnomAD: 0.07154 dbSNP:0.1172 Asia 100K: 0.014663	rs33936662	15.3	Tolerated (0.28)	Benign (0.003)		8.49E-02
c.1376-A-G	p.Pro325Pro	0.003	gnomAD: 0.2998 dbSNP:0.4 Asia 100K:0.152673	rs2297211	6.11	-	-	Benign	2.69E-01
c.935-C-T	p.Ser178Ser	0.001	gnomAD: 0.005917	rs5432433	12.2	-	-	Benign	NA

			dbSNP: 0.0014 Asia 100K: 0.00115	99					
c.1343+37 C>A		0.1	gnomAD:0.01096 dbSNP: 0.0026 Asia 100k:0.001437	rs150108327	7.2	-	-	Benign	NA
c.279-56A>G		0.007	gnomAD: 0.01544 dbSNP:0.056 Asia 100k:0.030477	rs41305912	4.45	-	-	Benign	NA
c.156+55A >G	-	0.001	gnomAD: 0.007890 dbSNP: 0.0016 Asia 100K:0.002302	rs529653677	4.86	-	-	Benign	NA

4.3.2 *CILK1* variants in Epi25

Epi25 is a global collaboration under which more than 9,000 epilepsy patients have been sequenced (https://epi25.broadinstitute.org/). In this database for *CILK1,* 70 variants are listed. Of these, 37 variants give rise to missense changes, 27 are synonymous, and two are loss of function changes. We checked for variants common between the Epi25K dataset with the variants identified in this study. Variants p.Ilu7Val, p.Val320Ile, and p.Pro325Pro are common. p.Pro325Pro occurs in 4931 cases out of the 18340 examined in the Epi25K data set and 4518 of the 16852 controls. This variant is found at an allele frequency of 0.001 in this study. The allele frequency for p.Ilu79Val is consistent across the two studies. It is found at a frequency of 0.001 in our study and has been identified in 2/18430 (allele frequency – 0.001) in the Epi25K database. The variation at p.Val320Ile was found at an allele frequency of 0.001 here, and this variant occurs at a frequency of 0.084. The differences in frequency of these alleles across the two datasets could be due to the differences in sample size.

We have identified p.Trp624Arg, a potentially pathogenic missense variant. In the Epi25 database, the same amino acid has been reported to be mutated to Gly – p.Trp624Gly, occurring in 1/18340 alleles. The novel and rare variants found exclusively in our study indicate the heterogeneity of variants across different populations.

4.3.3 Computational modeling of mutations

ATP structure (PDB ID: 3c4w) was docked onto the kinase domain of CILK1 (1-285aa) using PyMOL. From these, we predict phenylalanine at the 79th residue could be necessary for inter-molecular interaction of the adenine of the ATP. On mutating this residue to valine, the Van der Waals interactions are lost, and the number of non-hydrogen atoms

available for the interaction reduced from 3 to 1. We also see a slight increase in the intermolecular distances from 3.1A to 4.1A (Figure 4).

4.3.4 CILK1 expression analysis

In HEK293T cells, GFP-CILK1 localized to the cytoplasm and nucleus during interphase. Specific localization was seen with the mitotic spindle and midbody during cell division. The staining pattern overlapped with tubulins – α and -γ (Figure 5). The extent of localization of wildtype and mutants to centrosomes, spindle, or midbody was compared. The mutant p.Arg272Glu has a diffused cytoplasmic expression. The construct failed to localize to centrosomes, spindle, or midbody. However, the p.Phe79Val, p.Thr348Met, p.Trp624Arg localized to the spindle poles and midbody during mitosis. Their expression was comparable to the wildtype protein (Figure 6a and 6b). In Western analysis, the wildtype and mutant constructs' expression was similar (Figure 5c).

CILK1 is expressed both in the nucleus and the cytoplasm. The effect of mutations on the distribution of CILK1 was studied using immunohistochemistry. Cells expressing GFP tagged constructs were quantitated for localization in either the nucleus, cytoplasm or both (Figure 5d). We did not observe changes in the proportion of cells with nuclear v/s cytoplasmic v/s equal localization between wildtype and mutant proteins. This suggests that the protein distribution was not affected by the mutations identified in JME patients.

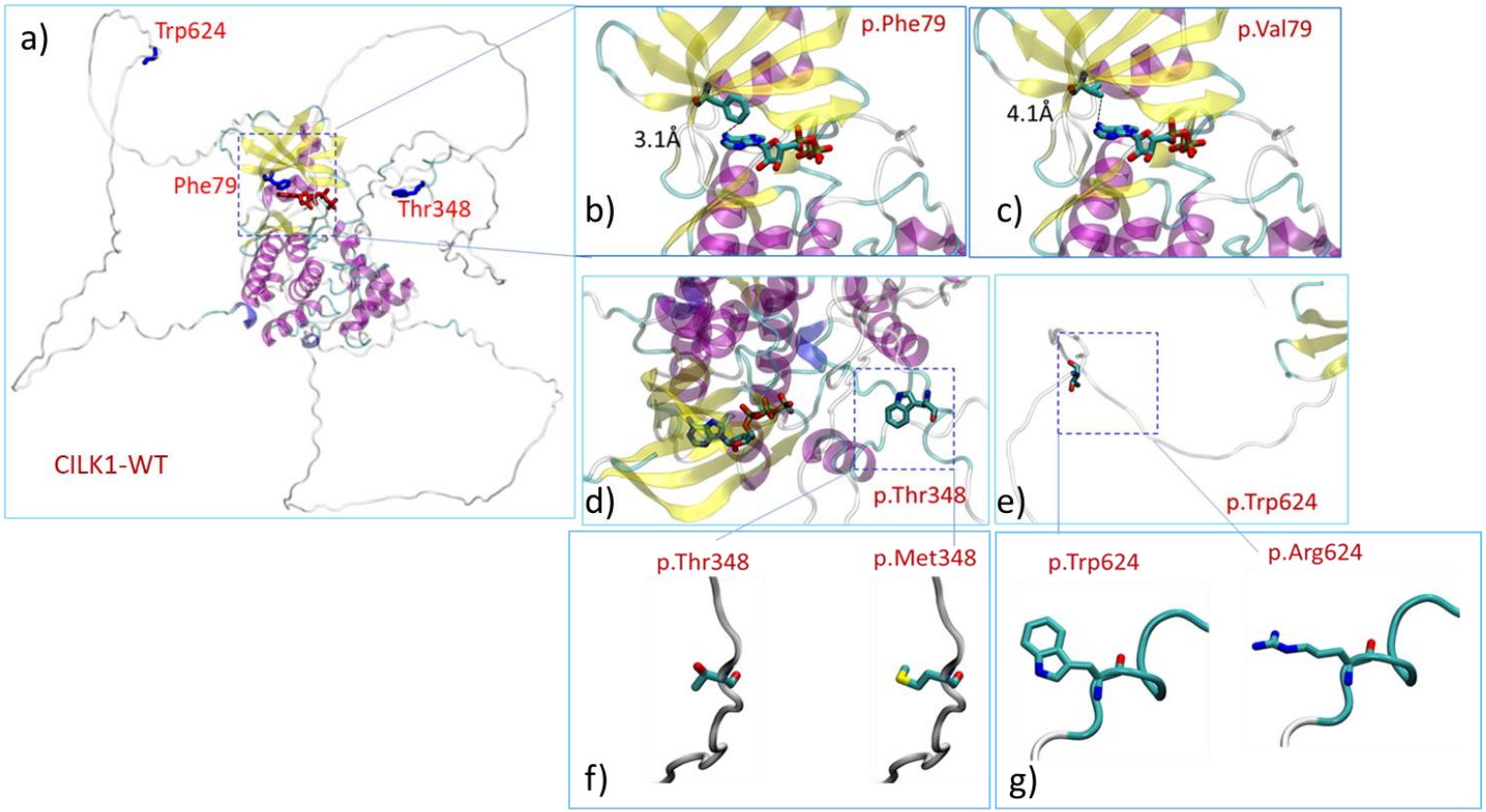


Figure 4: (a) Predicted structure of CILK1 protein docked with ATP in the kinase domain. **(b)** Kinase domain of CILK1 showing intermolecular distance between phenylalanine and ATP. **(c)** Kinase domain of CILK1 showing intermolecular distance between the mutant valine and ATP. **(d)** and **(e)** C-terminal domain with threonine at 348th and Trp at 624th position. Changes from p.Thr348Met and p.Trp624Arg are shown in **(f)** and **(g).** Credits: ATP docking by Mr. Sudarshan Behera, JNCASR

4.3.5 Effect of CILK1 variants on mitosis

During mitosis, we observed CILK1 at the spindle apparatus and midbody. In interphase, CILK1 co-localized with γ-tubulin at the centrosome. Cells expressing GFP tagged wildtype and mutants were quantitated for mitotic defects. The defects were multinucleate, multipolar, monopolar, and cells with lagging chromosomes or abnormal spindle (Figure 6). Compared to wildtype, we find moderate cell division defects due to p.Phe79Val. However, p.Thr348Met, p.Trp624Arg overexpression did not lead to an apparent increase in mitotic defects.

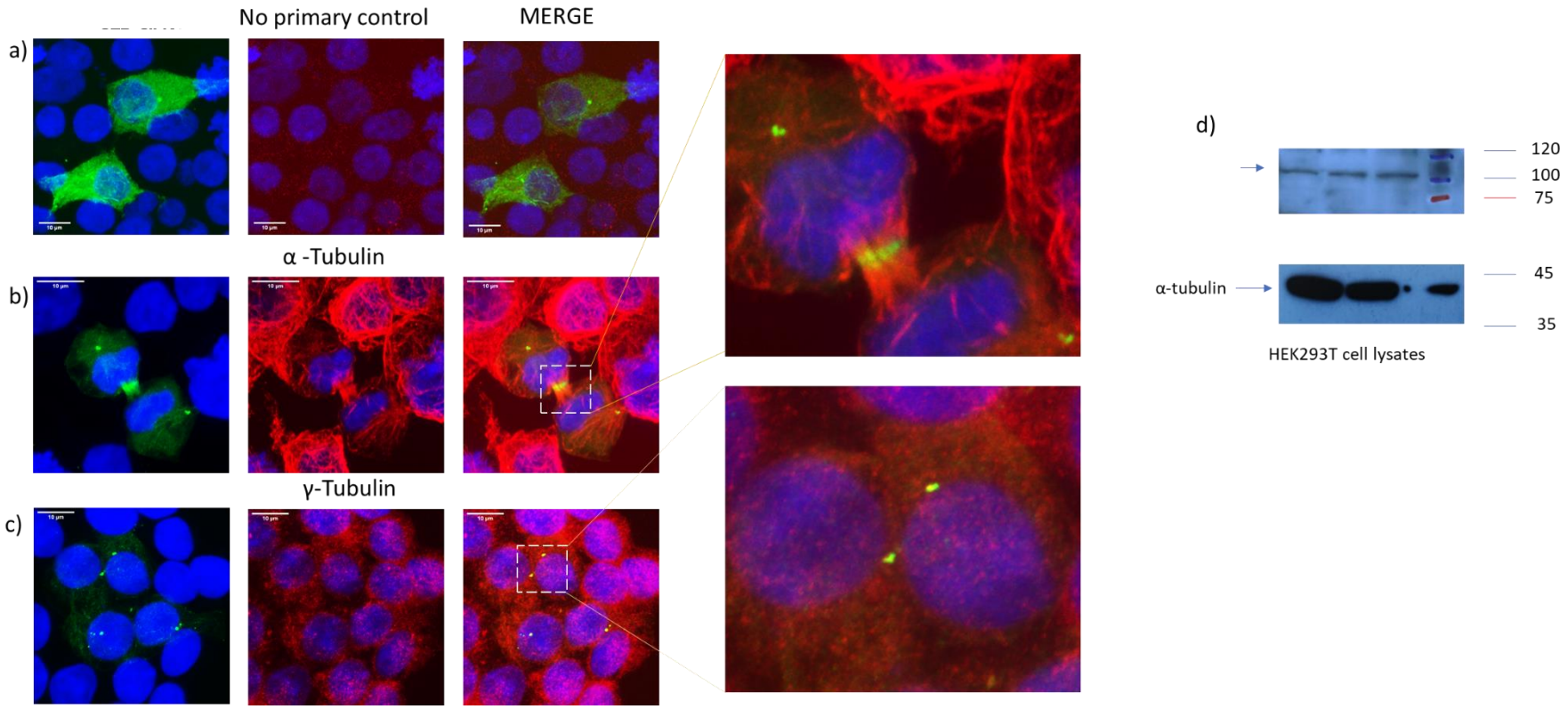


Figure 5: (a) HEK293T cells over expressing GFP-CILK1. **(b)** GFP-CILK1 and α-tubulin co-staining during cytokinesis. **(c)** GFP-CILK1 and γ-tubulin co-staining at the centrosomes. **(d)** Western blot showing expression of GFP-CILK1.

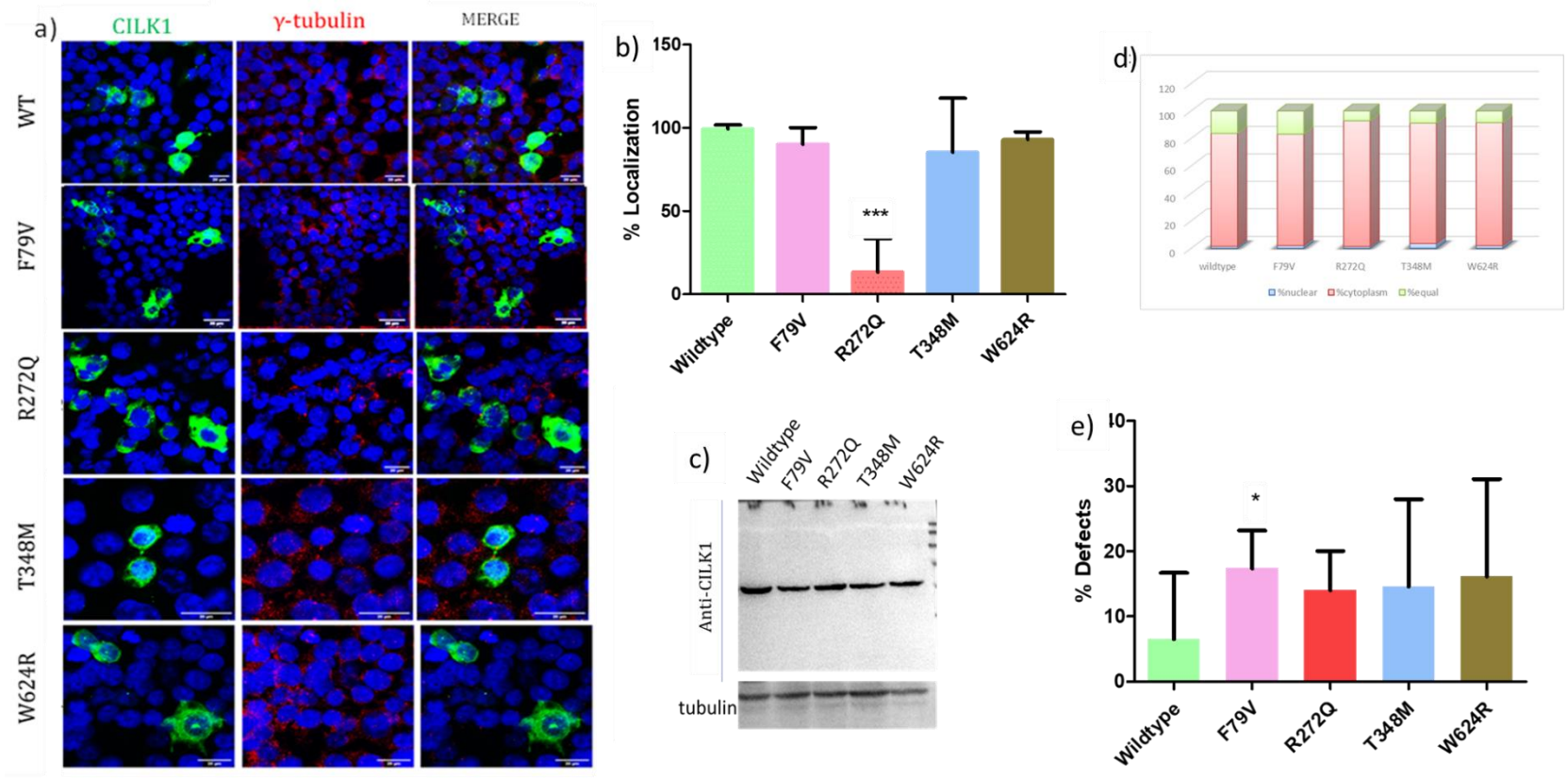


Figure 6: (a) HEK293T cells over expressing GFP-CILK1 wildtype and mutants co-stained with γ-tubulin. **(b)** Analysis of localization of CILK1 wildtype and mutant proteins to centrosomes, spindle poles or midbody. **(c)** Western analysis for expression of wildtype and mutant constructs. **(d)** Proportion of cells with the distribution of wildtype and mutant proteins in the nucleus, cytoplasm or in equally to both. **(e)** Mitotic defects due to CILK1 mutants (N=3, no. of cells – 80 per experiment, per construct)

4.3.6 Localization of CILK1, EFHC1, and EFHC2

Mutations in centrosome and cilia-associated proteins associated with JME reveal the importance of centrosome and cilia to brain development and function. The centrosome is composed of a dynamic and highly interconnected network of proteins allowing the centrosome to play numerous roles in a cell. Centrosomal proteins often have cell-type specific, cell cycle-dependent roles, regulated by specific protein-protein interactions, biochemical alterations, and binding partners. Considering that EFHC1, EFHC2, and CILK1 are ciliary proteins with functions in cell cycle, we tested for their co-localization and interaction.

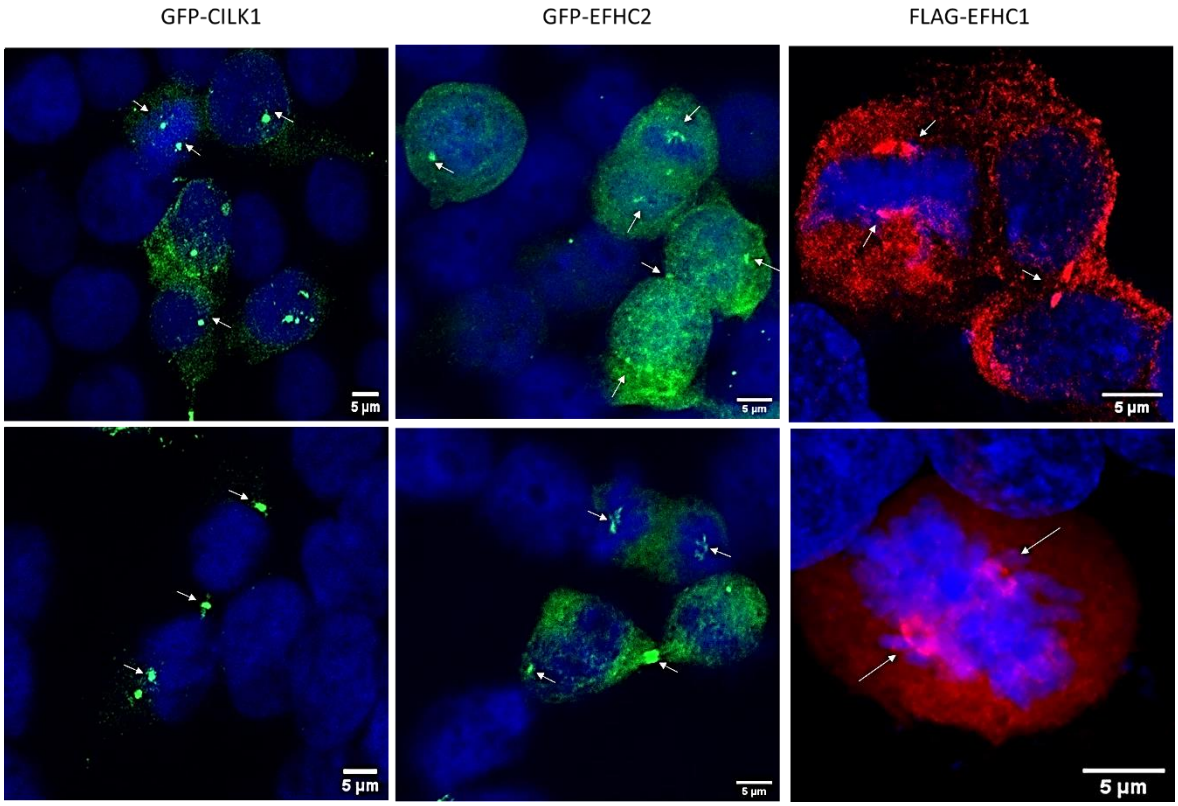


Figure 7: Localization of CILK1, EFHC2 and EFHC1 in HEK293T cells. GFP-CILK1 localizes to spindle poles, centrosomes and midbody. GFP-EFHC2 is detected in the cytoplasm, spindle microtubules and midbody. EFHC1 is localized to spindle microtubules and at the intercellular bridge during cytokinesis. White arrows mark the specific mitotic localization, DAPI in blue stains the nucleic acids.

In dividing HEK293T cells, CILK1, EFHC2, and EFHC1 exhibit distinct localization. In early mitotic cells, the proteins were observed at two distinct puncta suggestive of duplicated centrosomes. They decorated the spindle poles during metaphase and were retained at the spindle poles till early telophase. In a few cells, at the transition from late anaphase to telophase, their expression at spindle poles was reduced except at the intercellular bridge and midbody (Figure 7). At telophase, the localization of EFHC1 was prominent at the central spindle, flanking the Flemming body, while EFHC2 was found at the central spindle and midbody ring. CILK1 was present almost exclusively in the midbody ring and the midbody remnant.

In experiments testing the co-occurrence of EFHC2 and CILK1, we find that these proteins co-localized (Figure 8b). Although EFHC1 and CILK1 appeared at similar locations, they failed to co-localize (Figure 8a). Surprisingly, in co-immunoprecipitation experiments EFHC2 and CILK1, they did not co-elute. Consistent with the lack of co-localization between EFHC1 and CILK1, their interaction was not detected in Co-IP (Figure 8c).

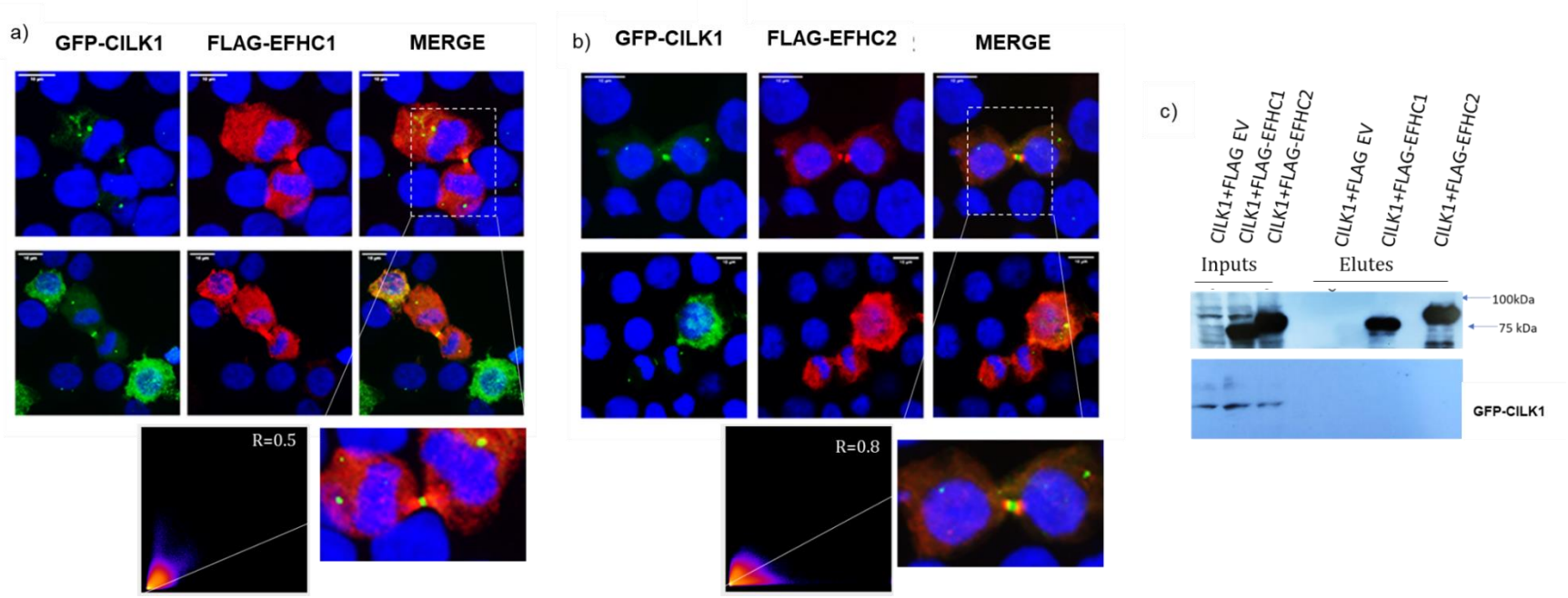


Figure 8: Co-localization and interaction of CILK1 with -EFHC2 and -EFHC1. **(a)** Co-expression of GFP-CILK1 and FLAG-EFHC1 (red). **(b)** Co-expression of GFP-CILK1 and FLAG-EFHC2 (red). R-value is Pearson correlation coefficient. DAPI (blue) stains nucleic acids. **(c)** Western analysis for co-immunoprecipitation of CILK1 with EFHC1 and EFHC2. The inputs and elutes were probed with anti-GFP antibody to detect CILK1 and anti-FLAG antibody to test for immuno-precipitation of EFHC1 and

4.4 Discussion

CILK1 is a MAP-like kinase belonging to the family of tyrosine kinase, in specific to the *v-ros* cross-hybridizing kinase (RCK) serine-threonine kinases (Togawa et al., 2000). The RCK kinases, consisting of male germ-cell associated kinase (MAK) and MAPK/MAK/MRK-overlapping kinase (MOK), are a group of highly conserved proteins. CILK1 was first localized in the intestinal crypt. It has a dual phosphorylation site typical of Mitogen-Activating protein (MAP) kinases (Togawa et al., 2000). This kinase has a multitude of cellular roles. CILK1 is a protein kinase involved in cell cycle regulation, cell proliferation, ciliogenesis, signaling, anterograde trafficking, and autophagy.

CILK1 is ubiquitously expressed and regulates proliferation by promoting G1-cell cycle proliferation. Its downregulation influences the expression and activity of cell cycle regulators and transcription factors such as cyclin D1, c-Myc, and p21, eukaryotic initiation factor 4E (eIF4E), and mTOR/Raptor. Thus, influencing signaling pathways essential for proliferation and differentiation of intestinal epithelial cells (Fu et al., 2009). CILK1 is also a substrate of cell cycle-related kinase (CCRK), a negative regulator of ciliogenesis and cilia length. CILK1 missense and loss of function mutations alter the ciliary length, ciliary numbers, and ciliary signaling (Broekhuis et al., 2014; Fu et al., 2009).

Inactivating mutations in CILK1 – p.Glu80Lys, p.Gly120Cys, p.Arg272Gln in the catalytic kinase domain of the protein cause ciliopathies. Homozygous mutation at Glu80Lys, Gly120Cys, Arg272Gln are causative of ECO- (endocrine-cerebro-osteodysplasia) and ECO-like syndromes. These manifest in a wide range of skeletal and nervous system defects during development resulting in neonatal lethality (Lahiry et al., 2009; Oud et al., 2016; Paige Taylor et al., 2016). Twenty-one pathogenic

heterozygous variants in *CILK1* were identified in patients with juvenile myoclonic epilepsy of Latin American and Japanese origin. Of the missense variants linked to epilepsy, two of them- p.Ile102Leu, p.Lys220Glu are in the catalytic kinase domain, and four in the C-terminal domain –p.Lys305Thr, p.Val320Ile, p.Ala615Thr, and p.Arg632Ter. Among these –p.Lys220Glu, p.Lys305Thr, p.Ala615Thr, and p.Arg632Ter were reported to impair mitosis, promote apoptosis, and delay cell cycle exit. They also affected radial neuroblast migration. These changes during development were speculated to underlie microdysgenesis and the pathological underpinnings of JME. This report also concluded that the variants in the gene *CILK1* cause 7% of cases of JME (Bailey et al., 2018).

The suggestion of the contribution of *CILK1* variants to 7% of JME cases was further tested in 1149 cases of genetic generalized epilepsy (GGE). This included 357 JME individuals and 5911 ethnically matched controls belonging to European and North American ethnicities. Among these samples, *CILK1* variants were not enriched in individuals with GGE or JME. Also, no rare, non-synonymous variants were identified in the JME patients examined (Lerche et al., 2019). These varied results were suggested to be consequences of methodologic issues, allelic heterogeneity, and population-specific observations. This called for studies from different ethnicities to validate or rebut these findings. Given this, we examined the association of variants in CILK1 in the JME cohort of Indian origin.

Among the 380 individuals sequenced for CILK1, twenty-three variants were identified. These include five missense changes, six synonymous, nine intronic, and three UTR rare variants. The missense changes occurring at minor allele frequency less than 0.005 are – p.Phe79Val,

p.Thr348Met, and p.Trp624Arg. Broadly, the incidence of missense changes 3 in 380 and 5 in 310 occurring respectively in the Indian cohort and the GENESS (Genetic Epilepsies Studies) consortium sites are comparable. In the Latin American and Japanese cohort of 310 cases, five missense variants p.Iso102Leu, p.Val320Iso, p.Lys220Glu, p.Lys305Thr, p.Ala615Thr and a nonsense variant p.Arg632Ter were reported. Of these four were classified as pathogenic p.Lys220Glu, p.Lys305Thr, p.Ala615Thr, and p.Arg632Ter. One was identified as likely pathogenic, p.Iso102Leu and p.Val320Iso as benign (Bailey et al., 2018). Among 310 individuals, 21 rare variants were found, accounting to 7%. Among the Indian cohort, 15 rare variants were identified out of 380 cases (4%). The difference in the number of variants in *CILK1* across these studies could be due to geographic and ethnicity differences and methods used in the analysis. For example, in our study, variants occurring at a frequency less than 0.005 across databases were considered rare, while 0.05 was used in the earlier study (Bailey et al., 2018).

Of the rare potentially pathogenic variants in CILK1 identified among Indian patients, the p.Phe79Val lies in the highly conserved catalytic kinase domain. It has been reported that Glutamine at the 80th residue forms a direct bond with ATP. p.Glu80Lys was reported to abolish kinase activity and alter the subcellular localization of the protein. Taking this into account, on performing ATP docking and coordinate transfer, we predict phenylalanine at the 79th residue is important for intermolecular interaction of the adenine of the ATP. On mutating this residue to valine, the Van der Waals interactions are lost, and the number of non-hydrogen atoms available for the interaction reduced from 3 to 1. We also see a slight increase in the intermolecular distances from 3.1A to 4.1A. These predictions need to be tested using biochemical assays to check if the mutation leads to a change in the kinase activity of the protein**.**

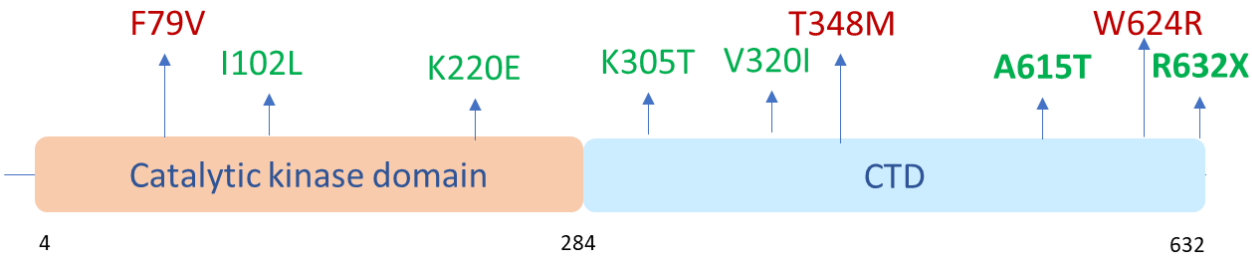


Figure 9: Rare CILK1 variants in epilepsy. CILK1 protein architecture is represented with the location of the three missense variants identified in Indian JME patients (in red) and pathogenic variants identified in JME patients of Latin American and Japanese origin (in green) (Bailey et al., 2018)

Studying the subcellular localization of CILK1, during non-dividing stages, the protein localized to the nucleus, cytoplasm, and co-localized with γ-tubulin at the centrosome. However, during mitosis, we observed the protein localization with the mitotic spindle and the midbody. The missense mutations –p.Phe79Val, p.Thr348Met, and p.Trp624Arg did not change the localization of the protein with centrosome or mitotic spindles. This was not surprising as mutations in EFHC1 or EFHC2 also do not cause mislocalization of the protein but affect cell division. Thus, we examined the effect of CILK1 mutants on cell division. We find that p.Phe79Val caused a moderate increase in mitotic defects while the other mutations –p.Thr348Met and -p.Trp624Arg did not. This could be because the C-terminal domain of the protein has been shown to have

important roles in ciliary localization and ciliogenesis and may not affect the mitotic roles of CILK1. The c-terminal noncatalytic region of CILK1 is important for interaction with the IFT machinery and facilitates the localization of the protein to the ciliary tip (Nakamura et al., 2020). It would be interesting to test if p.Thr348Met and p.Trp624Arg identified in JME patients in this study affect the ciliary localization and IFT mediated intraciliary protein trafficking.

CILK1 occurs both in the cytoplasm and nucleus. It has been reported earlier that the mutations affect protein distribution to the nucleus and cytoplasm or both. We examined the proportion of cells with nuclear, cytoplasmic, or equal distribution of CILK1. While we do not find differences between the wild type and mutants for relative protein distribution localization varied between our study and what was reported earlier. On careful examination of the proportion of cells with CILK1 wildtype equally distributed in the nucleus and cytoplasm of HEK293T cells, it was variable across studies. In experiments with c-terminal FLAG-tagged CILK1, it was 40% (Paige Taylor et al., 2016), with monomeric Red Fluorescent Protein (mRFP), it was 20 (Oud et al., 2016). In our experiments with N-terminal GFP-tag, it was 16%. The differences could be a consequence of the various tags used across studies and require replication with the same tags across cell lines and mutations.

CILK1 plays key role in regulating cilia formation and ciliary length (Fu et al., 2019). CILK1 point mutations identified in JME patients are shown to exert dominant-negative effects on cilia formation and ciliary growth in vitro (E. J. Wang et al., 2020). Here we report three point mutations. Their impact on ciliary growth and function remains to be elucidated. Cilia are generally assembled when cells reach a stationary, quiescent, or differentiated state. In the case of neurons, which are terminally

differentiated, this commitment is spatiotemporally regulated during development. Considering the mitotic and ciliary roles for CILK1, it would be worth studying if CILK1 and its mutants affect ciliary assembly and cell cycle progression in neurons that have primary cilia and ependymal cells with motile cilia.

Chapter 5

A preliminary study to identify potentially pathogenic gene variants in small families with genetic generalized epilepsy

Summary

Genetic generalized epilepsy (GGE) is a class of heterogenous epileptic disorders. They manifest as absences, myoclonic and generalized tonic-clonic seizures and are characterized by electrographic evidence of widespread cortical hyperexcitability. This hyperactivity leads to generalized spike-and-wave (SW) or polyspike-wave discharges and seizures. More recently, using quantitative neuroimaging and neuropsychological tests, cognitive impairment is correlated with GGE etiology (Loughman et al., 2014). GGEs account for 15-20% of all epilepsies and are highly heritable. The underlying genetic causes have been examined through linkage studies, genome-wide association studies (GWAS), and more recently using next-generation sequencing (NGS) analysis.

In the past, for genetic generalized epilepsy, linkage analyses in large families with multi-affected individuals have offered findings of linked regions and mutations. However, with the advent of newer methods of genetic analysis the next-generation sequencing platforms are increasingly employed to identify potentially causative variants. The next-generation sequencing (NGS) approaches offer high information to cost value, allow multiplexing, and are easily accessible. In a preliminary attempt to identify rare variants among 26 small GGE/JME families with two or more GGE affected members, we employed whole-exome sequencing and have

identified gene variants occurring at an allele frequency less than 0.005, for further analysis in the laboratory. We checked if a single variant/ gene could be identified associated with GGE. We did not find any such variant. However, within the members of the same family, an overlap of gene variants was observed. The absence of a common candidate gene/variant is in line with the proposal that a heterogeneous set of genetic variants shape risk for GGEs.

Considering the absence of a common candidate gene/variant, we considered variants in genes with a burden of ultra-rare deleterious variants as reported by the Epi25 collaborative. We found potentially pathogenic, rare variants in genes *KCNB1, PDE4DIP, PCNT, SLC9A1, XPO5, PCDH11X, PCDHGA8, RSPH4A, TTC7A, RWDD4.* These variants need to be confirmed by Sanger sequencing in the families and in additional GGE patient samples to gain confidence for their possible involvement with the disorder.

5.1 Introduction

Epilepsy is a neurological disorder affecting 50–100 million individuals globally, with 2–4 million new cases diagnosed annually (Pitkanen et al., 2016). Genetic generalized epilepsy accounts for 15-20% (Jallon & Latour, 2005). Population-based familial aggregation studies, twin studies examining the concordance of GGE in twins, and studies on large multi-affected families have provided evidence for the heritable nature of the disorder. Estimation of risk of inheritance of GGE among family members of probands showed an increased risk of six times (Peljto et al., 2014). In twin studies, higher concordance values were observed for monozygotic than dizygotic twins (Vadlamudi et al., 2014). These findings suggest the disorder is highly heritable and genetic. Thus, identifying the genetic

causes of this disorder adds value for both the diagnosis and understanding of the molecular basis of the disorder.

Several genetic studies have been undertaken on families and individuals with GGE. The approaches could broadly be classified into four categories: (a) Large families wherein variant or locus identification is guided by inheritance pattern and linkage, (b) evaluation of a cohort of cases and controls – for candidate-gene sequencing, (c) genome-wide association studies (GWAS) and (d) examination of cases and controls using NGS tools in a large cohort of patients across ethnically diverse populations.

Using genome-wide association studies, common variants in *CHRM3, VRK2/FANCL, PNPO, SCN1A, and ZEB2* were identified (Consortium et al., 2012) as disease-associated. Copy number variation with deletions chromosomes 15 (15q13.3 and 15q11.2) or 16 (16p13.11) have been identified in around 3% of all GGE cases (Helbig et al., 2013). More recently, the power of whole-exome-based sequencing tools has been harnessed to assess the burden of rare variants in GGE through three large projects. The Epi4K Consortium and Epilepsy Phenome/Genome Project. Here, an excess of ultra-rare pathogenic changes in known dominant epilepsy genes was identified. However, no single gene with an increased burden of mutations was identified. In a second study conducted by EPICURE, EuroEPINOMICSCoGIE, and EpiPGX Consortia examining 1092 cases, an enrichment of missense variants in *GABA-A* receptor subunits was found (May et al., 2018). In the third and the most extensive study to date, the Epi25K collaborative has examined 3108 individuals with GGE. This study reported enrichment of *GABA-A* receptor genes and the association of *CACNA1G, EEF1A2, and GABRG2*.

Here, I present results of WES of a small set of GGE/JME families, which has revealed certain rare variants with predicted damaging effects.

5.2 Materials and methods

5.2.1 Family ascertainment

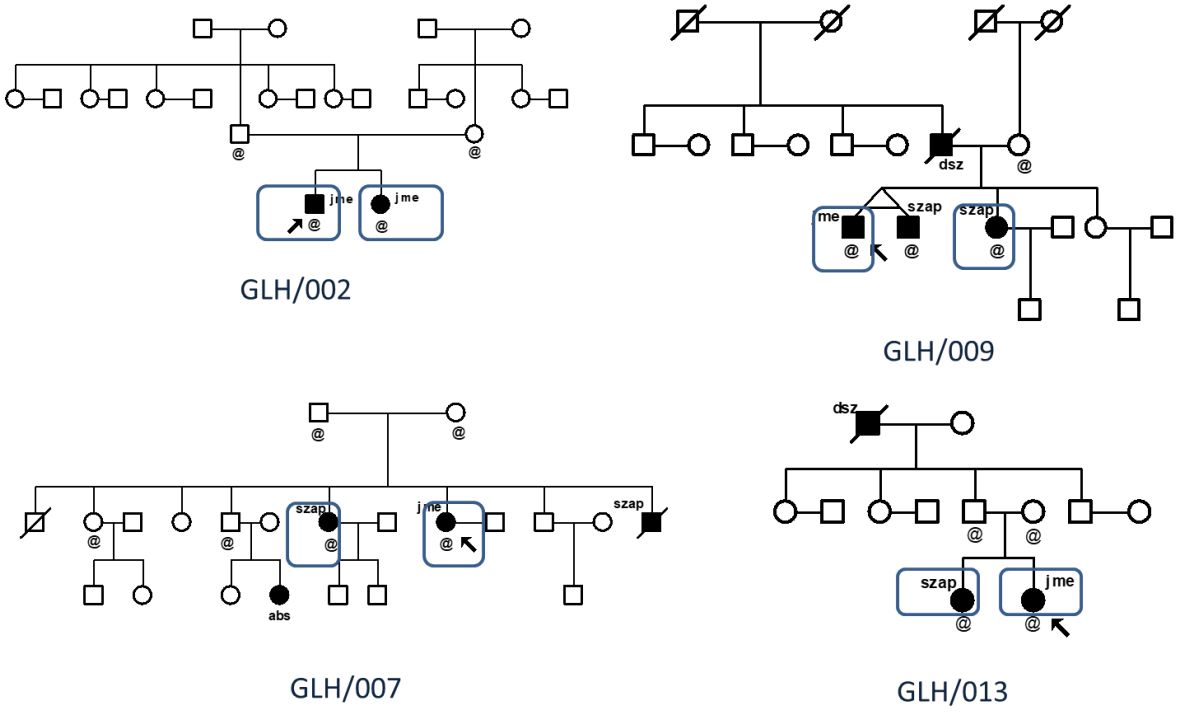
Two- or three-generation families with two or more affected members with clinical diagnoses of GGE/JME were considered for this study. General physical and neurological examinations found no additional clinical manifestation in the individuals examined. All members of these families were enquired about their clinical histories and neurological examination. For the members who were deceased or could not participate in the study, the diagnoses were made based on the information given by the relatives. All members who participated in the study have provided informed written consent. The diagnosis of GGE was established as per the International League Against Epilepsy guidelines (Scheffer et al., 2017b). Genomic DNA was extracted from the peripheral blood samples of the participants using the phenol-chloroform method. This study had the approval of Institutional Ethics Committees of JNCASR. Table-1 details the clinical details of the probands through which the samples were recruited for this sequencing study. The pedigrees of each family are shown below.

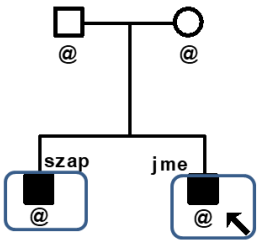
Sample ID	**EEG pattern**	**Photoparaxysomal response seizure**	**Treatment**	**Age of onset**	**Gender**
GLH/002/C1	Not available	Not available	Valproic acid	16	Male
GLH/007/B9	Generalized spike and wave	No	None	12	Female
GLH/009/C1	Generalized spike and wave	Not available	Valproic acid	17	Male

GLH/016/B 2	Generalized spike and wave	Not available	Valproic acid	16	Male
GLH/018/B 1	Generalized spike and wave	Not available	Valproic acid	17	Female
GLH/042/B 5	Generalized spike and wave	No	Valproic acid	12	Male
GLH/13/C2	Generalized spike and wave	Yes	Valproic acid	10	Female
GMW/013/ B3	Generalized spike and wave	Not available	Valproic acid	1	Female
MCC/017/ B1	Generalized spike and wave	Yes	Valproic acid	17	Female
MCC/O24/ B1	Generalized spike and wave	Yes	Valproic acid and Clonazepam	15	Female
NIH/007/C 1	Not available	Not available	Valproic acid	16	Male
NIH/123/C 5	Not available	Not available	Valproic acid	Not availabl e	Female
NIH/162/C 2	Not available	Not available	Not available	Not availabl e	Male
NIH/163/C 4	Not available	Not available	Not available	Not availabl e	Male
NIH/174/B 4	Not available	No	Valproic acid	Not availabl e	Male
NIMS/005/ C1	Generalized spike and wave	Yes	Valproic acid	11	Female
SCT/025/D 32	Generalized spike and wave	Yes	Valproic acid	12	Female
SCT/131/C 20	spike and wave 3Hz	No	Valproic acid	13	Male

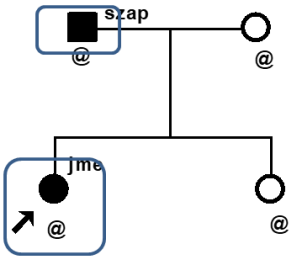
SCT/160/B 2	Not available	Not available	Not available	Not availabl e	Female
SCT/187/C 3	Normal	No	Valproic acid	13	Female
SCT/204/C 4	Generalized spike and wave	No	Valproic acid	13	Male
SCT/229/C 1	Generalized spike and wave	Yes	Valproic acid	12	Female
SCT/271/D 4	Generalized spike and wave	No	Valproic acid	15	Female
SCT/290/C 9	Generalized spike and wave	No	Phenytoin	8	Male

Table 1: Clinical details of the probands of the families analyzed

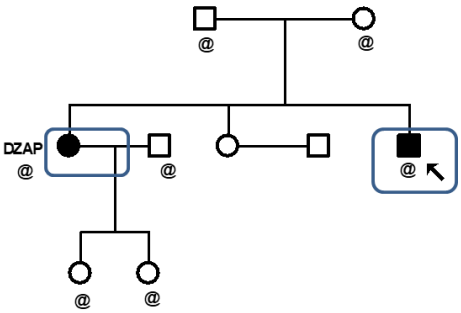




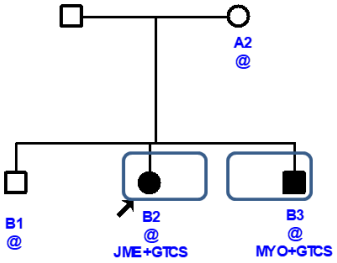
GLH/016



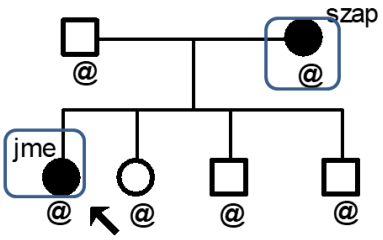
GLH/018



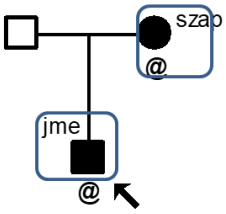
GLH/042



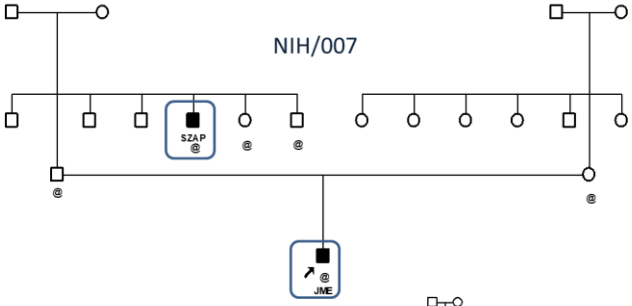
GMW/013

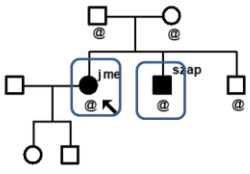


MCC/017

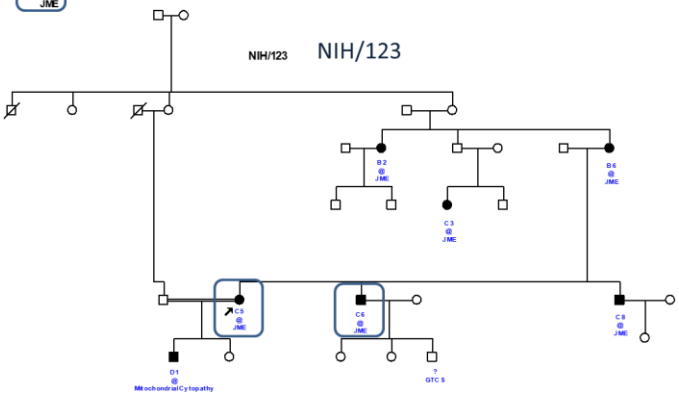


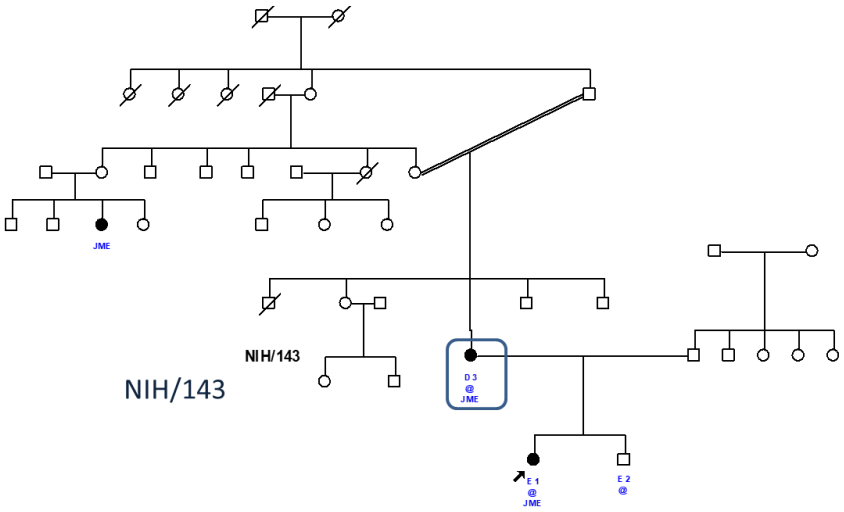
MCC/024

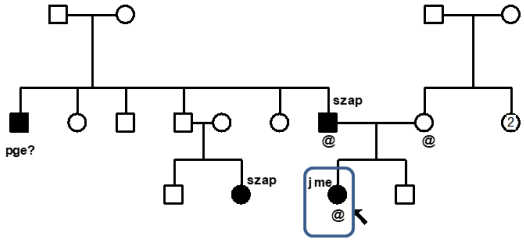




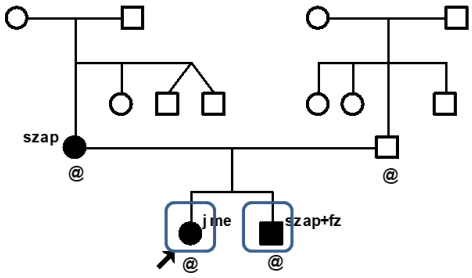
SCT/160



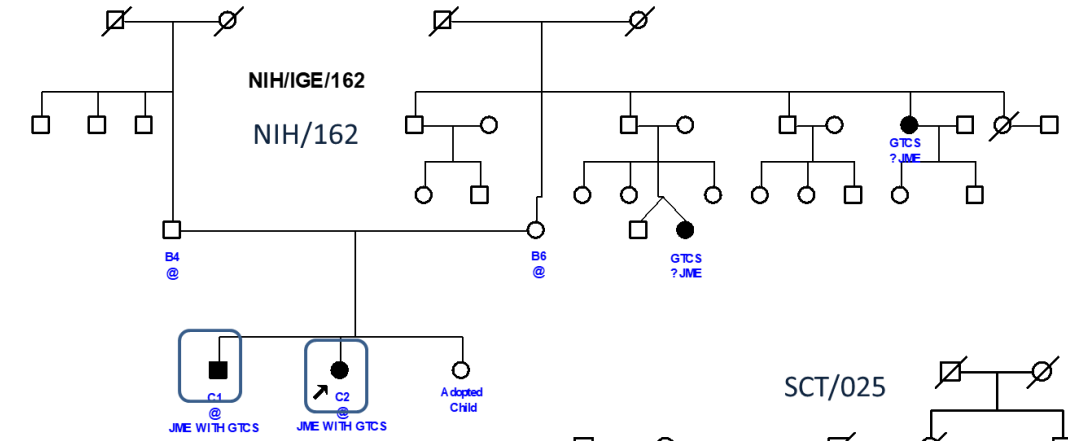


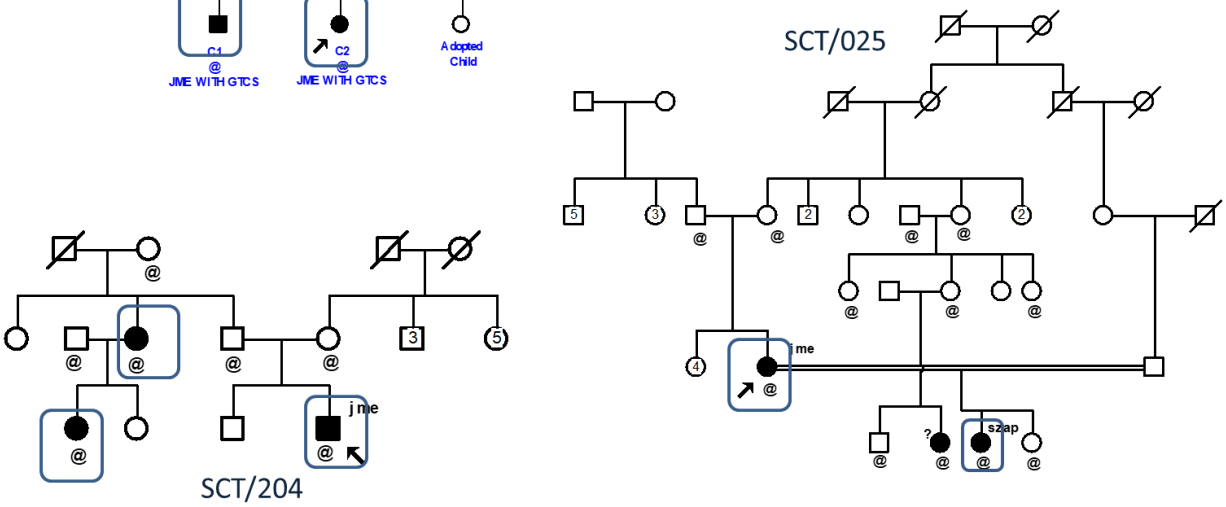


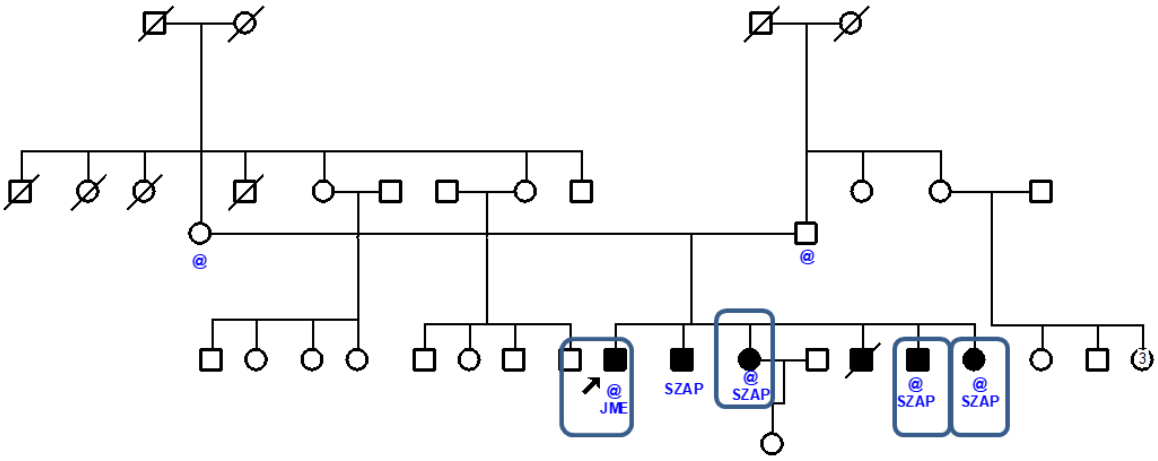
SCT/187



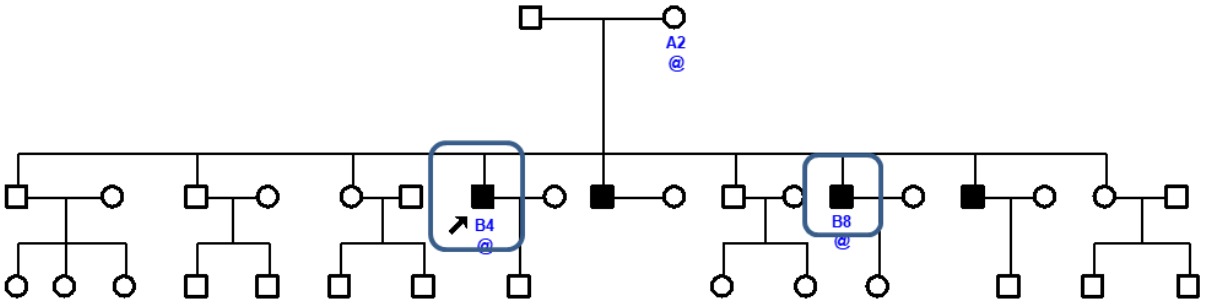
NIMS/005


NIH/IGE/162
NIH/162
B4
@
B6
@
GTCS
?JME
GTCS
?JME
C1
@
JME WITH GTCS
C2
@
JME WITH GTCS
Adopted
Child

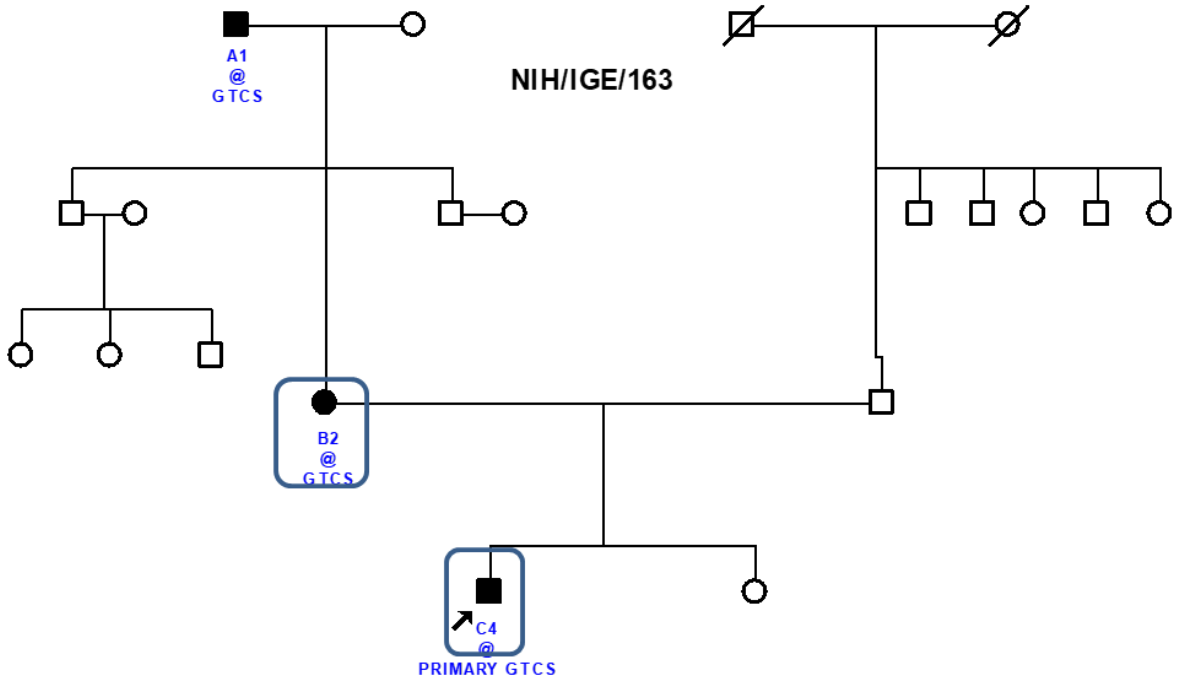

SCT/025
jme
szap
jme
SCT/204


JME
SZAP
SZAP
SZAP
SZAP

SCT/290



NIH/174



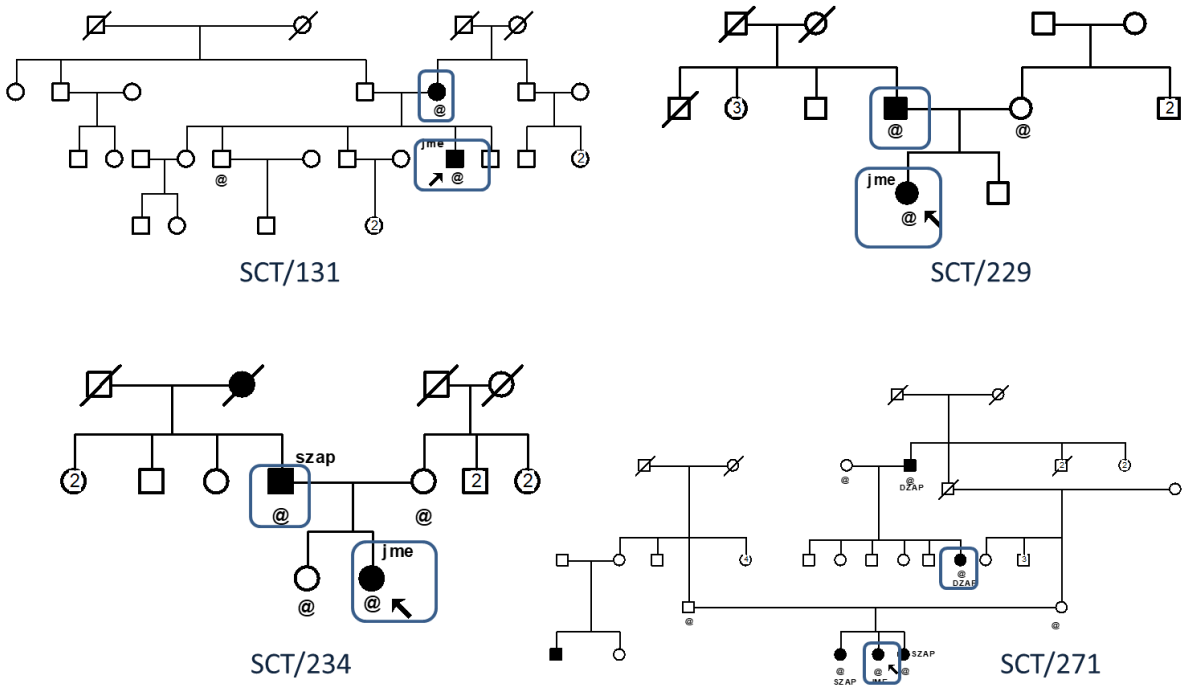


Figure 1: **Pedigrees of the families analyzed**. Squares denote males, and females are marked by circles. Filled symbols represent affected individuals and empty, unaffected ones. The blue boxes indicate the individuals subjected to whole-exome sequencing.

5.2.2 Library preparation

Genomic DNA was isolated from the whole blood of the participating individuals using the phenol-chloroform method (Green and Sambrook, 2017). DNA sample integrity was assessed using agarose gel electrophoresis (Figure 2a). 50 ng of genomic DNA was used for library preparation using NEXTERA rapid capture expanded exome as per manufacturer's protocols (Illumina, San Diego, CA, USA). The targeted

exome spans about 62 Mb (1.7%) of the human genome, corresponding to the consensus coding sequence database comprising a set of curated protein-coding regions that are of high quality and are consistently annotated; over 700 microRNA from the Sanger v.13 database; and over 300 additional human non-coding RNAs. The library preparation protocol in brief, is as follows: Genomic DNA was quality assessed on agarose gel electrophoresis (Figure 2) and normalized in a two-step process. After the initial quantification, gDNA samples are first normalized to 10 ng/ul and further to 5 ng/ul. The libraries were prepared in sets with 12 DNA samples in each set.

The gDNA was tagmented to obtain 300bp fragments. The tagmented DNA fragments were purified and analyzed on Agilent Technologies 2100 Bioanalyzer using a high sensitivity DNA chip. The post-tagmentation library consisted of fragments ranging from 150bp to 1kbp (figure 3b). Index adapters were added to these fragments using a PCR reaction. The PCR products are quantified using Qubit dsDNA broad range kit, Figure 2c (Thermo Fisher Scientific). The products were purified. The tagmented and indexed library was used for probe hybridization using Expanded Exome Oligos (EEX) containing biotinylated probes. The hybridized probes were captured using streptavidin magnetic beads. The beads were washed twice to remove non-specific binding. The enriched library was then eluted from the beads. The DNA binding with capture probes was repeated to enhance the specificity of the captured regions.

The enriched DNA library was eluted, purified, and amplified. The quality of the library was assessed using Agilent Technologies 2100 Bioanalyzer using a high sensitivity DNA chip. DNA fragments with a size of ~200 bp to ~1 kbp were observed, as shown in Figure 2d. The sample library was quantitated and was ~30 to ~45ng/ul. Sequencing (paired-end 2x100 bp)

was performed on an Illumina HiSeq2500 machine according to the recommended protocol (Illumina, CA, USA).

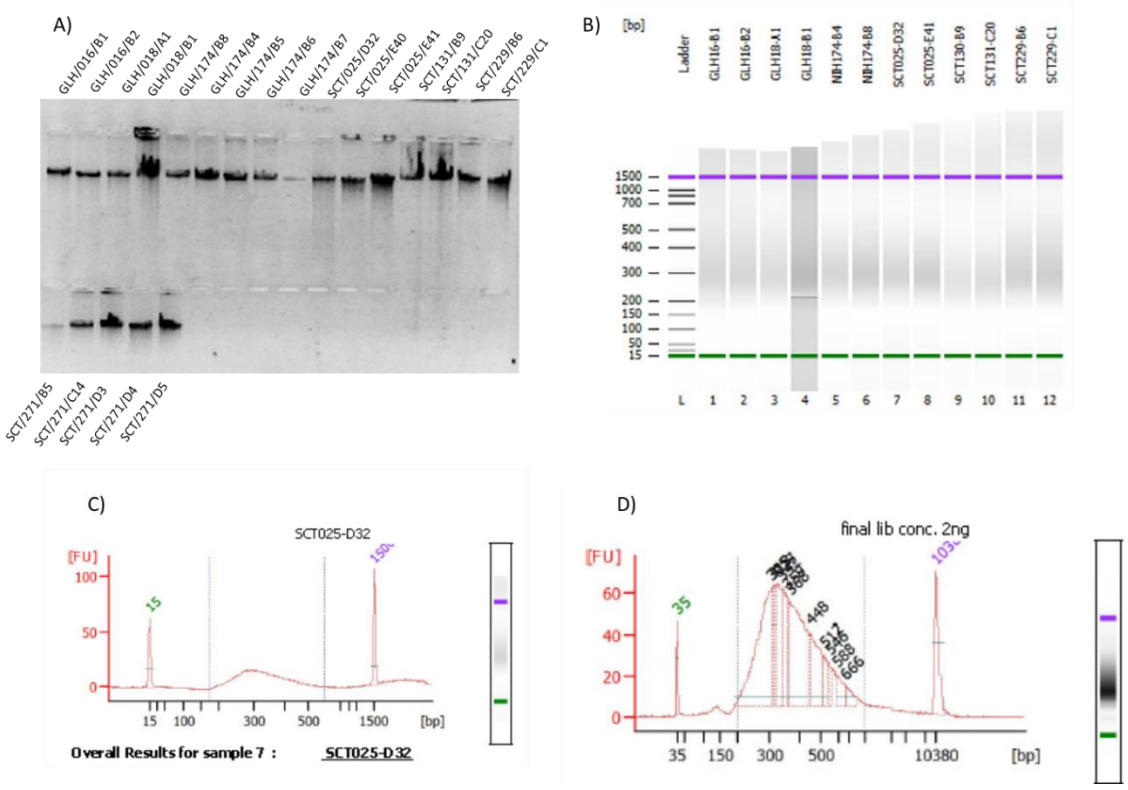


Figure 2: Quality control of genomic DNA and exome libraries. (A) Representative image of agarose gel electrophoresis of DNA sample quality. **(B)** Post-tagmentation, PCR products were analyzed using Qubit ds DNA broad range kit. **(C)** Representative image of quality assessment and quantitation of samples post adapter ligation. **(D)** Size distribution of the whole-exome sequencing library.

5.2.3 Data analysis

Raw intensity files from HiSeq 2000 platform were converted into readable paired-end sequence text files using bcl2fastq.pl program (Illumina, San Diego, CA, USA: http://support.illumina.com/downloads/bcl2fastq-

conversion-software-v217.html). Each read pair was searched for adapter contamination, and adapter sequences present in reads were trimmed using Cutadapt (https://cutadapt.readthedocs.io/en/stable/). After trimming low-quality bases (Phred score < 20) from 3'-ends, reads having at least 70% of bases with a minimum Phred score of 20 were considered for alignment. These trimmed reads were aligned to the human reference genome (hg19, UCSC Genome Browser) using Burrows-Wheeler Aligner (BWA) 0.7.5a (H. Li & Durbin, 2009). Duplicate reads, potentially resulting from PCR artifacts, were removed using SAMTools version: 0.1.19-44428cd (H. Li et al., 2009). Further, local realignment surrounding known indel regions was performed, and the quality score for each read was recalibrated with GATK. Variants were called by GATK HaplotypeCaller. Variants identified were annotated against dbSNP144 using in-house scripts and against 1000 Genomes phase 3 data and ExAC-0.3 database using the latest version of Annovar (K. Wang et al., 2010). SnpEff, SIFT, and Polyphen scores were generated using Annovar to prioritize single nucleotide variations. SnpEff software classifies variants into several categories. Variations within coding regions and variations within 5'-UTR regions were analyzed. Single nucleotide variants were further classified into non-synonymous SNVs, splice site variations (splice site acceptor and splice site donor) and start codon/stop codon related alteration (start gained, start lost, stop gained, stop lost). Insertions and deletions are sub-grouped to two categories: codon related alterations and frame shift mutations. Second filtering criteria were database frequencies. Three major databases 1000 Genomes (phase 3 data), dbSNP144 and ExAC server were searched for presence of each variant found in this study and variants with MAF <5% (occurrence 1 in 200) were taken forward for further analysis (Figure 3). At the third step of exclusion, common variants among family members were retained.

These gene lists were filtered using genes with burden of ultra-rare deleterious variants as reported by the Epi25 collaborative (Epi25 Collaborative. Electronic address & Epi, 2019). Evolutionary conservation of the genes and residue of interest was checked using NCBI and UCSC genome browsers.

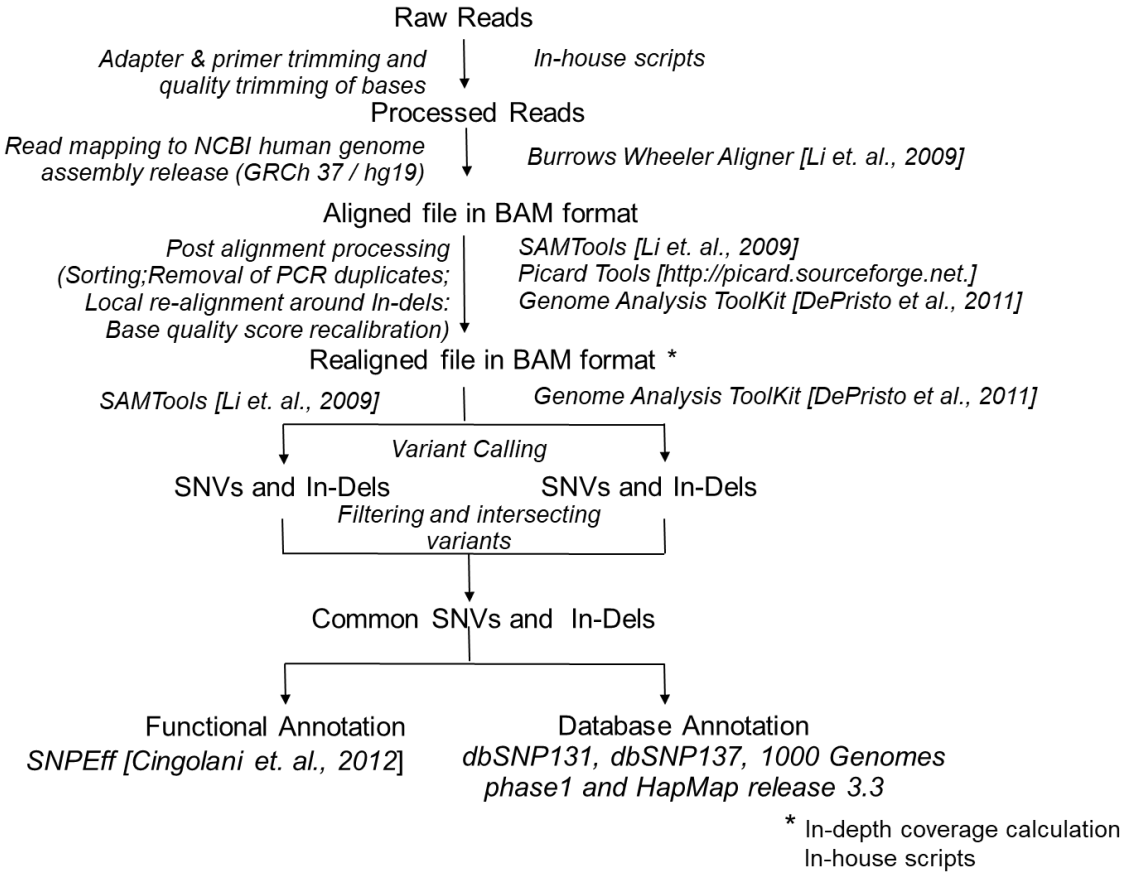


Figure 3: Flow chart with steps of data filtration and analysis. NGS data analysis by Sourav Nayak

5.3 Results

The whole-exome sequencing with the Nextera expanded exome capture resulted in an average of 7,000,000,000 reads each with forward and reverse primers. Of these, 150 million paired-end reads were found to pass the criteria for alignment after quality trimming. An average of 88.75

percent of reads were aligned, and variants were called from these alignments. Basic alignment statistics showed 98.62 percent coverage against probe for the data for 53 individuals. The alignment statistics are recorded in the appendix section. At 5X read depth, the mean value of coverage was 95 percent, with an average read depth of 84.12 against probe coordinates. Among the 55 genomic DNA samples, the library preparation for two samples did not meet the quality control and hence were excluded from the analysis.

We prioritized the variants based on frequency and functional consequence. Using snpEff and dbSNP144, changes occurring at an allele frequency of 0.005 were taken forward and around 3000 variants were identified in each affected individual. We then filtered for variants that were shared among the affected family members. This list constituted of about 1500- 2500 changes (Table -2). The variants were classified into non-synonymous coding, start gained, stop gained, stop lost, codon insertion, codon deletion, codon change plus codon deletion, codon change plus codon insertion, frameshift, splice site acceptor, splice site donor, UTR variants. Among the shared variants in each family, we found around 750-1000 non-synonymous coding variants, 250-400 UTR-ic changes, around 50-80 start and stop gained, 100-120 frameshift mutations. In these data sets, we examined if there was a common mutation. We queried for the occurrence of a common underlying variant in all the affected individuals examined and did not find a common mutation.

We then examined the list of novel single nucleotide variations (SNVs), shared between the affected members of each family for the occurrence of variants in genes the top 200 genes with the burden of deleterious ultra-rare variants (URVs) in genetic generalized epilepsy (Epi25 Collaborative.

Electronic address & Epi, 2019). Among the 26 families examined, we found variants in the top 200 genes with URVs in 6 families. These were in the genes *KCNB1, PDE4DIP, PCNT, SLC9A1, XPO5, PCDH11X, PCDHGA8, RSPH4A, TTC7A, RWDD4* (Table-3). All the variants identified here, except the PCDHGA8, occurring at an allele frequency of 0.000004, were found to be absent across databases such as 1000 Genomes, ExAC, and dbSNP, confirming their rarity. The variants were identified in reads with a quality score of over 600 and coverage in the range of 59-310.

Family-ID	Shared variant count	Sample 1	Sample 2	Sample 3	Sample 4
GLH002	2202	C1 – 3105	C2 – 3374		
GLH007	2340	B7 – 3182	B9 – 3255		
GLH009	2282	C1 – 3430	C3 – 3346		
GLH013	2415	C1 – 3213	C2 – 3164		
GLH016	2441	B1 – 3377	B2 – 3391		
GLH018	2392	A1 – 3398	B1 – 4093		
GLH042	2259	B1 – 3307	B5 – 3225		
GMW013	2265	B2 – 3295	B3 – 3290		
MCC017	2170	A2 – 3203	B1 – 3298		
MCC024	2294	A2 – 3254	B1 – 3251		
NIH007	2056	B5 – 3087	C1 – 3182		
NIH123	2434	C5 – 3337	C6 – 3289		
NIH143		D3 - 2947			
NIH162	2330	C1 – 3290	C2 – 3207		
NIH163	2343	B2 – 3339	C4 – 3298		
NIH174	2445	B4 – 3493	B8 – 3391		
NIMS005	1851	B1 – 3238	C1 – 3210	C2 – 3248	
SCT025	2282	D32 –3215	E41 – 3326		
SCT160	1769	B2 – 3295	B3 – 3282		

SCT187	1664	B6 – 3011	C3 – 3176		
SCT204	1709	B3 – 3045	C1 – 3157	C4 – 2734	
SCT229	2463	B6 – 3391	C1 - 3176		
SCT234	2214	B5 – 2989	C2 – 3271		
SCT271	1963	C14 –3373	C2 – 3254		
SCT290	1583	C9 – 3076	C11 – 3021	C14 –2966	C15 – 3126

Table-2: Total number of variants identified after data filtration steps among individuals. Shared variant counts are displayed in blue.

Genomic location[a]	**Gene[b]**	**Nucleotide change**	**Consequence[c]**	**Amino acid[d]**	**SIFT[e]**	**Polyphen[f]**	**RSID**	**CADD Score**	**Splice AI[g]**
Chr20: 47989812	*KCNB1*	T/G	Non-synonymous coding	D762A	Deleterious 0.02	Benign	Not available	24.1	No effect
Chr21: 47822339	*PCNT*	A/G	Non-synonymous coding	Q1686R	Tolerated 0.25	Benign	Not available	16.08	No effect
Chr1: 145075965	*PDE4DIP*	A/T	5'-UTR- Start-gained: ATG: 103 bases from TSS		Not applicable	Not applicable	Not available	12.19	No effect
Chr6: 116937974	*RSPH4A*	C/T	Non-synonymous coding	P63L	Deleterious 0.04	Possibly damaging	Not available	15.79	No effect
Chr5: 140773903	*PCDHGA8*	A/C	Non-synonymous coding	N508T	0	Possibly damaging	rs 1205005872	23	No effect
Chr6: 43543714	*XPO5*	T/G	5'-UTR: 113 bases from TSS		0	Possibly damaging	Not available	2.6	No effect
Chr2: 47184087	*TTC7A*	C/T	Non-synonymous coding	S153F	Deleterious 0.03	Possibly damaging	Not available	22.5	No effect
Chr4:	*RWDD4*	G/A	5'-UTR: 220		Not	Not	Not available	11.82	No

184580325			bases from TSS		applicable	applicable			effect
Chr1: 27481308	*SLC9A1*	A/G	5'-UTR: 483 bases from TSS		Not applicable	Not applicable	Not available	11.87	No effect
ChrX: 91090804	*PCDH11X*	C/G	Non-synonymous coding	P101A	Tolerated 0.3	Benign	rs1296070582	0.002	No effect

Table-3: List of variants identified in genes with burden of ultra-rare variants. [a]Genomic position of the nucleotide base according to GRCh37,NCBI. [b]Gene harboring the variant [c]Consequence of the variant, Nomenclature of variants in introns or UTR with respect to first base of the corresponding cDNA. The position of the variant is given according to cDNA of the longest protein-coding transcript of the respective gene. [d]Amino acid position for missense variants, [e,f,g] Predictions of pathogenicity of variants.

In this thesis, in chapters 2 and 4, we provide evidence of the involvement of microtubule-associated, centrosome, and cilia-associated molecules in juvenile myoclonic epilepsy, the predominant form of GGE. Our findings here of novel, rare, and potentially pathogenic variants in PCNT, RSPH4A, centrosome, and cilia-associated proteins, suggest a network of ciliary molecules underlying the disease.

Pericentrin (*PCNT*) gene is present on the 21st chromosome and codes for 3336 amino acid long- a highly conserved microtubule-associated protein. Pericentrin aids in microtubule organization. The protein associates with gamma-tubulin ring complex and other centrosomal proteins and is vital for microtubule nucleation and cell cycle regulation (Doxsey et al., 1994). PCNT is important for cilia formation and signaling (Jurczyk et al., 2004). Frameshift and nonsense mutations in PCNT have been associated with Seckel syndrome (Griffith et al., 2008) and microcephalic osteodysplastic primordial dwarfism (MOPD II) (Rauch et al., 2008), both of which display a reduction in body and brain sizes. In this study, a variant in *PCNT* gene was identified in two affected members of the family, GMW13. Here, the affected individuals, a male, and a female, were born to unaffected parents. The missense variant, c.5164A>G, corresponding to p.Gln1686Arg, occurs at a residue well conserved in primates, higher mammals, and Xenopus. Analysis using MutPred (Pejaver et al., 2020) predicts the gain of methylation on changing Gln at 1686 to Arg. Protein methylation is a covalent posttranslational modification often occurring on the carboxyl group of Gln, Lys, etc. The change of the Gln to Arg and gain of methylation at the residue could influence the protein-protein interactions as methylated Arg has been demonstrated to have an increased affinity for the indole ring of tryptophan residues (Bedford & Clarke, 2009). It would thus be interesting to test the influence of the mutation on the localization and cell division-related functions of PCNT.

Along with a potentially pathogenic variant in *PCNT*, in the family GMW13, we found novel, rare, and potentially pathogenic variants in genes *KCNB1, PDE4DIP* - genes with the burden of ultra-rare variants associated with epilepsy. *KCNB1* codes for a potassium-voltage gated channel subfamily B member 1. The gene is present on the reverse strand of chromosome 20 and is highly expressed in the brain, is associated with epileptic encephalopathy and Schizophrenia. The protein is 96-kDa and is made up of 858 amino acids. It forms the α-subunit of the voltage-gated potassium channel subfamily 2 (Kv2.1). The subunit consists of six transmembrane helices. Heterozygous missense, truncating variants arising de novo have been identified in individuals with epileptic encephalopathies (Bar et al., 2020; Torkamani et al., 2014). So far, The variant identified here, c. 2285A>C leading to p.Asp762Ala lies in the C-terminal of the protein beyond the transmembrane domains. The depth of coverage at this base during WES is 59. It is interesting to note that while mutations in the transmembrane regions, voltage sensor domain, or the pore region of the protein are associated with a more severe phenotype of epilepsy and global developmental delay, C-terminal domain variants are associated with a less-severe epileptic phenotype (Bar et al., 2020; de Kovel et al., 2017). It would be important to validate the identity of this variant through Sanger sequencing and use cell culture assays to test the localization and loss of ion selectivity in the mutant.

The human Myomegalin gene (*PDE4DIP*) encompasses 224 kb on chromosome 1q12 for phosphodiesterase 4D interacting protein. It codes for a 2346 amino acid protein implicated in Schizophrenia and adult pineoblastoma. The gene is conserved in rhesus monkey, dog, cow, mouse, rat, chicken, and zebrafish. In humans, this gene is predicted to encode 24 isoforms, 8 of which have evidence at the transcript sequence. Alternate splicing leads to 8 isoforms with differences in their N-terminus,

C-terminus, and internal sequences (Roubin et al., 2013). Intracellularly, the protein helps to anchor phosphodiesterase 4D to the Golgi/centrosome region of the cell (Bouguenina et al., 2017). It is also important for microtubule nucleation and cell migration (Roubin et al., 2013). The variant identified in this gene in the family GMW13 is in the 5'-UTR region, 103 bases upstream of the transcription start site (TSS). Across different isoforms, the variant at Chr1-145075965-A/T is unreported across 1000 Genomes, gnomAD, ExAC, etc., suggesting that it is a novel variant with a CADD score of 12.19. The variant occurs at a residue well conserved across humans and mammals such as rhesus monkey, dog, cow, mouse, rat, etc. The identity of the variant remains to be confirmed using Sanger sequencing. While bio-informatic tools predict no effects due to this variant, luciferase reporter assays in cell culture may be informative of the effect of the variant on transcription.

GLH13 is a three-generation family with three affected individuals. A male affected member in the first generation was deceased, and hence the sample was unavailable. The inheritance of GGE skipped a generation, and in the third generation, two siblings, a male, and a female, had GGE, and their samples were used for NGS analysis. Among these individuals, we identified variants overlapping with the EPI25K dataset for GGE genes with the burden of ultra-rare variants in two genes - *PCDHGA8* and *RSPH4A*. These variants had coverage depth of 147 and 53; quality scores of the bases were 1824 and 714, respectively. Variants in both the genes are non-synonymous coding variants.

The gene *PCDHGA8,* at the 5q31.3, forward strand, codes for Protocadherin gamma-A8, a brain enriched protein conserved across chimpanzee, rhesus monkey, dog, cow, mouse, and rat. It is predicted to have two isoforms and is suggested to be involved in the establishment

and maintenance of specific neuronal connections in the brain. The cellular roles of the protein remain to be examined. It is predicted to localize to the cell membrane. The variant in the gene identified here at c.1523A>C leads to p.Asp508Thr. The CADD score is 25. The change at Chr5-141394336-A/C occurs at a conserved location among humans, primates, dog, mouse, rat, and xenopus. The variant is identified with rs1205005872 and is a rare variant occurring at a frequency of 6.993e-06. The change is predicted to be possibly damaging or tolerated using Polyphen and SIFT prediction tools, respectively. Considering that the cellular and functional correlates of the gene and further the variant remain unknown, it would be challenging and important to validate our findings and expand the study.

Another variant of significance in this family is in the gene *RSPH4A.* The gene at 6q22.1 encodes seven exons for the protein; radial spoke head component 4A. The protein shows tissue enrichment in the brain, fallopian tube, and pituitary gland, as seen in the human protein atlas (Uhlen et al., 2015). The protein is a component of the evolutionarily conserved axonemal radial spoke head playing important roles in ciliary motility. Mutations in the gene are associated with ciliopathies (Oud et al., 2017). In this study, we identify a rare, heterozygous variant in the gene associated with GGE. The variant at Chr6- 116937974-C/T is a missense variant, p.Pro63Leu. The CADD score is 15.79. The amino acid proline at this position is well conserved across human, chimpanzees, rhesus monkey, sheep, dog, elephant, etc. It is poorly conserved across rodents. Bio-informatic predictions suggest that variant is deleterious to the protein function. Considering that the cellular localization, function, and effects of human mutations on the cellular attributes of this protein are unexplored. It is an interesting, novel, and possibly clinically relevant gene for further exploration. On validation, the variant and further the

cellular functions of the protein could help strengthen the possibility of microtubule-associated protein networks that are relevant to GGE and brain development. This network could possibly comprise additional ciliary proteins such as EFHC1, EFHC2, PCNT, CDC20B, CILK1, etc.

In the family, NIH174, four individuals, all males, in the same generation were affected by GGE. DNA samples from the proband and an additional affected member were subjected to WES. The variants were identified in common with the EPI25 database in the genes *TTC7A* and *XPO5*. *TTC7A* codes for tetratricopeptide repeat domain 7A, a ubiquitously expressed gene conserved among chimpanzee, rhesus monkey, dog, cow, mouse, rat, chicken, zebrafish, and frog. The gene encodes a protein containing tetratricopeptide repeats. Mutations in this gene are reported in early-onset inflammatory bowel disease and intestinal atresia (Ashton et al., 2020; Sharafian et al., 2018). The mutant proteins disrupt intestinal development. Alternative splicing results in multiple transcript variants. The protein is nuclear-localized and is important for chromatin compaction, genome stability, and cell viability (El-Daher et al., 2018). This gene has a brain- and muscle-expressed paralog tetratricopeptide repeat domain 7B (TTC7B) with 49.47% sequence identity, and both molecules are components of the conserved PI4KIIIα complex (Jardine et al., 2019). TTC7 protein contains TPR domains, which are helix-turn-helix structurally conserved motifs, with roles in protein scaffolding. The variant identified in this family is p.Ser153Phe, which occurs in the first TPR domain of the protein. The amino acid serine is well conserved among humans, rhesus monkey, mouse, dog, elephant, and Xenopus. The CADD score is 22.5, and the change is predicted to be possibly damaging to the function of the protein. Considering that other mutations in the gene have not been reported in brain development or disease, it would be important to confirm this variant examine other possible causative

variants in the family. Also, a scan for the occurrence of the mutations in this gene additional GGE case and controls to establish the significance of our finding.

XPO5 codes for Exportin 5, a protein involved in nuclear export of micro-RNA precursors, which form short hairpins (Yi et al., 2003); it also mediates nuclear export of proteins bearing a double-stranded RNA binding domain (dsRBD) and double-stranded RNAs (cargos). XPO5 mutations and polymorphisms are implicated in a wide range of cancers such as head and neck cancer (Ozdas et al., 2021), thyroid cancer (Wen et al., 2017), hepatocellular carcinoma (D. Lin et al., 2020). In this family, NIH174 and in members of SCT229, we identified a 5'-UTR variation in the XPO5 at Chr6- 43543714-T/G, occurring 113 bases from the transcription start site. The CADD score at this base is 11.33, the reference allele, T, is well conserved across primates, rat, mouse, zebrafish etc. Although the roles of SNPs in XPO5 are well studied in miRNA processing and cancer, the role of this protein and its mutations in brain development and neurological disorders remain unknown. Considering that the same variant was identified in 4 patients of the 55 samples examined here, increasing the sample size to confirm the rarity of this allele would be necessary and important. Along with this variant, in the family SCT229, with two affected members sequenced here, we find variations in genes SLC9A1 and PCDH11X.

SLC9A1, solute carrier family 9 member A1, is an antiporter for sodium/hydrogen ions. It localizes to the plasma membrane and is ubiquitously expressed across tissues. Mutations in the gene have been identified in patients with ataxia–deafness Lichtenstein–Knorr syndrome (Guissart et al., 2015). The variant identified here is a heterozygous 5'-UTR variant at Chr1-27481308-A/G, c.-483T>G,483 bases away from the

transcription start site. The reference allele, A while, is well conserved in humans, rhesus monkey, mouse, dog, and elephant; the alternate allele G is reported in species such as naked mole rat, guinea pig, and hedgehog. The CADD score here is 11.87. The poor conservation, along with the distance from the TSS, may hint that this allele, as suggested by bio-informatic predictions, is not deleterious.

In the family, SCT229, we identified a variant at X-91835805-C/G in the gene *PCDHX11*, Protocadherin-11 X-linked. Data from the human protein atlas shows that the protein is enriched in the brain, ovary, and placenta. The gene is conserved in chimpanzee, dog, cow, mouse, rat, chicken, zebrafish, and frog. It is identified as a candidate gene for developmental dyslexia (Veerappa et al., 2013). In this study, the variant allele reported leads to p.Pro101Ala change at a residue with CADD score 0.002 and is reported at an allele frequency of 0.0001048, among South Asian population, has been reported with ID rs1296070582. The individuals sequenced here are a father-daughter pair, and we find mutation as hemizygous in the father and heterozygous in the daughter, both affected with GGE suggestive of a possible X-linked dominant inheritance and effect of the variant. The variant, however, is predicted to be tolerated and has a very low CADD score, and is unlikely to be a candidate variant in this family. Thus, in this family, the overlap of our data with the Epi25 GGE candidate gene list did not identify a promising candidate gene variant.

SCT290 is a three-generation family with four males and two females in the third generation affected by GGE. Of these four males, one individual was deceased. We performed whole-exome sequencing on four individuals, two males and two females, all affected by GGE. Among these individuals, the total number of shared variants was 1583. Of these, 43

were shared novel single nucleotide variants. We asked if any of the genes with SNVs in the family were observed in the Epi25 GGE list of genes with the ultra-rare variant burden. We identified a novel, heterozygous variant in the gene *RWDD4*, at Chr4- 184580325-G/A, a 5'-UTR variant occurring 220 bases upstream of TSS. The reference allele is well conserved among primates and mammals - humans, rhesus, mouse, dog, elephant, etc. The CADD score is 11.87. The variant identity and segregation remain to be confirmed by Sanger sequencing. Analysis across databases such as human protein atlas and human RNA-seq data shows that the gene is expressed ubiquitously (Fagerberg et al., 2014). The cellular localization, biological roles, and functions of the protein are unknown. The RWD domain of about 110 amino acids is predicted to facilitate protein-protein interactions. Thus, a possible association of a novel gene to GGE was identified in the affected individuals.

5.3 Discussion

Identifying a genetic cause in small families with limited family history is cumbersome. In these cases, finding candidate protein-altering variants is possible using next-generation sequencing technologies. Heritable or de novo protein-coding changes represent most disease-causing variants accounting for about 60% of total causative genome variations (Botstein & Risch, 2003; Petersen et al., 2017). Exome sequencing analysis allows for targeted sequencing of every protein-coding region of the genome. It leverages the well sequenced, mapped reads with *in silico* predictions of the effect of variants on protein function to identify candidate gene variants. The evolution of efficient sequencing tools with high coverage, read depth, and quality supported by powerful bioinformatic tools has shifted the field of gene discovery in epilepsy from multi-step loci discovery

followed by resequencing to the multiplexed analysis of nearly every protein-coding gene simultaneously.

Initial application of this technology to GGE involved the exome sequencing analysis of 118 IGE patients and 242 controls. The candidate variants identified were further screened through genotyping, in a larger cohort of 878 IGE-affected subjects and 1830 controls (Heinzen et al., 2012). In this study, no single gene variant reached the statistical significance level of association or accounted for more than 1% of the cases. However, variants in a few genes were enriched in patients. These were in *GREM1, OR10S1, PPEF2 and CHD1*. Further, using gene set-based analysis of exome sequencing data, GABAA receptors were associated with GGE (May et al., 2018).

Consortia-based large-scale studies such as Epi4K (Ref) and Epi25K (Ref) have revolutionized the field by reporting rare, ultra-rare variants and genes with the burden of disease causation. For example, in the ongoing Epi25 Collaboration, whole-exome sequencing analysis of 13,487 epilepsy-affected individuals and 15,678 control individuals has been performed. Among these, 3108 individuals with GGE have been sequenced. In this analysis, compared to controls, individuals with epilepsy carried an excess of ultra-rare, deleterious variants in constrained genes and in genes previously associated with epilepsy. Even though no single gene surpassed exome-wide significance among individuals with GGE, highly constrained genes and genes encoding ion channels, such as *CACNA1G, EEF1A2, and GABRG2,* were top candidate genes. Also, the top 200 ranked genes associated with GGE were identified (Epi25 Collaborative. Electronic address & Epi, 2019). The lack of a single significant finding illustrated the need for NGS analysis in many samples or co-segregation analyses in multiplex families. The WES dataset also allows for the

probing of concurrent variations, for grouping of variants into a gene set to examine the association of any or all variants within a group of genes to the disorder, as has been reported for GABAA receptors and genetic generalized epilepsy (May et al., 2018).

Here, we performed whole-exome sequencing of 55 individuals from 26 unrelated families with genetic generalized epilepsy. Our aim here was two-fold. To identify rare, potentially pathogenic gene variants in the individuals sequenced. The second aim was to examine the data set for potential disease-causing variants in genes reported as carrying a burden of ultra-rare variants in GGE. The first step towards these was library preparation and sequencing. Here, the whole-exome sequencing library preparation resulting in high-quality reads with at least 70% of bases with a minimum Phred score of 20 was obtained in 53/55 samples sequenced. This accounts for a success rate of 96%, with over 96% of high-quality reads and around 90% of high-quality bases. The average read length was around 95 across samples, and over 80% of the reads aligned to the reference exome serving as indicators of confidence of the data set used for analysis. The in-house pipeline used for bioinformatic analysis with stringent filtering criteria allowed the elimination of commonly occurring variants, poor quality reads, and non-segregating variants within the individuals of the same family. Upon multiple filtering steps, around 1500 shared variants were identified within members of each family. The data was analyzed to prioritize variants of significance for a novel gene variant occurring commonly in the GGE patients sequenced here. No commonly occurring novel disease variant was found. In future studies, the examination of common gene sets or pathways using gene ontology analysis may be informative.

As the first step towards prioritizing variants, we searched for our data set for genes with the burden of ultra-rare variants, as reported in the EPI25K database for GGE. Among the 200 candidate genes, we found variants in five families in the genes *KCNB1, PDE4DIP, PCNT, SLC9A1, XPO5, PCDH11X, PCDHGA8, RSPH4A, TTC7A, RWDD4.* Of the ten variants, eight were novel, and two were reported at minor allele frequency less than 0.005. In this subset of ten genes, we found that three genes are microtubule-associated and are possibly crucial for microtubule nucleation, cell division, migration, and cilia formation - *PDE4DIP, PCNT, RSPH4A.* A growing body of evidence indicates that centrosomes, cilia, and cytoskeleton components play crucial roles in some forms of epilepsy. For example, mutations in EFHC1, a microtubule-associated protein involved in cell division and corticogenesis during brain development, causes JME (de Nijs et al., 2009; de Nijs et al., 2012).

Similarly, variants in the cilia expressed *CILK1/ICK* causing impaired mitosis, cell-cycle exit, and promotion of apoptosis have been identified in JME (Bailey et al., 2018). Animal models of CILK1 and EFHC1 showed tonic-clonic seizures. Also, the mutant mice displayed polyspikes on EEG more frequently than did wild-type littermates (Bailey et al., 2018; Suzuki et al., 2009). EFHC2, reported in this thesis, is a microtubule-associated protein, mutants of which display cell division defects are associated with JME. Notably, all these proteins localize at similar sub-cellular locations, and mutants cause cellular phenotypes of cell division defects, defects in cell migration, and cortical development. The correlates of these, in humans could be the regional abnormalities of the subcortical and cortical grey matter that have been observed in quantitative MRI studies in patients with JME (J. H. Kim, 2017). Mutations leading to spindle misorientation as shown for EFHC1, EFHC2, CILK1 have been more recently identified in patient-derived induced pluripotent stem cells (iPSCs)

with mutated cells PCDH19 allele for clustering epilepsy. Using iPSCs and 3D-brain organoids, it was demonstrated that misoriented mitotic spindles, increased asymmetric cell division of NPCs and misregulated mitosis are the likely underlying molecular events leading to decreased brain organoid size (Borghi et al., 2021). Thus, various lines of evidence point to critical roles of spindle-associated proteins - brain anomalies in humans with *CILK1* mutations, *Efhc1* and *Cilk1* mutant mouse with aberrant cortical development and further *PCDH19* iPSCs and brain organoids showing organizational defects. Along these lines, it would be critical to validate the variants in *PDE4DIP, PCNT, RSPH4A* and examine if these microtubule-associated proteins also have similar spindle pole defects in cells and if they affect cortical development. It is also enticing to expand these studies to generate mutations in combinations as informative experimental tools to understand the pathogenesis of GGE.

In the future, this study would benefit from a functional enrichment analysis of the list of shared rare variants to discover biological processes overrepresented in our gene list. These could be done using the WEB-based Gene Set Analysis Toolkit (Web-Gestalt) [http://www.webgestalt.org] and Reactome Knowledgebase [www.reactome.org]. The results of the investigations could be informative of an underlying molecular network contributing to GGE. Also, a limitation is that sequential filtering steps used here to narrow down the number of variants of interest may have been too strict, and other pathogenic variants might have been missed.

In conclusion, we have described variant prioritization and generated disease-gene lists that empower the search for GGE- genes in family-based sequencing studies. This study forms a platform for much of the further analysis. It would benefit from variant validation, segregation analysis,

analysis of additional novel genes and variants in each of the families sequenced here.

Appendix Chapter 1

Generation of *Sox30* loss-of-function mouse

Summary

Sox30 is a member of the SRY-related high-mobility-group (HMG)-box-containing (SOX) transcription factors. Mutations in SOX30 have been associated with juvenile myoclonic epilepsy (Ratnapriya, 2009) and male infertility (Jaishankar, 2021). In mouse, SOX30 was reported to play a critical role in mouse spermatogenesis (C. A. Feng et al., 2017). Sox30 is important for meiosis in both sexes. However, in the absence of SOX30, the acrosomal vesicle formation is disrupted in the male reproductive system, and spermatid development is arrested (Bai et al., 2018). Earlier work in the laboratory involving a family-based linkage study in a large family affected by juvenile myoclonic epilepsy, SOX30 was identified as a causative gene (Ratnapriya, 2009). On sequencing SOX30 in 480 individuals with juvenile myoclonic epilepsy, 15 additional rare SOX30 coding variants were identified. A recent study in the lab aimed at screening SOX30 gene mutations in male infertility led to the identification of 15 rare potentially pathogenic variants in individuals with azoospermia (Jaishankar, 2021). These data have suggested the involvement of SOX30 in epilepsy and male fertility among humans. To facilitate functional elucidation of SOX30 and the molecular mechanisms underlying its contribution to development and disease, we embarked on generating SOX30 mutant mice. In agreement with the findings in the literature, Sox30 homozygous mutant male mice were found to be infertile. Homozygous mutant female mice are fertile. In this chapter, I detail the

approach used to generate the Sox30 mutant mice and their genetic characterization.

A1.1 Introduction

Sox proteins are highly conserved DNA-binding proteins belonging to the HMG family of transcription factors. Based on conservation of the DNA-binding HMG domain, the *SOX* family branches into eight sub-categories (Sox A-H; Figure 1). To date, in vertebrates, thirty proteins of this family have been identified. In the mouse and human genomes, twenty orthologous pairs of *SOX* genes have been reported (Schepers et al., 2002).

The SOX family of proteins employ sequence specificity to bind to DNA. They bind to a heptameric sequence, 5′-(A/T)(A/T)CAA(A/T)G-3′, and its variants (Harley et al., 1994). While the binding of DNA does not cause alterations to the structure of the SOX proteins, it results in bending of the DNA – resulting in the widening of the minor groove and unwinding of DNA. This feature allows for the recruitment of other transcription factors and DNA modifiers, leading to gene regulation (Wegner, 1999). The SOX family of proteins show spatiotemporally regulated tissue-specific expression patterns. While each SOX gene has a unique expression pattern, within each subfamily, there is an overlap permitting additive and redundant functions (Angelozzi & Lefebvre, 2019). This can be elucidated using the *SoxB1* members as an example: *SoxB1* consists of *Sox1-3*. *Sox1* is necessary to differentiate neural progenitors from the ganglionic eminence and dorsal telencephalic wall (Kan et al., 2007). *Sox2* expressed in multipotent cells is vital for maintaining cell proliferation potential, and its overexpression inhibits neurogenesis (Bani-Yaghoub et al., 2006). *Sox3* is transiently expressed by proliferating and differentiating neural progenitors in the olfactory bulb and dentate gyrus (T. W. Wang et al., 2006). While there exists redundancy between members of this group, the

SoxB2 members *Sox14* and *Sox21* counteract the functions of SoxB1. While the SoxB2 group allows for maintenance of proliferating progenitor population, SoxB1 proteins inhibit SoxB2 and allow progression from progenitor cell to neurons (Whittington et al., 2015). In other examples, members across subfamilies can substitute for the function, as in the case of Sox2. The roles of *Sox2* in reprogramming can be substituted by *Sox18* and *Sox17* (Nakagawa et al., 2008). Thus, the Sox proteins have complex regulation within and across this group. They also show possible redundancy. These factors allow for fine, contextual transcriptional regulation by Sox proteins.

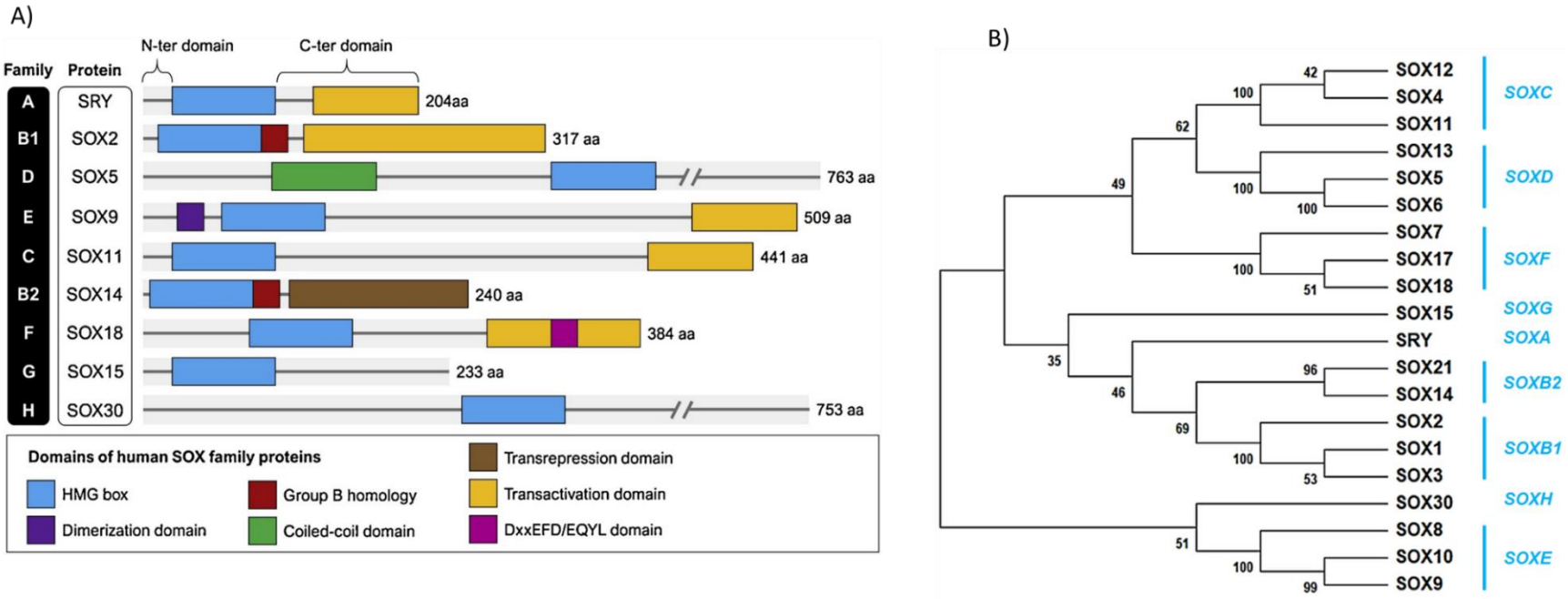


Figure 1: SOX families and their evolution. **(A)** The highly conserved and characteristic HMG box is specified alongside other functional domains, including the transactivation domain. Domains of SOX family proteins are shown in the box (Grimm et al., 2020). **(B)** A rooted phylogenetic neighbor-joining tree for the human SOX full-length proteins showing SOX30 as a recently evolved protein (Mehta et al., 2019)

Mutations in SOX genes have been reported to be associated with neurodevelopment, spermatogenesis, and cancers, among other human disorders. *SOX11* deletion or mutation causes Coffin-Siris phenotype (Hempel et al., 2016). This disorder is marked by developmental delay/intellectual disability, feeding difficulties, facial dysmorphology, and microcephaly. Haploinsufficiency of *SOX5* causes intellectual disability and behavioral deficits (Schanze et al., 2013). Duplications in *SOX3* have been identified in patients with intellectual disabilities (Arya et al., 2019). Within the same subgroup, the phenotypes caused due to mutants are varied. For example, the loss of Sox8 and Sox9 leads to male infertility in mice (Barrionuevo & Scherer, 2010). However, mutation in *Sox10* affects neural crest development (Pingault et al., 2013). *SOX9* mutations in humans cause campomelic dysplasia and XY sex reversal (Wagner et al., 1994), while *SOX8* mutations are found in increased frequencies in azoospermic men and in women with primary ovarian insufficiency (Portnoi et al., 2018). *SOX10* is important for neural crest specification and is downregulated in individuals with schizophrenia (Iwamoto et al., 2005). Thus, examination of roles of SOX proteins across cell types and tissues is important to understand mutant phenotypes.

It is also known that aberrant expression of Sox proteins can lead to cancer development. Overexpression of SOX4 has been observed in multiple tumor types - breast, prostate, bladder, hepatocellular carcinoma, medulloblastoma, and small cell lung carcinoma. SOX4 upregulation also causes tumor aggressiveness (Mehta et al., 2019). There exists a correlation of upregulation of a few SOX proteins and cancer progression SOX2, SOX4, SOX5, SOX8, SOX9, and SOX18 (Grimm et al., 2020). Conversely, a few SOX proteins are downregulated during the progression of cancers– SOX1 (L. Song et al., 2016), SOX7 (Stovall et al., 2013), and SOX17 (L. Li et al., 2018). Interestingly, upregulation of

SOX11 in glioblastoma (Korkolopoulou et al., 2013) and SOX30 in lung adenocarcinoma (Hao et al., 2018) offers favorable outcomes and increased survival in patients. These data are imperative to the multifaceted roles of SOX proteins which have been explored using a combination of approaches.

Animal models are a versatile tool to study the roles of such multifaceted molecules. They allow for the examination of the protein and its function across stages of development and disease. One of the Sox group members recently identified as essential for male fertility in mice is Sox30. Sox30 belongs to the SOX-H subgroup. Sox30 in the mouse genome is present on the 11th chromosome, on the positive strand at the location Chr11:45871137-45908821. The transcript generated is 2997bp long and codes for a 782 amino acid protein. Sox30 consists of a DNA binding HMG domain and can act as a transcriptional activator and repressor. It binds to its promoter and activates its transcription (D. Zhang et al., 2018).

Sox30 was first identified in a yeast two-hybrid screening of Wilms' tumor suppressor (Wt1) -associated proteins using mouse testis cDNA library (Osaki et al., 1999). In humans, SOX30 is expressed in the human brain regions (Ratnapriya, 2009). It is expressed specifically in mouse fetal ovaries, pubertal testis, and adult male germ cells (C. A. Feng et al., 2017). DNA methylation at CpG islands of Sox30 is reported to regulate the expression by silencing. Sox30 is hypo-methylated in adult mouse testis, epididymis, and lung, where Sox30 is expressed. Sox30's hypermethylation is correlated with reduced expression (Han et al., 2014). In zebrafish (Danio rerio), SOX30 is expressed in the central nervous system. SOX30 is predominantly expressed in early embryos at the presumptive midbrain-hindbrain boundary (mhb). The mhb is a putative

organizing centre important for midbrain induction and patterning (De Martino et al., 1999). In common carp (Cyprinus carpio), Sox30 is expressed abundantly in female and male gonads, especially in spermatocytes and spermatid/sperm of carp testis. Lower expression is seen in the brain, liver, muscle, and kidney (Anitha & Senthilkumaran, 2020). These data suggest that Sox30 may have evolved species-specific roles and corresponding differential expression patterns. The cellular functions of Sox30 have been studied in further detail in the mouse reproductive tissues. Sox30 null male mice are infertile (C. A. Feng et al., 2017). They have impaired testis, germ-cell differentiation arrest, and irregular Leydig cell proliferation, resulting in the complete absence of spermatozoa. These defects are restored on re-expression of Sox30 in Sox30-null mice (Han et al., 2021). Further, at the molecular level, Sox30 regulates the expression of post-meiotic genes such as *Fhls, Tnp1, Tnp2, Prm1, Prm2 and 3, Odf1, Dnmt1, Catsper1, Smcp, Hils1, Ccdc54,* and *Tsks* that play roles in round spermatid maturation (Bai et al., 2018; S. Zhang et al., 2018). More recently, Sox30 was shown to promote male differentiation via sex differentiation genes - *Rspo1, Foxl2, Sox9,* and *Ctnnb1*. It also regulates key meiotic genes such as *Cyp26b1, Stra8,* and *Rec8.* While the molecular details of the involvement of SOX30 in meiosis and spermatogenesis have been studied, the brain-related roles remain less explored. To examine the involvement of SOX30 in the brain and to study the effect of the human pathogenic alleles for epilepsy and azoospermia in the SOX30 null background, we generated SOX30 mutant mouse lines.

A1.2 Materials and methods

A1.2.1 Animal breeding and maintenance

C57BL/6NCr, B6.β-actin Cre, *Sox30*$^{flox/flox}$, and *Sox30*$^{lox/lox}$ mice strains were housed in groups under a 12-hour light/dark cycle with access to food and water ad libitum. Transgenic lines were maintained as homozygotes. All the progenies were weaned post 21 days, earmarked, and tail clipped for genotyping. Mating was set up either in the ratio of 1:1 or 1:2 (male: female). All experimental procedures followed the guidelines set by the Institutional Animal Ethics Committee (IAEC) of the JNCASR.

A1.2.2 Generation of *Sox30* floxed mice

The homozygous floxed *Sox30 (Sox30* $^{flox/flox}$) mice were generated in collaboration (Dr. Shyam K. Sharan, National Cancer Institute, National Institute of Health, USA) using the following strategy: The gene-targeting vector (Figure 2) introduced into mouse embryonic stem cells to generate floxed *Sox30*, tm1a allele. The vector consists of 5.8Kb of the 5'-homology arm mapping to intron 1 of *Sox30* and 3' homology arm of 4.3Kb mapping to intron 3. The critical region consisting of exon2 and exon3 is flanked by LacZ reporter flanked by FRT and loxP recognition sites for flippase and Cre recombinase enzymes. Random integrations were selected against using negative selection for Diptheria toxin. The recombinants resistant to neomycin were chosen to obtain floxed alleles in the 129SVJ X C57BL/6NCr background. The heterozygous *Sox30* $^{flox/+}$ mice were mated to get homozygous *Sox30* $^{flox/flox}$ mice. The boundaries of homology arms, reporters, and the DNA between recombinase sites were confirmed using PCR amplification and sequencing. All mouse experiments were performed following the guidelines of CPCSEA (Committee for the Purpose of Control and Supervision of Experiments on animals).

A1.2.3 Generation of *Sox30* loxed mice

The homozygous floxed *Sox* $^{flox/flox}$ was mated to β-actin driven Cre recombinase expressing mice to generate constitutive deletion of exon 3 of *Sox30* to generate β-actin Cre; *Sox30* $^{lox/+}$ and β-actin Cre; *Sox30* $^{lox/lox}$ mice. Sox30$^{+/-}$ males and Sox30$^{-/-}$ females were intercrossed to obtain additional homozygous loxed alleles, Figure 3.

A1.2.3 Genomic DNA isolation and genotyping

1 cm tail-clippings were chopped into 4-5 pieces and resuspended in 180µl of 50mM NaOH. The samples were placed on a thermo shaker at 90°C, 1200 RPM, for 15 minutes. 20µl of 1M Tris at pH 8.0 was added, vortexed thoroughly, and centrifuged at 13000 RPM for 15 minutes. The supernatant was collected and stored at -20°C. This protocol is a modification of the HotSHOT protocol (Truett et al., 2000). 1µl of the supernatant was used for genotyping. The primer pairs used are detailed in Table 1 in Appendix of the chapter. The PCR products were separated on 0.8 or 2% agarose gel. The products from genotyping PCR were confirmed to be specific by sequencing analyses.

A1.2.4 Cell culture and transfection

HEK293T cells were maintained in DMEM (high glucose, Sigma, D1152) supplemented with 10% heat-inactivated fetal bovine serum (MP Biomedicals, USA, 29167), 2mM L-glutamine (Sigma, G7513), and antibiotics (100U/ml Penicillin and 10mg/ml streptomycin; Sigma, P4333) in a humidified atmosphere of 5% CO_2 at 37°C. Cells were grown to 60-70% confluence in 6-well culture dishes or on poly-L-lysine (Sigma, P2658) coated 18 mm glass coverslips and transiently transfected as follows: An hour before transfection, 10% complete DMEM was replaced with DMEM with 2 mM L-glutamine. Transfections were carried out using

Lipofectamine®2000 (Invitrogen, Thermo Scientific, Massachusetts, USA) at plasmid DNA: Lipofectamine ratio of 1:1.5. After 5-6 hours, transfection media was replaced by 10% complete DMEM. Cells were harvested for lysate preparation or fixed for imaging after 24-48 hours of transfection.

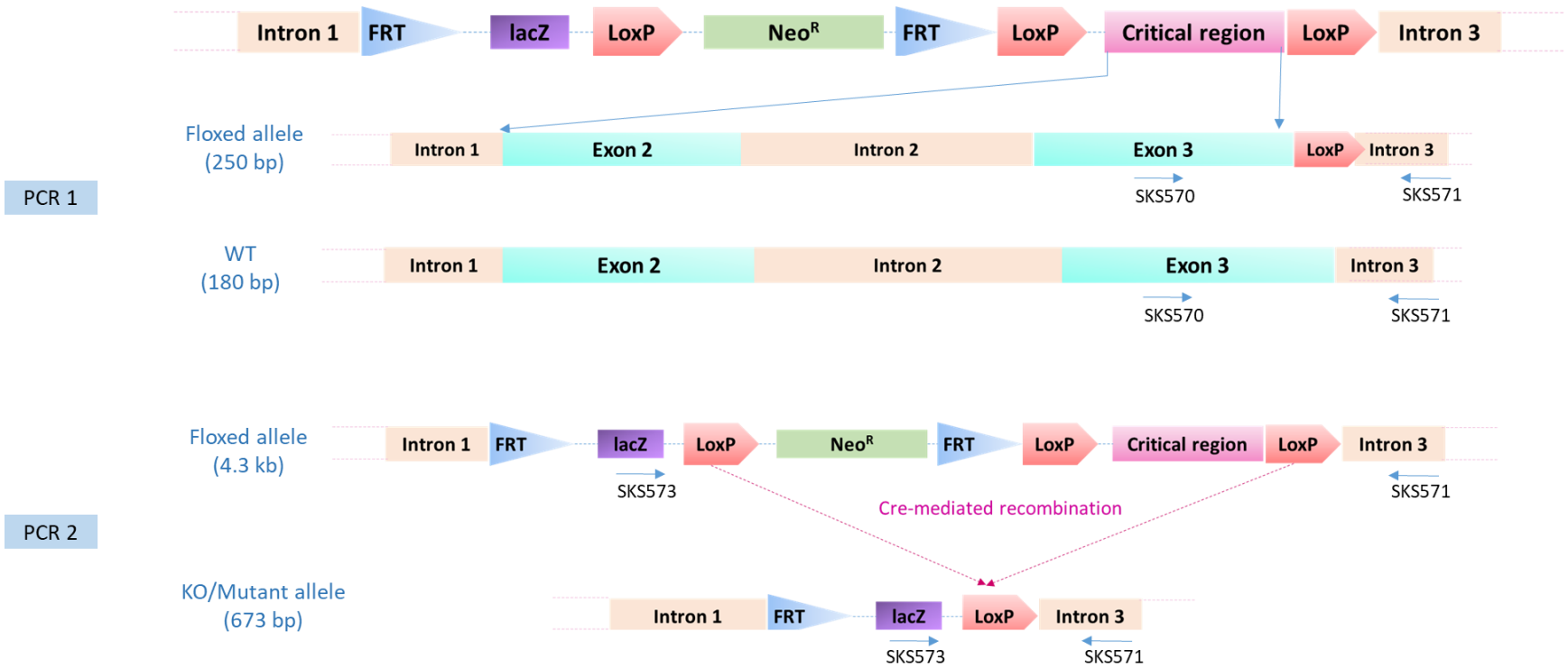


Figure 2: Illustration of the gene-targeting vector with primer binding sites marked in blue arrows for genotyping. The outcome of Cre-mediated recombination that deletes the critical regions is shown with the genotyping strategy.

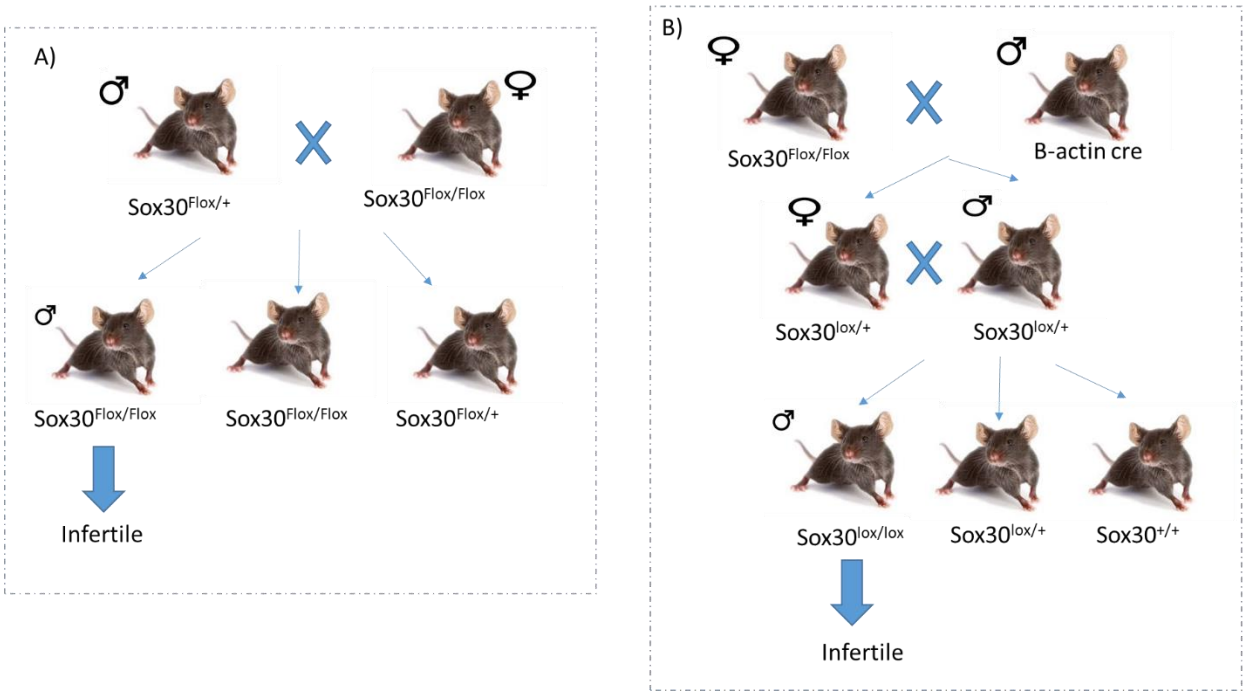


Figure 3: Breeding and outcome: (A) Breeding to obtain Sox30 floxed alleles. (B) Mating pattern used to generate Sox30 loxed alleles.

A1.2.5 Western analysis

Animals were decapitated, and whole brain was extracted. The tissue was held in ice-cold 1X PBS on a filter paper on ice; cortex, hippocampus, and cerebellum were separated using sterile scalpels. The separated tissue was weighed and resuspended in RIPA buffer. The tissue was homogenized using a handheld homogenizer in ice. The solution obtained was centrifuged at 13000 RPM for 30 minutes at 4°C. The supernatant collected was used as tissue lysate, snap-frozen in liquid nitrogen, and stored at -80°C.

Total protein concentration in the lysates was estimated using the bicinchoninic acid assay (BCA) reagents (Sigma). For Western analysis,

aliquots of 50-100 µg of protein were mixed with SDS gel-loading buffer, boiled for 5 minutes, and then resolved electrophoretically on a 10% or 8% SDS-polyacrylamide gel (Laemmli, 1970). The proteins on the gel were subsequently electrotransferred (Bio-Rad, USA) to equilibrated nitrocellulose membrane (Amersham Protran, GE or BioTrace™ NT, Pall life sciences) at constant 400mA for one hour and fifteen minutes. The transfer was confirmed by Ponceau S staining followed by washes in distilled water. Blocking was performed using 3% skimmed milk powder (Sigma) and 2% Bovine Serum Albumin (BSA, Sigma) in 1X PBS with 0.05% Tween 20, for 12 hours at 4°C. The membrane was probed with a primary antibody for 10 hours at 4°C. The membrane was washed in 1X PBS with 0.05% Tween 20 at room temperature for 5 minutes. The blot was further incubated for one hour at room temperature with a secondary antibody in 1X PBS containing 1% BSA, 1% skimmed milk powder, and 0.05% Tween 20. The membrane was washed twice for 10 minutes each at room temperature in 1X PBS and 0.05% Tween 20. The protein bands were detected using an enhanced chemiluminescent substrate for HRP (Pierce-Thermo Fischer Scientific, USA or Clarity™ Western ECL Substrate - Bio-Rad).

A1.2.6 Generation of custom antibody

To identify hydrophilic regions, the proteins were scanned for peptide regions of length 20-25 amino acids. The epitopes were chosen using Protein Hydroplotter software from protein lounge (proteinlounge.com), Kyte-Doolittle plot (Kyte & Doolittle, 1982), and Hopp-Woods plots (Hopp & Woods, 1981). Antigenic regions were identified using the Kolaskar Tongaonkar antigenicity scale (Kolaskar & Tongaonkar, 1990). Selected sequences were validated in the antibody epitope prediction tool from the immune epitope database (IEDB). For SOX30, the peptide

RDYPFRDYPDEHTHSEDSR (697-715) at the C-terminus was chosen for antibody generation in two rabbits. The peptide synthesis, antibody generation, and affinity purification were conducted with Abgenex, Bhuvaneshwar, India.

A1.3 Results

A1.3.1 Generation and validation of *Sox30* floxed allele

To generate the floxed mice for *Sox30*, two loxP sites flanking the second and the third exon were inserted by homologous recombination in mouse embryonic stem cells. Genotyping PCR resulted in the amplification of a specific band of 250bp in the $Sox30^{flox/flox}$ and 180bp in $Sox30^{+/+}$. The identity of these bands was confirmed by sequencing analysis. The $Sox^{flox/flox}$ females were viable and fertile with no gross abnormalities. However, $Sox^{flox/flox}$ males were found to be infertile. Thus, the floxed lines were maintained by mating heterozygous floxed males and females.

A1.3.2 Generation and validation of *Sox30* loxed allele

$Sox30^{flox/flox}$ females were mated to Tg. β- actin Cre mice. The resultant litters were heterozygous for deletion of critical exons of Sox30. The genotypes were confirmed, as shown in Figure 4b. The absence of band in PCR1 with primer pairs SKS 570-571 and the presence of a band at 673bp by PCR2 with primer pairs SKS573-571 confirms the deletion of exon 2 and exon 3 in Sox30. The genotype PCR band indicative of lox P-mediated recombination was sequence confirmed. The $Sox30^{lox/lox}$ females and $Sox30^{lox/+}$ male mice were viable, fertile, and with no gross abnormalities. However, $Sox30^{lox/lox}$ males were infertile.

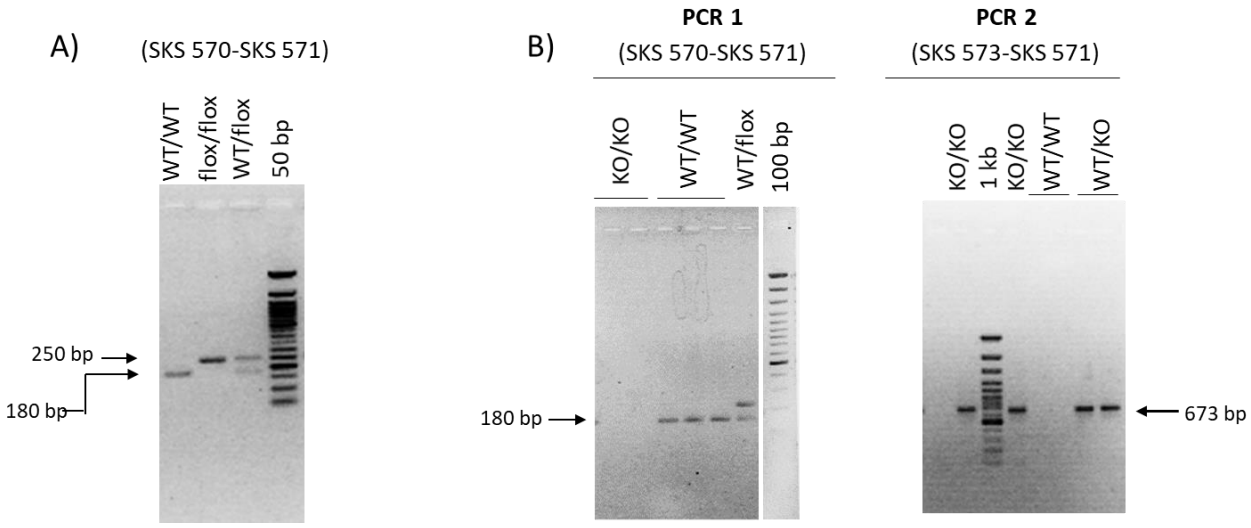


Figure 4: (A) Genotyping PCR for Sox30 floxed alleles confirming vector integration at the locus. A band of 250bp represents the mutant allele here. **(B)** Two-step PCR-based genotyping to validate excision of critical exons in the loxed alleles. The absence of a band in PCR1 and the presence of a band at 673bp by PCR2 confirms the deletion of exon 2 and exon 3 in Sox30. The same were also confirmed by sequencing

A1.3.3 Western analysis validates deletion of SOX30 in mouse testis

In the absence of a suitable commercial antibody against SOX30, we attempted to develop rabbit-raised polyclonal antibodies against mouse SOX30 protein. We selected the peptide epitope approach - a 20 amino acid stretch that is hydrophobic and is predicted to be antigenic was synthesized, purified, and injected with KLH-adjuvant into rabbits. The third immune sera were purified using the peptide affinity column. The details of the epitope chosen are as follows:

Selected Fragment: RDYPFRDYPDEHTHSEDSR (697-715)

Antigenicity Index: 0.972

Hydrophobicity Index: -2.432

Hydrophilicity Index: 1.005

The purified antibody staining was confirmed to overlap with GFP-tagged SOX30 using immunofluorescence (Figure 5).

SOX30 has a predicted molecular mass of 86kDa. We probed the testis lysates from Sox30 wildtype and Sox30 null mice. The antibody showed immuno-reactivity to the lysate from wildtype. However, no band was detected at comparable size in knockout tissue lysates. This confirms the specificity of the antibody for SOX30 in testis and validates the SOX30 loxed allele to be a null allele. However, in the brain tissue lysates, we detected a non-specific band at a higher molecular weight (closer to 100kDa) in both the wildtype and mutant tissues. We could not determine the identity of this band.

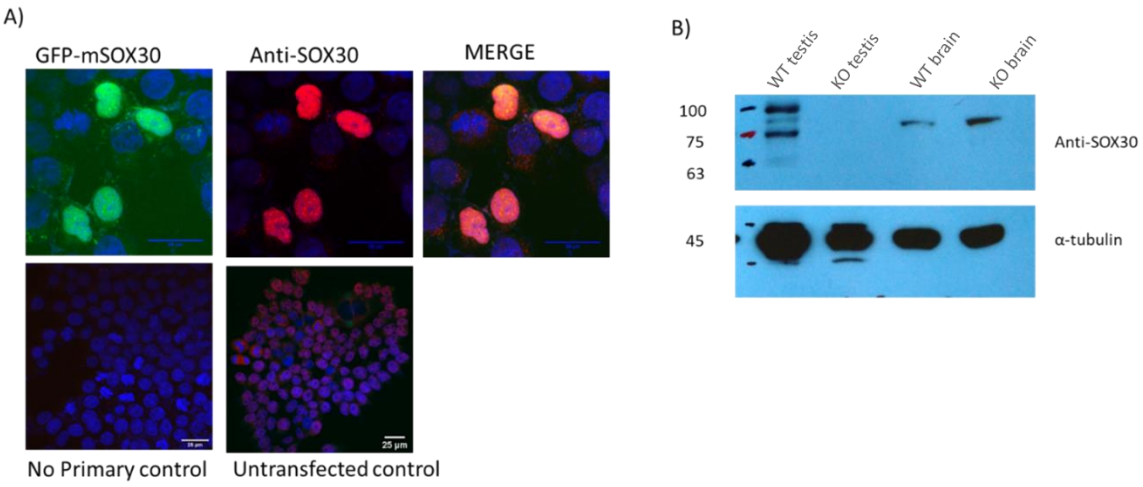


Figure 5: (A) Immunocytochemical analysis with GFP-tagged mouse SOX30 co-stained with anti-SOX30 antibody in red. No primary antibody controls and untransfected controls are shown in the bottom panel. **(B)** Western analysis of wildtype and Sox30 null testis and brain tissue lysates

A1.4 Discussion

Sox proteins are transcription factors with key roles in embryogenesis, development of testis and fertility, neurogenesis, central nervous system

development, oligodendrocyte development, chondrogenesis, neural crest cell development, among other respects (Jiang et al., 2013). In this study, our focus has been to generate resources to examine the possible roles of SOX30 in association with a brain disorder and male infertility. Interestingly, in our attempt to generate the SOX30 null mice carrying a deletion of exon 2 and exon 3 of the gene encoding the HMG domain, SOX30 homozygous mutant male mice were found to be viable but infertile. The female mice were viable and healthy.

Mouse models with gene disruption are a useful tool to understand gene functions. The knockout-first allele offers a flexible strategy to produce reporter knockouts, conditional knockouts, and null alleles following exposure to site-specific recombinases Cre and Flp. Considering several SOX proteins are essential for development, we used the knockout-first strategy to generate Sox30 mutant alleles. This allowed us to benefit from SOX30 being a reporter-tagged and a conditional mutation. Here, the cassette used consists of a mouse *En2* splice site, a β-galactosidase reporter, a neomycin resistance gene, and a transcriptional termination SV40 polyadenylation sequence, into the intronic sequence. The initial allele, the $Sox30^{Flox/Flox,}$ has the potential to generate a null allele on Cre-mediated recombination or conditional-ready (wildtype) allele on the removal of the cassette flanked by the FRT sites. In the $Sox30^{Flox/Flox}$ mice, transcription could be terminated by the polyadenylation sequence present in the gene-targeting vector. This generally happens before the entire gene sequence is transcribed, resulting in truncated transcription of a nonfunctional reporter protein. Alternatively, on Cre-mediated recombination, a Knockout-first can cause the deletion of critical exons and lead to loss of function allele – $Sox30^{Lox/Lox}$. We find that homozygous males with for the knockout-first and the null allele while viable are infertile. $Sox30^{Flox/Flox}$ and $Sox30^{Lox/Lox}$ homozygous females, heterozygous

animals of both sexes, are fertile. It is possible that the Sox30 mutants generated here are hypomorphs. But the DNA sequencing and Western analysis on $Sox30^{Lox/Lox}$ confirm that the allele is indeed a null allele. Also, the infertility phenotype of the SOX30 mutant mice generated here replicate the earlier findings by Feng and colleagues (C. A. Feng et al., 2017).

While the evidence of the association of Sox30 with infertility in humans is limited, the molecular underpinnings of the contribution of Sox30 to the phenotype in mice are beginning to be explored. In a mutational analysis for SOX30 among 38 Japanese individuals with human male infertility expressed as Sertoli cell-only syndrome, no rare pathogenic alleles were identified (Toshinobu Miyamoto, 2020). Among 494 male infertile patients of Indian ancestry, with majority of them with a clinical diagnosis of nonobstructive azoospermia, 15 rare heterozygous variants have been recently identified (Jaishankar, 2021). It is interesting to note that mice with Sox30 $^{+/-}$ are fertile, while in humans, variants identified so far are heterozygous. It would be interesting to examine if the expression of human SOX30 in the null mice would restore the phenotype in males. The mutant mice generated here shall help explore the effects of epilepsy and infertility mutations in the human *SOX30* gene. The wildtype and mutant human SOX30 could be expressed in the adult mouse testis facilitating in vivo gene function analysis. This could be achieved using microinjection and electroporation of mouse testis (Michaelis et al., 2014). Thus, the mutant mice generated here offer a robust system to tease out the roles of SOX30.

The expression of SOX30 in the mouse brain is debated with no conclusive evidence to date. Sox30 was initially reported to be expressed in the mouse embryonic tissues at E13.5 using RT-PCR SOX30 transcript was

seen in lungs, brain, testis, ovary, heart, etc. (Osaki et al., 1999). However, in a recent report using qRT-PCR, expression of SOX30 was detected only in the germ cells of gonads and postnatally only in the testis (Feng et al., 2017). In our preliminary immune-histochemical analysis of mouse embryonic brain with anti-SOX30 antibody, we detected signals in the cortex, thalamus. In these regions, nuclear staining was observed (Appendix, Figure 1). While these signals could correspond to SOX30, we need further validation of this observation considering the observation of a possibly non-specific band detected in brain lysates in Western analysis. Experiments using RNA *in situ* hybridization and staining for LacZ in $Sox30^{Flox/Flox}$ mice shall be informative of *Sox30* expression across various mouse tissues.

Appendix

Pairwise sequence alignment of human- and mouse- SOX30.

```
Aligned_sequences: 2
# 1: mSox30
# 2: hSOX30
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 782
# Identity:     595/782 (76.1%)
# Similarity:   633/782 (80.9%)
# Gaps:          29/782 ( 3.7%)
# Score: 3047.0#
#
#=======================================

mSox30             1 MERARPEPPPPPPPPRQPPRPTPPRPLRPAPPAQPVEAATFRAAAAERS     50
                     |||||||||             .|||||||||..|||..:|.|||.|..
hSOX30             1 MERARPEPPP-------------QPRPLRPAPPPLPVEGTSFWAAAMEPP     37

mSox30            51 QSPSAQATAAMAAVASSCGEAAAAGAQAAGTRRLLQVKPEQVLLLPPGGP    100
                     .|....:.||.|.:|||||||.|:|.|.| .|||||||||||||.
hSOX30            38 PSSPTLSAAASATLASSCGEAVASGLQPA-VRRLLQVKPEQVLLLPQ---     83

mSox30           101 GVPPAPDEGAAAAAAAAAAAASSAQARLLQLRPELLLLPPQSAADGGPCR    150
                       |.|.:|         .||||||||||||.||:|.||.|.:|:||...|
hSOX30            84 --PQAQNE---------EAAASSAQARLLQFRPDLRLLQPPTASDGATSR    122

mSox30           151 PELHPMQPRTLLVKAEKQELGAGLDLSVGSRRTTEAGPRASRAAKLDGTG    200
                     |||||:||..|.|||:||:||..||.|||.|...|.||||||..||:|.|
hSOX30           123 PELHPVQPLALHVKAKKQKLGPSLDQSVGPRGAVETGPRASRVVKLEGPG    172

mSox30           201 KALDGRRSDEKKAKLEAEEAPRDALKGGEGKSLLAIGEGVIKTEEPDRPR    250
                     .||...|.|| |.|||||..||:::||.|||..||.|||||||||:|..
hSOX30           173 PALGYFRGDE-KGKLEAEEVMRDSMQGGAGKSPAAIREGVIKTEEPERLL    221

mSox30           251 DDCRLGTEATSNGLVHSSKEAILAQPPSAFGPHQQDLRFPLTLHTVPPGA    300
                     :|||||.|..|||||.|.|.|||....||||||||||.||||||||||
hSOX30           222 EDCRLGAEPASNGLVHGSAEVILAPTSGAFGPHQQDLRIPLTLHTVPPGA    271

mSox30           301 RIQFQGPPPSELIRLSKVPLTPVPIKMQSLLEPSVKIETKDVPLTVLPSD    350
                     ||||||.||||||||:||||||||.|||||||||||||||||||||||||
hSOX30           272 RIQFQGAPPSELIRLTKVPLTPVPTKMQSLLEPSVKIETKDVPLTVLPSD    321

mSox30           351 AGIPDTPFSKDRNGHVKRPMNAFMVWARIHRPALAKANPAANNAEISVQL    400
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
hSOX30           322 AGIPDTPFSKDRNGHVKRPMNAFMVWARIHRPALAKANPAANNAEISVQL    371

mSox30           401 GLEWNKLSEEQKKPYYDEAQKIKEKHREEFPGWVYQPRPGKRKRFPLSVS    450
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
hSOX30           372 GLEWNKLSEEQKKPYYDEAQKIKEKHREEFPGWVYQPRPGKRKRFPLSVS    421

mSox30           451 NVFSGTTQNIISTNPTTIYPYRSPTYSVVIPGLQNTITHPVGEAPPAIQL    500
                     ||||||||||||||||:|||||||||||||.|||.|||||||..|||||
hSOX30           422 NVFSGTTQNIISTNPTTVYPYRSPTYSVVIPSLQNPITHPVGETSPAIQL    471
```

```
mSox30          501 PTPAVQRPSPITLFQPSVSSTGPVAVPPPSLTPRPSLPPQRFSGPSQTDI    550
                    ||||||.|||:|||||||||...|||..|||...|:||||||:||||||.
hSOX30          472 PTPAVQSPSPVTLFQPSVSSAAQVAVQDPSLPVYPALPPQRFTGPSQTDT    521

mSox30          551 HRLPSGSSRSVKRSTPGSLESTTRIPAGASTAHARFATSPIQPPKEYASV    600
                    |:|.|.::.:||:.||.|||||..||.:.||||||||||.||||:||:||
hSOX30          522 HQLHSEATHTVKQPTPVSLESANRISSSASTAHARFATSTIQPPREYSSV    571

mSox30          601 STCPRSTPIPPATPIPHSHVYQPPPLGHPATLFGTPPRFSFHHPYFLPGP    650
                    |.||||.|||.|:||||.||||||||||||||||||||||||||||||||
hSOX30          572 SPCPRSAPIPQASPIPHPHVYQPPPLGHPATLFGTPPRFSFHHPYFLPGP    621

mSox30          651 HYFPSSTCPYSRPPFGYGNFPSSMPECLGYYEDRYQKHEAIFSALNRDYP    700
                    |||||||||||||||||||||||||||.||||||.|||.|||.||||||.
hSOX30          622 HYFPSSTCPYSRPPFGYGNFPSSMPECLSYYEDRYPKHEGIFSTLNRDYS    671

mSox30          701 FRDYPDEHTHSEDSRSCESMDGPPYYSSHGHGGEEYLNAMPTLDIGALEN    750
                    ||||..|.||||:|||||:|:|..||:||.|.|||.||.:|.||||.|||
hSOX30          672 FRDYSSECTHSENSRSCENMNGTSYYNSHSHSGEENLNPVPQLDIGTLEN    721

mSox30          751 VFTAPASAPSGVQQVNVTDSDEEEEEKVLRNL    782
                    |||||.|.||.:||||||||||||||||:|
hSOX30          722 VFTAPTSTPSSIQQVNVTDSDEEEEEKVLRDL    753
```

Table 1: Primers pairs used for genotyping.

Primer	Sequence 5'-3'	Length
SKS-570	CAGTTGGTGAGTTGGTCGTA	20
SKS-571	CCACATGGATCTGCTATTGC	20
SKS-573	GTACGTCTTCCCGAGCGAAAAC	22

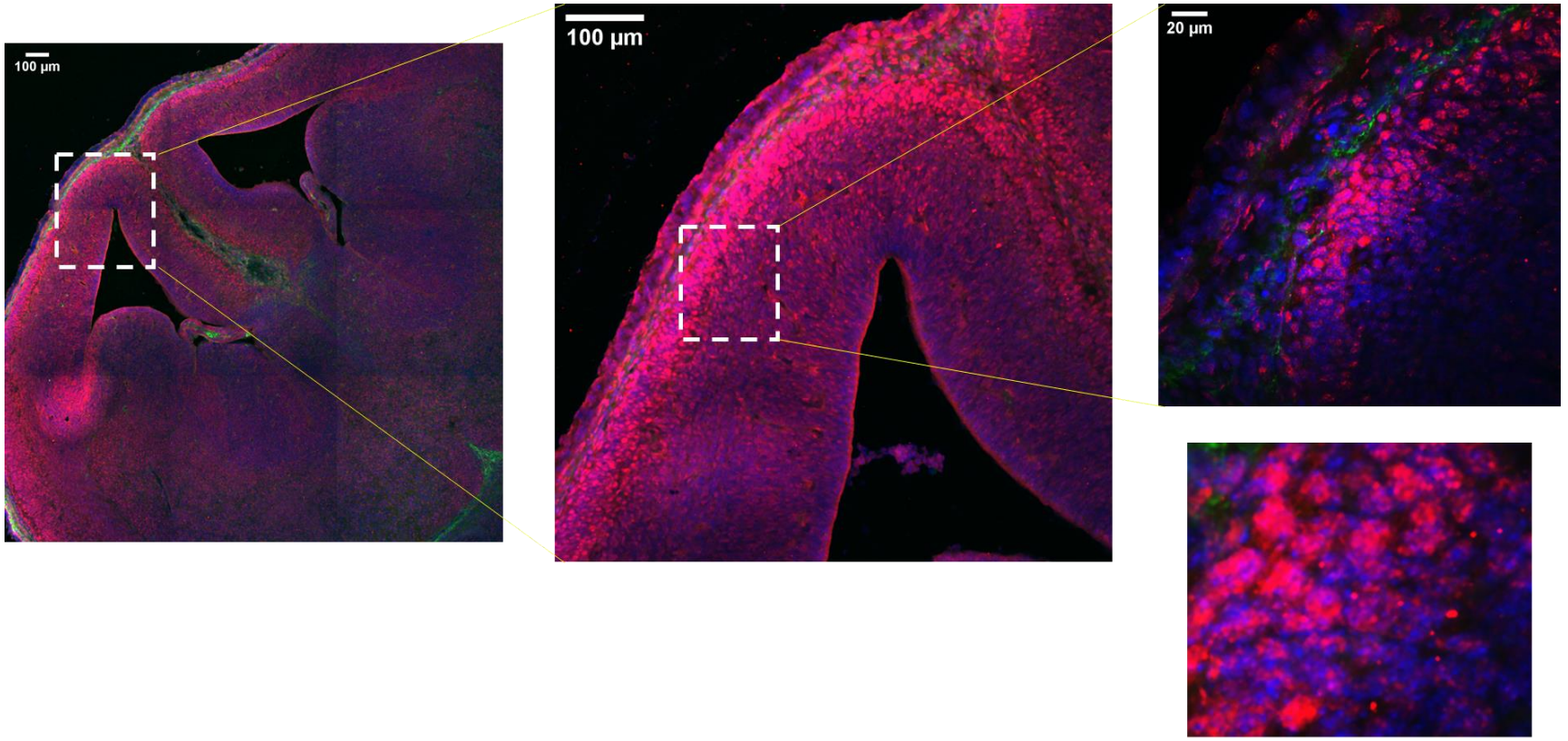


Figure 1: Mouse embryonic brain E12.5 stained with anti-SOX30 antibody. 25µm sections were stained with anti-Sox30 antibody in red at 1:500 dilution. Nuclear stain DAPI is blue. Zoomed in panels show antibody staining in nuclei but not in the nucleoli

Table 2: cDNA and cloning primers for Sox30

Primer	Sequence	Tm	Length
Sox30-cDNA-EcoRI-F	5'-CGGAATTCAATGGAGAGAGCC AGGCCG-3'	60.8	27
mSox30-cDNA-1R	5'- CTCTTGCTTCTCGGCCTTGA -3'	60.5	20
mSox30-cDNA-2F	5'-CTGCTGTTGCCTCCGCAGT-3'	61.6	19
mSox30-cDNA-2R	5'- CAAGGACTGCATTTTAATAGG -3'	55.4	21
mSox30-cDNA-3F	5'- ACTCCACACCGTCCCACCT-3'	61.6	19
mSox30-cDNA-3R	5'- GAGTAGGTAGGTGAGCGATAA-3'	59.5	21
mSox30-cDNA-4F	5'- ATCAGCCTCGTCCAGGGAA-3'	59.5	19
mSox30-cDNA-4R	5'- TGATAGACGTGTGAGTGTGGA-3'	59.5	21
mSox30-cDNA-5F	5'- CAGATTTGCAACCTCGCCTAT-3'	59.5	21
mSox30-cDNA-BamHI-R	5'- AAAGTGCTCAGGAATTTATAG GGATCCCG-3'	53.4	29

Appendix for chapter 2

Appendix 2

Table 1: Primer pairs used for PCR and sequencing of *EFHC2*.

Exon	Forward primer (5'-3')	Length	Reverse primer (5'-3')	Length	Product size(bp)
1	ttcccgtctccaggcaac	18	ggtccaagttgccaccaa	18	281
2	gccactttatttgcattttgtgag	24	ggcaaaaagaaaaagaactagcc	23	436
3	cgcatggctgatgtgtagaa	20	cagcaggttagacagcctaaa	21	418
4	tgctgcttcctagaacgagtc	21	gcccaaagaagaacaaaactgg	22	498
5	gctggtcttgaactcctgaa	20	gccttaacttcaatgttccaac	22	650
6	tgacatttacagtgtgccaaca	22	gacatttcagcatgtttattactg	24	434
7	cagtaataaacatgctgaaatgtc	24	caaacggctggaagatagca	20	528
8	tcacaggattttgctgttttg	21	tctccctggacttgcaatta	20	436
9	ggttgaactctttcactttgg	21	gaaccctattgttcagggttt	21	423
10	cactgcctcattcttccctt	20	cccacatgtgtgaatgcac	19	482
11	agtcatgccttctcggtgtt	20	ttcgtggccaaaactttacc	20	440
12	ttcctgctgctcccaaatta	20	tcactgggaaatatctgtacc	21	557
13	gaaatgtgactgttcttcctt	21	gaggcaggtcaccaataata	20	395
14	gagggtcagaagcaataagaa	21	aatgcccagaaacacataagca	22	401
15	ggagggaatgaagaatcaaga	21	acgtcggcaatgtaacatga	20	452
16	tggaaatccataaaccacaa	20	tgcatatgtttccagtgtgt	20	484
17	cacactggaaacatatgcag	20	gtagtccctgaaaggtaaatg	21	486

Table 2: Primer pairs for EFHC2 cDNA analysis

Sl. No	Forward primer (5'-3')	Length	Reverse primer (5'-3')	Length	Product size
1	ggcaacagcttcaaccgcaa	20	tcaggaggcggaagagtaat	20	404
2	ccctgaagatgacacaattca	21	ctgagctgtgtggaagcaat	20	491
3	gaactcatcctgcattacttc	21	ggtgtgggcttgaggtctat	20	500
4	cttacaacggttttggttctga	22	gtttccaacagtcaaagacatc	22	605
5	agagagctcaagcaggtattt	21	cgtccttcaaaaaggtccag	20	547

Table 3: EFHC2 site-directed mutagenesis primers

Name	**Sequence (5'-3')**	**Length**
EFHC2-G8X-F	cctgcctctgctgccgtagaacagcttcaaccgca	35
EFHC2-G8X-R	tgcggttgaagctgttctacggcagcagaggcagg	35
EFHC2-Ser70Ter-F	aaggagatggaagtgatgtaccataatgggtagccttt	38
EFHC2-Ser70Ter-R	aaaggctacccattatggtacatcacttccatctcctt	38
EFHC2-Cys191Ter-F	gtgaatcccccagtgcaatgaccagaagatccttaca	37
EFHC2-Cys191Ter-R	tgtaaggatcttctggtcattgcactgggggattcac	37
EFHC2-Arg200Ter-F	cagaagatccttacatgaagatttagagagaggttgtagaacacgt	46
EFHC2-Arg200Ter-R	acgtgttctacaacctctctctaaatcttcatgtaaggatcttctg	46
EFHC2-Arg294Ter-F	ttgcccacctagagtctattaaccaggccagataa	35
EFHC2-Arg294Ter-R	ttatctggcctggttaatagactctaggtgggcaa	35
EFHC2-Ser377Ter-F	gagaactttacctcagtttgatgcaagcctccttctc	37
EFHC2-Ser377Ter-R	gagaaggaggcttgcatcaaactgaggtaaagttctc	37
EFHC2-Lys547Ter-F	gtatcctttcagtaacctctaacttgccctacaaaagct	39
EFHC2-Lys547Ter-R	Agcttttgtagggcaagttagaggttactgaaaggatac	39

Appendix for chapter 3

Appendix 2

Pairwise sequence alignment of human and mouse EFHC2 proteins

```
# Aligned_sequences: 2
# 1: Human
# 2: Mouse
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 751
# Identity:     556/751 (74.0%)
# Similarity:   642/751 (85.5%)
# Gaps:           3/751 ( 0.4%)
# Score: 3082.0
#
#
#=======================================

Human               1 MALPLLPGNSFNRNVGKEKFHKSQHWGFCNNVMMLVSDEKPGIGGEPLLG     50
                      ||||.|||||||||:|||:|||||||||||||.||||:.|||:||:.|.|
Mouse               1 MALPFLPGNSFNRNIGKERFHKSQHWGFCNNVRMLVSENKPGVGGDLLYG     50

Human              51 QKIKPKCSIYPKGDGSDVPSWVAFDKQVLSFDAYLEEEVLDKSQTNYRIR    100
                      ||||||.|::|||||:|.||||||||||||||||:|:.||.|..:|||
Mouse              51 QKIKPKHSVFPKGDGTDAPSWVAFDKQVLSFDAYLEDEISDKRQEIFRIR    100

Human             101 YYKIYFYPEDDTIQVNEPEVKNSGLLQGTSIRRHRITLPPPDEDQFYTVY    150
                      |||||||.||||||||||||.||||.|||||||.||..|||::|||||||
Mouse             101 YYKIYFYLEDDTIQVNEPEVINSGLPQGTSIRRQRIPYPPPNDDQFYTVY    150

Human             151 HFNVGTEVVFYGRTFKIYDCDAFTRNFLRKIGVKVNPPVQCPEDPYMKIR    200
                      .||:...|||||||||||||.||:|||:|||:|:|||.|||.|||||:|
Mouse             151 DFNINISVVFYGRTFKIYDCDPFTKNFLKKIGIKLNPPGQCPLDPYMKMR    200

Human             201 REVVEHVEPLRPYESLDTLKQFLQYHGKILCFFCLWDDSVSMFGDRRELI    250
                      ||.:|.|:|.|||:|.||||:|:||.||:|.||||||||.|:||||||.:
Mouse             201 RETLEFVDPFRPYQSFDTLKRFIQYDGKVLRFFCLWDDSTSLFGDRREFV    250

Human             251 LHYFLCDDTIEIKELLPHSSGRDALKMFLRRSKLPKNCPPRVYQPGQITD    300
                      |||||||.|:||:|:||.:|||||:..||||.|||||..||.:||||||||
Mouse             251 LHYFLCDGTVEIREVLPSNSGRDAMSSFLRRGKLPKYGPPGIYQPGQITD    300

Human             301 RAVLNSYGDFIKNQADGYLFDRYKLGKVDQEFYKDSDLSLGVTINVWGRK    350
                      |||||.||...:.:||||||.|:|:||||:|:||.|.|||:|.||||||||
Mouse             301 RAVLNVYGGLSEWRADGYLLDKYQLGKVEQDFYTDQDLSIGATINVWGRK    350

Human             351 VLLYDCDEFTKSYYKSKYGIENFTSVSCKPPSPPPKIERKFPPYNGFGSE    400
                      |||.|||||||:||::|||::|||.:|||||. .||||||:|||.|||||
Mouse             351 VLLCDCDEFTKTYYRTKYGVDNFTPISCKPPH-LPKIERKYPPYTGFGSE    399

Human             401 EDSLRNCIDLKPTPHRRNFKKFMEKDSYGSKSNILRFFAKLVTDKCVDLD    450
                      |||.|:|:.|||||||:||||||.||:|:.|||||:|.||:|.||.|:|
Mouse             400 EDSFRSCVGLKPTPHRKNFKKFMELDSFGNISNILRYFGKLITHKCADVD    449
```

```
Human           451 RMFVISYYLGDDTISVFEPIERNSGIAGGMFLKRSRVKKPGQEVFKSELS    500
                    |:|||::||.||||||||||.|||.|||||||||||||||||||||.|
Mouse           450 RIFVIAFYLSDDTISVFEPIENNSGNAGGMFLKRSRVKKPGQEVFKSEFS    499

Human           501 EYIKAEELYIGVTVNVNGYLFRLLNADEYTLNYMEQNTDKYPFSNLKLAL    550
                    ||||||||||.|||:|||||.||||||||||||.||||:|:||.:||:
Mouse           500 EYIKAEELYIGATVNINGYLFILLNADEYTLNYMENNTDKFPYSNFELAI    549

Human           551 QKLKQEEGKSRELKQVFKAADSKHTNMVDYNTFRDILMSLTVGNLAEQEF    600
                    ||||||:.||||:.|||.|||..||.:|.||||||||||:|:|.|.:||.
Mouse           550 QKLKQEKSKSREITQVFAAADYNHTKVVPYNTFRDILMSITMGKLIDQEL    599

Human           601 VTIARHYRVPEGTCSDMDFLIALAHEKFKKNMFENFDTFIYSCVYEDREK    650
                    :||||||||||....|:.:|||.||||||||:|||||.|||:||||||||
Mouse           600 ITIARHYRVPEIMDPDLAYLIARAHEKFKKNIFENFDMFIYNCVYEDREK    649

Human           651 KNVLPTKDIKRLCKSSRLPLSDDLLESLLSRFEDSEKQIDYKSFFSALNW    700
                    |.|||||||:|:||||||||.||.|:.|||||||.:.||:|:.|||.|||
Mouse           650 KGVLPTKDIRRMCKSSRLPLDDDFLDCLLSRFEDKDHQINYEIFFSVLNW    699

Human           701 RKNPVPELQPASYLKERCEDVWLGMPSPIPAKYIDYWTFLKDAFGLEEE-    749
                    |.||.|:||...||||:||||:|||||||.||:.|..||.|.:|||:.
Mouse           700 RMNPTPDLQAPPYLKEKCEDVWVGMPSPIPVKYVRYLDFLIDVYGLEDNM    749

Human           750 -    749

Mouse           750 L    750
```

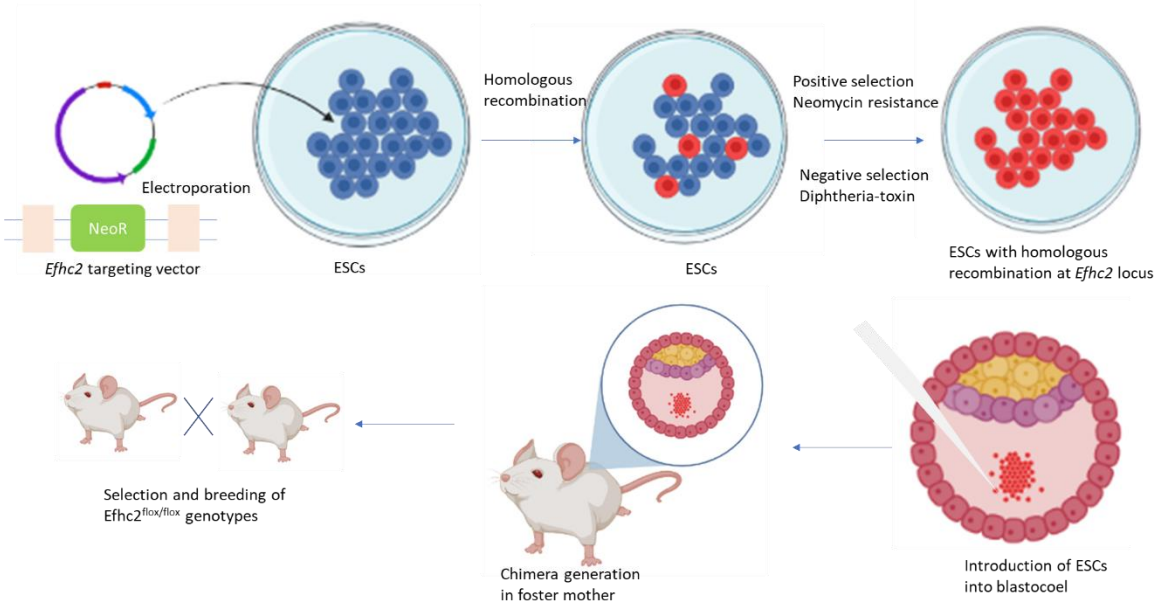


Figure 1: Strategy used for the generation of Efhc2 floxed allele. Illustration created using BioRender.

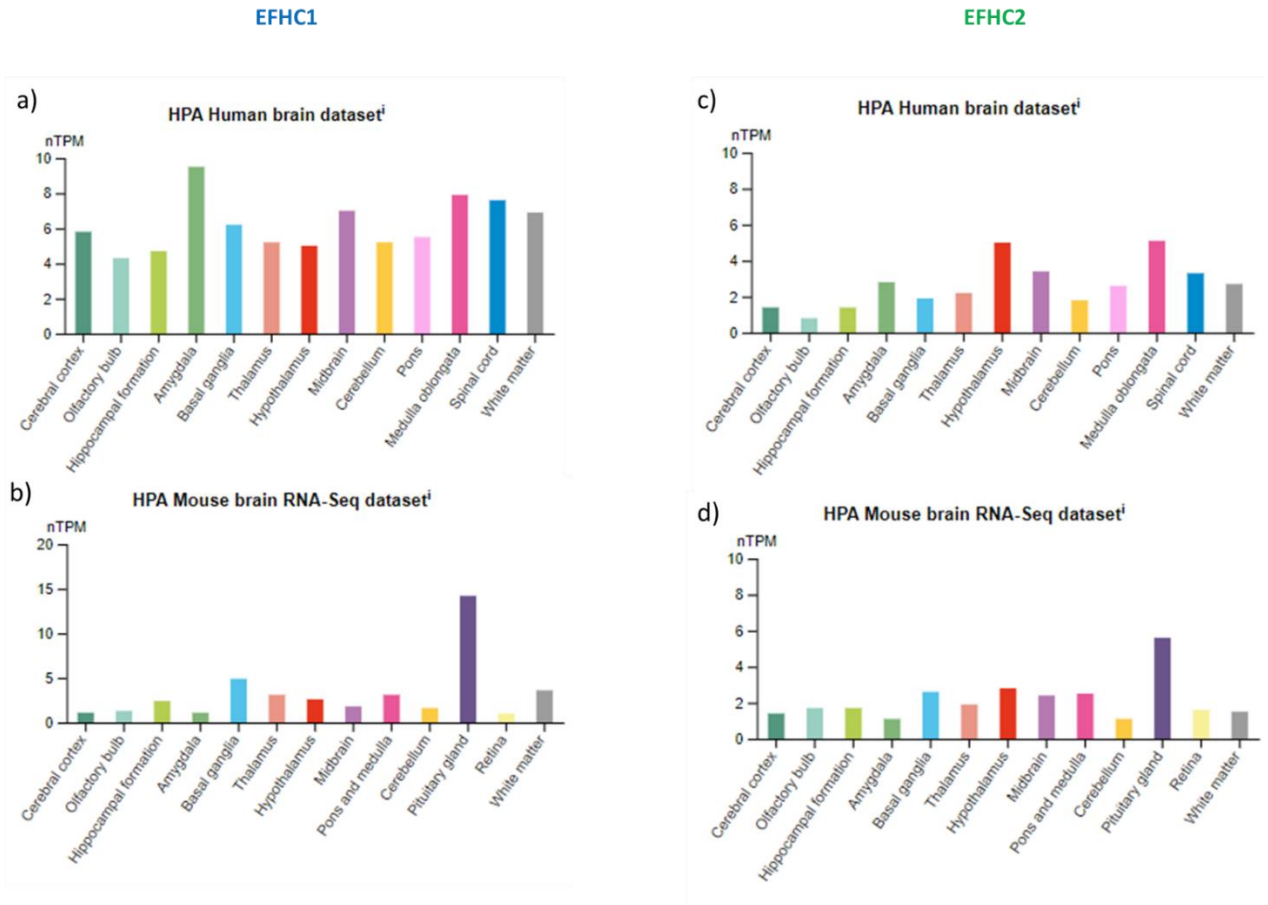


Figure 2: Expression of EFHC1 and EFHC2 protein across regions of the human and mouse brain. EFHC1 in **(a)** human brain **(b)** mouse brain; EFHC2 in **(c)** human brain **(d)** mouse brain. Image credit: Human Protein Atlas, version 21.0; Information for EFHC2 sourced from https://www.proteinatlas.org/ENSG00000183690-EFHC2/brain and EFHC1 from https://www.proteinatlas.org/ENSG00000096093-EFHC1/brain

2 Pair-wise sequence alignment of EFHC2 and EFHC2-En2 insertion

```
# Aligned_sequences: 2
# 1: EFHC2-En2
# 2: EFHC2
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 758
# Identity:     708/758 (93.4%)
# Similarity:   717/758 (94.6%)
# Gaps:          26/758 ( 3.4%)
# Score: 3774.5
#
#
#=======================================

EFHC2-En2          1 MALPFLPGNSFNRNIGKERFHKSQHWGFCNNVRMLVSENKPGVGGDLLYG     50
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EFHC2              1 MALPFLPGNSFNRNIGKERFHKSQHWGFCNNVRMLVSENKPGVGGDLLYG     50

EFHC2-En2         51 QKIKPKHSVFPKGDGTDAPSWVAFDKQVPGPE-------NQRRRT-----     88
                     ||||||||||||||||||||||||||||...:       :.:|:.
EFHC2             51 QKIKPKHSVFPKGDGTDAPSWVAFDKQVLSFDAYLEDEISDKRQEIFRIR    100

EFHC2-En2         89 -------LTKRTSGLAQPSLLSSSRGSRLSFRPTGTSIRRQRIPYPPPND    131
                             |...|..:.:|.:::|.       .|.|||||||||||||||
EFHC2            101 YYKIYFYLEDDTIQVNEPEVINSG-------LPQGTSIRRQRIPYPPPND    143

EFHC2-En2        132 DQFYTVYDFNINISVVFYGRTFKIYDCDPFTKNFLKKIGIKLNPPGQCPL    181
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EFHC2            144 DQFYTVYDFNINISVVFYGRTFKIYDCDPFTKNFLKKIGIKLNPPGQCPL    193

EFHC2-En2        182 DPYMKMRRETLEFVDPFRPYQSFDTLKRFIQYDGKVLRFFCLWDDSTSLF    231
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EFHC2            194 DPYMKMRRETLEFVDPFRPYQSFDTLKRFIQYDGKVLRFFCLWDDSTSLF    243

EFHC2-En2        232 GDRREFVLHYFLCDGTVEIREVLPSNSGRDAMSSFLRRGKLPKYGPPGIY    281
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EFHC2            244 GDRREFVLHYFLCDGTVEIREVLPSNSGRDAMSSFLRRGKLPKYGPPGIY    293

EFHC2-En2        282 QPGQITDRAVLNVYGGLSEWRADGYLLDKYQLGKVEQDFYTDQDLSIGAT    331
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EFHC2            294 QPGQITDRAVLNVYGGLSEWRADGYLLDKYQLGKVEQDFYTDQDLSIGAT    343

EFHC2-En2        332 INVWGRKVLLCDCDEFTKTYYRTKYGVDNFTPISCKPPHLPKIERKYPPY    381
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EFHC2            344 INVWGRKVLLCDCDEFTKTYYRTKYGVDNFTPISCKPPHLPKIERKYPPY    393

EFHC2-En2        382 TGFGSEEDSFRSCVGLKPTPHRKNFKKFMELDSFGNISNILRYFGKLITH    431
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EFHC2            394 TGFGSEEDSFRSCVGLKPTPHRKNFKKFMELDSFGNISNILRYFGKLITH    443

EFHC2-En2        432 KCADVDRIFVIAFYLSDDTISVFEPIENNSGNAGGMFLKRSRVKKPGQEV    481
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EFHC2            444 KCADVDRIFVIAFYLSDDTISVFEPIENNSGNAGGMFLKRSRVKKPGQEV    493

EFHC2-En2        482 FKSEFSEYIKAEELYIGATVNINGYLFILLNADEYTLNYMENNTDKFPYS    531
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
EFHC2           494 FKSEFSEYIKAEELYIGATVNINGYLFILLNADEYTLNYMENNTDKFPYS    543

EFHC2-En2       532 NFELAIQKLKQEKSKSREITQVFAAADYNHTKVVPYNTFRDILMSITMGK    581
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
EFHC2           544 NFELAIQKLKQEKSKSREITQVFAAADYNHTKVVPYNTFRDILMSITMGK    593

EFHC2-En2       582 LIDQELITIARHYRVPEIMDPDLAYLIARAHEKFKKNIFENFDMFIYNCV    631
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
EFHC2           594 LIDQELITIARHYRVPEIMDPDLAYLIARAHEKFKKNIFENFDMFIYNCV    643

EFHC2-En2       632 YEDREKKGVLPTKDIRRMCKSSRLPLDDDFLDCLLSRFEDKDHQINYEIF    681
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
EFHC2           644 YEDREKKGVLPTKDIRRMCKSSRLPLDDDFLDCLLSRFEDKDHQINYEIF    693

EFHC2-En2       682 FSVLNWRMNPTPDLQAPPYLKEKCEDVWVGMPSPIPVKYVRYLDFLIDVY    731
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
EFHC2           694 FSVLNWRMNPTPDLQAPPYLKEKCEDVWVGMPSPIPVKYVRYLDFLIDVY    743

EFHC2-En2       732 GLEDNML*    739
                    |||||||
EFHC2           744 GLEDNML*    751
```

Table 1 – Primer sequences used for genotyping and sequencing

PRIMER NAME	SEQUENCE (5'-3')	LENGTH
SKS 574 -F	TGGACTGCCTCAAGGTATCT	20
SKS 575 -R	ATCCTGGATGATGGTACTGG	20
SKS 573 -F	GTACGTCTTCCCGAGCGAAAAC	22
FLOX-1F	GGATGCCTCTGGCTGTACTT	20
FLOX-V-1R	TGACCTTGGGCAAGAACATA	20
FLOX-V-2F	CCCCAAACACTGCCAACTAT	20
FLOX-V-2R	ATGTGCCCACTGACCAGAA	19
FLOX-V-3F	GCATTCCTTCCGCTAGCTTT	20
FLOX-V-3R	AAAGACGGCAATATGGTGGA	20
FLOX-V-4F	GGTACCCGGAAGATCTGGAC	20
FLOX-V-4R	GTTTTCCCAGTCACGACGTT	20
FLOX-V-5F	TCGGTGCACATGCTTTACAT	20

FLOX-V-5R	CGGTTTATGCAGCAACGAG	19
FLOX-V-6F	CGTCAGTATCGGCGGAAT	18
FLOX-V-6R	AGGAGGAGGGAGAGTTTTGG	20
FLOX-V-7F	AGTGCCCAAGAGATGTCCAC	20
FLOX-V-7R	ATACTTTCTCGGCAGGAGCA	20
FLOX-V-8F	CCCCCGGATCTAAGCTCT	18
FLOX-V-8F1	CGTTGGCTACCCGTGATATT	20
FLOX-V-8R	ACTTTCACCGCAGTCTGGTT	20
FLOX-V-9F	TGAGATTCCAAAGATGGACAAA	22
FLOX-V-10F	GCATCTCAGTTATAGGCAGCAG	22
FLOX-V-10R	CGGAAATCCATCAGCATAATC	21
FLOX-LACZ-F	GACGTCTCGTTGCTGCATAA	20
FLOX-LACZ-R	GATACAGCGCGTCGTGATTA	20
FLOX-EN2-F	GGTCCCGAAAACCAAAGAAG	20
FLOX-V-10F	GCATCTCAGTTATAGGCAGCAG	22
FLP-IC-0042	CTAGGCCACAGAATTGAAAGATCT	24
FLP-IC-0043	GTAGGTGGAAATTCTAGCATCATCC	25
FLP-TG--1348	CACTGATATTGTAAGTAGTTTGC	23
FLP-TG--1349	CTAGTGCGAAGTAGTGATCAGG	22
SKS573	GTACGTCTTCCCGAGCGAAAAC	22
B-CRE-F	CCATCTGCCACCAGCCAG	18
B-CRE-R	TCGCCATCTTCCAGCAGG	18

Table 2 – Primers used for cDNA cloning

PRIMER NAME	SEQUENCE (5'-3')	LENGTH
EFHC2-NOTI-FP	ATAAGAATGCGGCCGCAATGGCTCTACCGTTTCTG CCTGGC	41
EFHC2-ECORI-RP	GGGCGAATTCTCATAGCATGTTGTCCTCTAAGCC	34
EFHC2-XHOI-FP	CCGACTCGAGCCATGGCTCTACCGTTTCTGCC	32
MEFHC2_OVERLAP-A	TGATAGAAGTGCCTGCTTATCAAAAGCAAC	30
MEFHC2_OVERLAP-B	TTGCTTTTGATAAGCAG GCACTTCTATCAGGCGTC	35
MEFHC2-ISH-KPNI-RP	GGGGTACCCTCTCTCCGCATCTTCATGT	28
MPAX6_ISH_FP	GGGCGAATTCAGTCCCCAGTCAGACCTC	28
MPAX6_ISH_RP	GGGGTACCTCACACGCTCTCTCCTTC	26

Table 3 – Primers used for site-directed mutagenesis

PRIMER NAME	SEQUENCE (5'-3')	LENGTH
mEFHC2-SDM-F74X-FP	GCACCATCCTGGGTTGCTTAGGATAAGCAGGTG TTATCTT	40
mEFHC2-SDM-F74X-RP	AAGATAACACCTGCTTATCCTAAGCAACCCAGGA TGGTGC	40
mEFHC2-SDM-G189X-FP	ATCAAATTGAACCCTCCATGACAATGTCCACTCG ATC	37
mEFHC2-SDM-G189X-RP	GATCGAGTGGACATTGTCATGGAGGGTTCAATTT GAT	37
mEFHC2-SDM-T374X-FP	TATTATAGGACAAAGTATGGAGTTGATAACTTTTA GCCCATTTCATGCAAGC	52
mEFHC2-SDM-T374X-RP	GCTTGCATGAAATGGGCTAAAAGTTATCAACTCC ATACTTTGTCCTATAATA	52

mEFHC2-SDM-E546X-FP	GACAAGTTTCCATATAGCAACTTTTAACTTGCCATACAAAAACTGAAG	48
mEFHC2-DM10-2,3-XhoI-FP	CTTCAGTTTTTGTATGGCAAGTTAAAAGTTGCTATATGGAAACTTGTC	48
mEFHC2-CTER-BglII-FP	CCGCTCGAGCAATGCGGAGAGAGACATTAGAGT	33
mEFHC2-DM10-3,EF-BglII-FP	ACTCTAAGATCTATGGCCATACAAAAACTGAAG	33

Table 4 – Primers used for cDNA amplification and sequencing

Primer Name	Sequence (5'-3')	Length
mEfhc2-e4-2F	ATTGAACCCTCCAGGACAA	20
mEfhc2-e8-2R	GCTTCAGGCCTACACAGGAG	20
mEfhc2-e5-5R	GGTTGAATCATCCCACAAGC	19
mEfhc2-e8-3F	ACTTTACGCCCATTTCATGC	20
mEfhc2-e12-3R	GCAATGGTTATAAGCTCCTGGT	20
mEfhc22-e10-4F	GCTACCTGTTTATTCTGCTTAATG	22
mEfhc2-cdna-E2-RP	CTGGGTTGCTTTTGATAAGCAG	22

Table 5 – Primers used validation of sites of integration of the gene targeting vector at the genomic locus

Primer Name	Sequence (5'-3')	Length
mefhc2_ 5 arm conf-FP	CCCCCATTCTACTCTTTTTCC	21
mefhc2_ 5 arm conf-RP	TCATGGAGGGTTTGGGTTTA	21
mefhc2_ 3 arm conf-FP	TGCCTGTCGTCTTTATCAATG	21

mefhc2_ 3 arm conf-RP	GCTGAGAATCAAAAAGCAGGA	21

Table 6: Dosage and phenotypes on intra peritoneal injection of Pentylenetetrazol

Dosage mg/kg	**Response**
12.5	Mild tail stiffening that lasted less than 10 seconds.
25	Head jerks, increased digging
	GTCS, tail stiffening, hind limb jerks, loss of balance
50	Severe GTCS, MJ, loss of balance, eventually leading to respiratory arrest

Appendix for chapter 4

Appendix 3

Table 1 – *CILK1* gene sequencing primers

Exon	Forward primer (5'-3')	Reverse primer (5'-3')	Product size(bp)
5'UTR-A	gaacaggacgcagtgcag	aggatcgcggaaagaaacta	492
5'UTR-B	acaaggggtctgagcaaaaa	agcacaccctgaaactccac	480
Exon-1	ctgtactttgggctttgcag	tcatctaacccacagcaatca	595
Exon-2	caccactgctactggcaaaa	tatcatttgcctccctttcc	394
Exon-3	ggcagccagaaataccagtt	gaaaagaaaagcacaccagca	595
Exon-4	gctatgttggaggaaaaact	tccattgtcagcctctca	400
Exon-5	aagcctttggtgccatgc	cccagtgtctcaaagttttcac	600
Exon-6	gcagttgccaagttctcaca	gtgggcagtggagaatgact	467
Exon-7	gggtgggtaatgacatggtt	catgccaattttcaaagcaa	497
Exon-8	aaggggtgttaaaaattcaagg	ccatgctgttgcaaacataaa	599
Exon-9	gttgccccttcagataagca	ggcttcaggtccaaaacact	473
Exon-10	ccctcagcaggattgacct	tctatgatgttcccctcagttc	461
Exon-11	tctatgatgttcccctcagttc	gccgagaaagagaggagga	431
Exon-12	ggatgcttggctctgaactc	tccctacctcctgtctgctg	497
Exon-13	ttttctcatgtttttgtcacca	ctggaaccccaaaggaaaat	595
3'UTR-A	gggtcccattattttcttaaac	tgagagccttttgaagcaga	596
3'UTR-B	ccctttaatgtggttgcaga	tggatgagactggacctttg	546
3'UTR-C	aactttatttgttgggcagagg	ccagtactttgggaggtgga	540

3'UTR-D	accacgcccagctaattttt	tgggccaggtttcacatta	630
3'UTR-E	tgtgttttcagagatgagtcct	aatggctgcaggttctcagt	574
3'UTR-F	tcctccccttctctgaaatg	aactccagaagatgacacagt	599
3'UTR-G	acccagcccatgtgtgatag	agacttgatgcaggtcattt	593
3'UTR-H	aatttattgtgaggaccttct	caggagaagttagaaaacatcg	490

Table 2 – Primer pairs for *CILK1* clone sequencing

NAME	SEQUENCE (5'-3')	LENGTH
pEGFP-C2-ICK-XhoI-FP	CCGACTCGAGCCATGAATAGATACACAACA	30
pEGFP-C2-ICK-EcoRI-RP	CCGACTCGAGCCATGAATAGATACACAACA	29
ICK-cDNA-Ex4-FP	ATTATGTATCTACCAGATG	19
ICK-cDNA-Ex5-RP	GCCTGAGGGTGTAAACTTC	19
ICK-cDNA-Ex7-FP	GAGGTCTCCAGGACAGAT	18
ICK-cDNA-Ex7-RP	CGACTGTGGATGCTTGTT	18
ICK-cDNA-Ex9-FP	GCACTCTCGATACTTGCCTG	20
ICK-cDNA-Ex10-RP	CTGAATTTACTTTGCTGATTAC	22

Table 3 – *CILK1* site-directed mutagenesis primers

Name	Sequence (5'-3')	Length
ICK-F79V-FP	attttccttcatgtactcgacgataaaataaagatgatcattttccctg	49
ICK-F79V-RP	attttccttcatgtactcgacgataaaataaagatgatcattttccctg	49
ICK-R272Q-FP	cctgactagctgttggttgtttcttgggatcccac	35
ICK-R272Q-RP	gtgggatcccaagaaacaaccaacagctagtcagg	35
ICK-W624R-FP	gtacttggaagcccggtctgtccggccat	29
ICK-W624R-RP	atggccggacagaccgggcttccaagtac	29
ICK-T348M-FP	gctttgtaggggtacatgagatgcagagggg	31
ICK-T348M-RP	cccctctgcatctcatgtacccctacaaagc	31

Appendix for chapter 5

Appendix 4

Alignment statistics of samples subjected to WES

Family ID	**GLH016**	**GLH016**	**GLH018**	**GLH018**
	Sample: b1	**Sample: b2**	**Sample: a1**	**Sample: b1**
Total number of reads	160,234,715	161,836,182	150,167,174	177,878,448
Total number of reads aligned	139,389,896	145,236,606	134,103,232	154,573,492
% of reads aligned	86.99	89.74	89.30	86.90
Target length	62,085,286	62,085,286	62,085,286	62,085,286
Target covered	61,447,676	61,540,322	61,417,551	61,467,013
% of target covered	**98.97**	**99.12**	**98.92**	**99.00**
% of target covered with at least 5X read depth	96.10	96.47	96.13	96.10
% of target covered with at least 10X read depth	93.86	94.39	93.94	93.94
% of target covered with at least 15X read depth	91.80	92.45	91.89	92.06
% of target covered with at least 20X read depth	89.75	90.52	89.83	90.26
Average read depth (against covered coordinates)	105.31	108.32	104.67	116.68
Average read depth (against bed coordinates)	**104.23**	**107.37**	**103.54**	**115.52**

Family ID	**GLH042**	**SCT187**	**GMW013**	**GMW013**
	Sample: b1	**Sample: c3**	**Sample: b2**	**Sample: b3**
Total number of reads	131,433,762	128,680,848	138,606,731	136,020,692
Total number of reads aligned	114,817,348	113,077,524	123,883,855	122,976,767
% of reads aligned	87.36	87.87	89.38	90.41
Target length	62,085,286	62,085,286	62,085,286	62,085,286
Target covered	61,302,837	61,374,049	61,345,306	61,429,181
% of target covered	**98.74**	**98.85**	**98.81**	**98.94**
% of target covered with at least 5X read depth	95.17	95.50	95.22	95.64
% of target covered with at least 10X read depth	92.24	92.63	92.37	92.98
% of target covered with at least 15X read depth	89.49	89.94	89.74	90.45
% of target covered with at least 20X read depth	86.67	87.18	87.05	87.86
Average read depth (against covered coordinates)	79.15	78.58	84.02	85.19
Average read depth (against bed coordinates)	**78.15**	**77.68**	**83.02**	**84.29**

Family ID	**MCC017**	**MCC017**	**MCC024**	**MCC024**
	Sample: a2	**Sample: b1**	**Sample: a2**	**Sample: b1**
Total number of reads	114,518,694	131,021,157	130,732,277	135,741,427
Total number of reads	104,487,268	115,555,955	119,636,448	122,917,699

aligned				
% of reads aligned	**91.24**	**88.20**	**91.51**	**90.55**
Target length	62,085,286	62,085,286	62,085,286	62,085,286
Target covered	61196387	61,224,392	61,342,687	61,419,411
% of target covered	98.57	98.61	98.80	98.93
% of target covered with at least 5X read depth	94.52	94.61	95.26	95.34
% of target covered with at least 10X read depth	91.25	91.58	92.43	92.46
% of target covered with at least 15X read depth	88.19	88.81	89.80	89.81
% of target covered with at least 20X read depth	84.93	85.91	87.09	87.07
Average read depth (against covered coordinates)	70.78	79.00	80.58	83.06
Average read depth (against bed coordinates)	**69.77**	**77.91**	**79.62**	**82.16**

Family ID	**NIH163**	**NIH163**	**NIH174**	**NIH174**
	Sample: b2	**Sample: c4**	**Sample: b4**	**Sample: b8**
Total number of reads	138,830,464	130,064,654	169,505,739	170,820,767
Total number of reads aligned	126,206,801	114,845,543	153,899,876	153,133,849
% of reads aligned	**90.91**	**88.30**	**90.79**	**89.64**
Target length	62,085,286	62,085,286	62,085,286	62,085,286
Target covered	61,438,466	61,360,590	61,546,003	61,495,987

% of target covered	98.96	98.83	99.13	99.05
% of target covered with at least 5X read depth	95.63	95.15	96.37	96.26
% of target covered with at least 10X read depth	92.99	92.17	94.35	94.15
% of target covered with at least 15X read depth	90.59	89.39	92.53	92.29
% of target covered with at least 20X read depth	88.13	86.47	90.73	90.46
Average read depth (against covered coordinates)	85.79	79.63	114.46	115.17
Average read depth (against bed coordinates)	**84.90**	**78.70**	**113.47**	**114.08**

Family ID	**SCT025**	**SCT025**	**SCT131**	**SCT131**
	Sample: d32	**Sample: e41**	**Sample: b9**	**Sample: c20**
Total number of reads	125,771,298	149,434,932	16,336,296	19,041,359
Total number of reads aligned	105,870,276	133,477,906	3,294,299	5,963,186
% of reads aligned	**84.18**	**89.32**	**20.16**	**31.32**
Target length	62,085,286	62,085,286	62,085,286	62,085,286
Target covered	61,049,855	61,324,240	43,383,413	51,304,590
% of target covered	98.33	98.77	69.88	82.63
% of target covered with at least 5X read depth	94.86	95.68	23.29	43.37
% of target covered with at	92.07	93.34	4.73	16.08

least 10X read depth				
% of target covered with at least 15X read depth	89.44	91.25	0.88	5.14
% of target covered with at least 20X read depth	86.72	89.15	0.20	1.64
Average read depth (against covered coordinates)	83.36	102.13	4.06	6.09
Average read depth(against bed coordinates)	**81.97**	**100.88**	**2.84**	**5.03**

Family ID	**SCT229**	**SCT229**	**SCT271**	**SCT271**
	Sample: b6	**Sample: c1**	**Sample: c14**	**Sample: d4**
Total number of reads	164,692,912	170,667,068	148,964,022	129,016,269
Total number of reads aligned	146,938,555	150,071,277	132,641,225	113,225,543
% of reads aligned	**89.22**	**87.93**	**89.04**	**87.76**
Target length	62,085,286	62,085,286	62,085,286	62,085,286
Target covered	61,496,183	61,459,265	61,449,815	61,237,443
% of target covered	99.05	98.99	98.98	98.63
% of target covered with at least 5X read depth	96.36	96.28	95.55	94.86
% of target covered with at least 10X read depth	94.31	94.24	92.84	91.87
% of target covered with at least 15X read depth	92.42	92.41	90.46	89.08
% of target covered with at least 20X read depth	90.56	90.62	88.07	86.20

Average read depth (against covered coordinates)	112.61	114.38	88.27	79.43
Average read depth (against bed coordinates)	**111.54**	**113.23**	**87.37**	**78.35**

Family ID	**GLH013**	**GLH013**	**NIH007**	**NIH007**
	Sample: C1	**Sample: C2**	**Sample: B5**	**Sample: C1**
Total number of reads	144,828,533	136,451,138	120,511,259	114,651,609
Total number of reads aligned	132,482,284	125,574,247	111,219,024	105,791,996
% of reads aligned	**91.47**	**92.03**	**92.29**	**92.27**
Target length	62,085,286	62,085,286	62,085,286	62,085,286
Target covered	61,435,989	61,387,349	61,364,793	61,324,306
% of target covered	98.95	98.88	98.84	98.77
% of target covered with at least 5X read depth	95.89	95.74	95.65	95.52
% of target covered with at least 10X read depth	93.51	93.29	92.93	92.75
% of target covered with at least 15X read depth	91.32	91.03	90.28	90.01
% of target covered with at least 20X read depth	89.14	88.74	87.51	87.14
Average read depth (against covered coordinates)	95.18	90.64	80.27	77.45
Average read depth (against bed coordinates)	**94.18**	**89.62**	**79.33**	**76.50**

Family ID	**NIH123**	**NIH123**	**NIH162**	**NIH162**
	Sample: C5	**Sample: C6**	**Sample: C1**	**Sample: C2**
Total number of reads	129,344,453	125,102,010	140,777,583	140,229,083
Total number of reads aligned	120,245,219	115,234,652	130,753,154	130,042,533
% of reads aligned	**92.96**	**92.11**	**92.88**	**92.74**
Target length	62,085,286	62,085,286	62085286	62,085,286
Target covered	61,458,891	61,429,140	61495922	61,427,527
% of target covered	98.99	98.94	99.05	98.94
% of target covered with at least 5X read depth	96.05	95.92	96.18	96.03
% of target covered with at least 10X read depth	93.59	93.38	93.86	93.78
% of target covered with at least 15X read depth	91.20	90.89	91.68	91.65
% of target covered with at least 20X read depth	88.74	88.32	89.46	89.49
Average read depth (against covered coordinates)	83.44	84.28	93.88	94.03
Average read depth (against bed coordinates)	**82.60**	**83.39**	**92.99**	**93.03**

Family ID	**NIH123**	**NIH123**	**NIH162**	**NIH162**
	Sample: C5	**Sample: C6**	**Sample: C1**	**Sample: C2**
Total number of reads	129,344,453	125,102,010	140,777,583	140,229,083

Total number of reads aligned	120,245,219	115,234,652	130,753,154	130,042,533
% of reads aligned	**92.96**	**92.11**	**92.88**	**92.74**
Target length	62,085,286	62,085,286	62085286	62,085,286
Target covered	61,458,891	61,429,140	61495922	61,427,527
% of target covered	98.99	98.94	99.05	98.94
% of target covered with at least 5X read depth	96.05	95.92	96.18	96.03
% of target covered with at least 10X read depth	93.59	93.38	93.86	93.78
% of target covered with at least 15X read depth	91.20	90.89	91.68	91.65
% of target covered with at least 20X read depth	88.74	88.32	89.46	89.49
Average read depth (against covered coordinates)	83.44	84.28	93.88	94.03
Average read depth (against bed coordinates)	**82.60**	**83.39**	**92.99**	**93.03**

Family id	**NIH143**	**SCT187**	**SCT234**	**SCT234**
	Sample: D3	**Sample: B6**	**Sample: B5**	**Sample: C2**
Total number of reads	118,572,852	79,507,450	85,587,760	124,359,662
Total number of reads aligned	104,704,840	66,682,686	69,463,876	112,769,193
% of reads aligned	**88.30**	**83.87**	**81.16**	**90.68**
Target length	62,085,286	62,085,286	62,085,286	62,085,286

Target covered	61,144,577	60,948,344	60,845,167	61,289,498
% of target covered	98.48	98.17	98.00	98.72
% of target covered with at least 5X read depth	94.89	94.01	93.57	95.51
% of target covered with at least 10X read depth	91.96	89.66	89.22	92.82
% of target covered with at least 15X read depth	89.19	84.94	84.62	90.26
% of target covered with at least 20X read depth	86.31	79.55	79.52	87.62
Average read depth (against covered coordinates)	78.49	50.51	52.29	81.96
Average read depth (against bed coordinates)	**77.30**	**49.58**	**51.24**	**80.91**

Family ID	**GLH002**	**GLH002**	**GLH007**	**GLH007**
	Sample: C1	**Sample: C2**	**Sample: B7**	**Sample: B9**
Total number of reads	107,941,546	140,236,695	129,651,473	132,913,595
Total number of reads aligned	90,977,974	128,070,223	114,541,544	118,825,470
% of reads aligned	**84.28**	**91.32**	**88.35**	**89.40**
Target length	62,085,286	62,085,286	62,085,286	62,085,286
Target covered	60,871,173	61,302,108	61,378,015	61,170,733
% of target covered	98.04	98.74	98.86	98.53
% of target covered with at least 5X read depth	93.17	95.51	95.11	94.73
% of target covered with at	89.39	93.18	92.07	92.08

least 10X read depth				
% of target covered with at least 15X read depth	85.70	91.12	89.25	89.74
% of target covered with at least 20X read depth	81.70	89.03	86.28	87.33
Average read depth (against covered coordinates)	66.588	95.76	74.79	87.96
Average read depth (against bed coordinates)	**65.29**	**94.56**	**73.93**	**86.67**

Family ID	**NIMS005**	**NIMS005**	**NIMS005**
	Sample: B1	**Sample: C1**	**Sample: C2**
Total number of reads	123,476,006	97,567,450	126,858,033
Total number of reads aligned	104,743,009	85,441,256	109,969,901
% of reads aligned	**84.83**	**87.57**	**86.69**
Target length	62,085,286	62,085,286	62,085,286
Target covered	60,894,572	60,675,235	60,950,013
% of target covered	98.08	97.73	98.17
% of target covered with at least 5X read depth	94.12	93.44	94.42
% of target covered with at least 10X read depth	91.26	90.10	91.70
% of target covered with at least 15X read depth	88.60	86.73	89.17

% of target covered with at least 20X read depth	85.78	83.07	86.54
Average read depth (against covered coordinates)	83.86	68.63	89.75
Average read depth (against bed coordinates)	**82.25**	**67.07**	**88.10**

Family ID	**SCT204**	**SCT204**	**SCT204**
	Sample: B3	**Sample: C1**	**Sample: C4**
Total number of reads	115,428,360	132,734,001	77,273,186
Total number of reads aligned	101,642,575	120,469,639	62,458,754
% of reads aligned	**88.06**	**90.76**	**80.83**
Target length	62,085,286	62,085,286	62,085,286
Target covered	60,914,924	61,225,108	60,048,582
% of target covered	98.11	98.61	96.72
% of target covered with at least 5X read depth	94.21	95.35	91.35
% of target covered with at least 10X read depth	91.32	92.95	86.36
% of target covered with at least 15X read depth	88.59	90.75	81.07
% of target covered with at least 20X read depth	85.70	88.51	75.51

Average read depth (against covered coordinates)	80.45	92.61	52.85
Average read depth (against bed coordinates)	**78.93**	**91.33**	**51.12**

Family ID	**SCT290**	**SCT290**	**SCT290**	**SCT290**
	Sample: C9	**Sample: C11**	**Sample: C14**	**Sample: C15**
Total number of reads	81,633,625	101,000,325	88,591,550	120,018,216
Total number of reads aligned	72,272,980	87,460,613	74,701,456	100,719,677
% of reads aligned	**88.53**	**86.59**	**84.32**	**83.92**
Target length	62,085,286	62,085,286	62,085,286	62,085,286
Target covered	60,518,635	60,630,084	60,380,068	60,803,100
% of target covered	97.48	97.65	97.25	97.93
% of target covered with at least 5X read depth	92.80	93.09	92.53	93.97
% of target covered with at least 10X read depth	88.64	89.57	88.52	91.02
% of target covered with at least 15X read depth	84.31	86.05	84.40	88.22
% of target covered with at least 20X read depth	79.62	82.28	80.00	85.26
Average read depth (against covered coordinates)	58.44	68.81	62.24	81.91
Average read depth (against bed coordinates)	**56.96**	**67.20**	**60.53**	**80.22**

References

Accardi, M. V., Huang, H., & Authier, S. (2018). Seizure liability assessments using the hippocampal tissue slice: Comparison of non-clinical species. *J Pharmacol Toxicol Methods, 93*, 59-68. doi:10.1016/j.vascn.2017.11.003

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., . . . Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nat Methods, 7*(4), 248-249. doi:nmeth0410-248 [pii]

10.1038/nmeth0410-248

Alieva, I. B., & Uzbekov, R. E. (2008). The centrosome is a polyfunctional multiprotein cell complex. *Biochemistry (Mosc), 73*(6), 626-643. doi:10.1134/s0006297908060023

Angelozzi, M., & Lefebvre, V. (2019). SOXopathies: Growing Family of Developmental Disorders Due to SOX Mutations. *Trends Genet, 35*(9), 658-671. doi:10.1016/j.tig.2019.06.003

Anitha, A., & Senthilkumaran, B. (2020). Role of sox30 in regulating testicular steroidogenesis of common carp. *J Steroid Biochem Mol Biol, 204*, 105769. doi:10.1016/j.jsbmb.2020.105769

Arain, F. M., Boyd, K. L., & Gallagher, M. J. (2012). Decreased viability and absence-like epilepsy in mice lacking or deficient in the GABAA receptor alpha1 subunit. *Epilepsia, 53*(8), e161-165. doi:10.1111/j.1528-1167.2012.03596.x

Arya, V. B., Chawla, G., Nambisan, A. K. R., Muhi-Iddin, N., Vamvakiti, E., Ajzensztejn, M., . . . Kapoor, R. R. (2019). Xq27.1 Duplication Encompassing SOX3: Variable Phenotype and Smallest Duplication Associated with Hypopituitarism to Date - A Large Case Series of Unrelated Patients and a Literature Review. *Horm Res Paediatr, 92*(6), 382-389. doi:10.1159/000503784

Ashton, J. J., Mossotto, E., Beattie, R. M., & Ennis, S. (2020). TTC7A Variants Previously Described to Cause Enteropathy Are Observed on a Single Haplotype and Appear Non-pathogenic in Pediatric Inflammatory Bowel Disease Patients. *J Clin Immunol, 40*(1), 245-247. doi:10.1007/s10875-019-00726-0

Badawy, R. A., Curatolo, J. M., Newton, M., Berkovic, S. F., & Macdonell, R. A. (2006). Sleep deprivation increases cortical excitability in

epilepsy: syndrome-specific effects. *Neurology, 67*(6), 1018-1022. doi:10.1212/01.wnl.0000237392.64230.f7

Badawy, R. A., Macdonell, R. A., Jackson, G. D., & Berkovic, S. F. (2009). Why do seizures in generalized epilepsy often occur in the morning? *Neurology, 73*(3), 218-222. doi:10.1212/WNL.0b013e3181ae7ca6

Bai, S., Fu, K., Yin, H., Cui, Y., Yue, Q., Li, W., . . . Ye, L. (2018). Sox30 initiates transcription of haploid genes during late meiosis and spermiogenesis in mouse testes. *Development, 145*(13). doi:10.1242/dev.164855

Bailey, J. N., de Nijs, L., Bai, D., Suzuki, T., Miyamoto, H., Tanaka, M., . . . Delgado-Escueta, A. V. (2018). Variant Intestinal-Cell Kinase in Juvenile Myoclonic Epilepsy. *N Engl J Med, 378*(11), 1018-1028. doi:10.1056/NEJMoa1700175

Bailey, J. N., Patterson, C., de Nijs, L., Duron, R. M., Nguyen, V. H., Tanaka, M., . . . Delgado-Escueta, A. V. (2017). EFHC1 variants in juvenile myoclonic epilepsy: reanalysis according to NHGRI and ACMG guidelines for assigning disease causality. *Genet Med, 19*(2), 144-156. doi:10.1038/gim.2016.86

Bani-Yaghoub, M., Tremblay, R. G., Lei, J. X., Zhang, D., Zurakowski, B., Sandhu, J. K., . . . Sikorska, M. (2006). Role of Sox2 in the development of the mouse neocortex. *Dev Biol, 295*(1), 52-66. doi:10.1016/j.ydbio.2006.03.007

Bar, C., Barcia, G., Jennesson, M., Le Guyader, G., Schneider, A., Mignot, C., . . . Nabbout, R. (2020). Expanding the genetic and phenotypic relevance of KCNB1 variants in developmental and epileptic encephalopathies: 27 new patients and overview of the literature. *Hum Mutat, 41*(1), 69-80. doi:10.1002/humu.23915

Baraban, S. C. (2007). Emerging epilepsy models: insights from mice, flies, worms and fish. *Curr Opin Neurol, 20*(2), 164-168. doi:10.1097/WCO.0b013e328042bae0

Barak, P. (2021). *Dissecting the molecular genetic basis of juvenile myoclonic epilepsy.* (Ph.D.), JNCASR, Bangalore.

Bargmann, C. I. (1998). Neurobiology of the Caenorhabditis elegans genome. *Science, 282*(5396), 2028-2033. doi:10.1126/science.282.5396.2028

Barrionuevo, F., & Scherer, G. (2010). SOX E genes: SOX9 and SOX8 in mammalian testis development. *Int J Biochem Cell Biol, 42*(3), 433-436. doi:10.1016/j.biocel.2009.07.015

Bedford, M. T., & Clarke, S. G. (2009). Protein arginine methylation in mammals: who, what, and why. *Mol Cell, 33*(1), 1-13. doi:10.1016/j.molcel.2008.12.013

Bender, A. C., Morse, R. P., Scott, R. C., Holmes, G. L., & Lenck-Santini, P. P. (2012). SCN1A mutations in Dravet syndrome: impact of interneuron dysfunction on neural networks and cognitive outcome. *Epilepsy Behav, 23*(3), 177-186. doi:10.1016/j.yebeh.2011.11.022

Benito-Kwiecinski, S., & Lancaster, M. A. (2020). Brain Organoids: Human Neurodevelopment in a Dish. *Cold Spring Harb Perspect Biol, 12*(8). doi:10.1101/cshperspect.a035709

Berg, A. T., Berkovic, S. F., Brodie, M. J., Buchhalter, J., Cross, J. H., van Emde Boas, W., . . . Scheffer, I. E. (2010). Revised terminology and concepts for organization of seizures and epilepsies: report of the ILAE Commission on Classification and Terminology, 2005-2009. *Epilepsia, 51*(4), 676-685. doi:10.1111/j.1528-1167.2010.02522.x

Betting, L. E., Mory, S. B., Li, L. M., Lopes-Cendes, I., Guerreiro, M. M., Guerreiro, C. A., & Cendes, F. (2006). Voxel-based morphometry in patients with idiopathic generalized epilepsies. *Neuroimage, 32*(2), 498-502. doi:10.1016/j.neuroimage.2006.04.174

Blaya, C., Moorjani, P., Salum, G. A., Goncalves, L., Weiss, L. A., Leistner-Segal, S., . . . Smoller, J. W. (2009). Preliminary evidence of association between EFHC2, a gene implicated in fear recognition, and harm avoidance. *Neurosci Lett, 452*(1), 84-86. doi:10.1016/j.neulet.2009.01.036

Borghi, R., Magliocca, V., Petrini, S., Conti, L. A., Moreno, S., Bertini, E., . . . Compagnucci, C. (2021). Dissecting the Role of PCDH19 in Clustering Epilepsy by Exploiting Patient-Specific Models of Neurogenesis. *J Clin Med, 10*(13). doi:10.3390/jcm10132754

Botstein, D., & Risch, N. (2003). Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat Genet, 33 Suppl*, 228-237. doi:10.1038/ng1090

Bouguenina, H., Salaun, D., Mangon, A., Muller, L., Baudelet, E., Camoin, L., . . . Badache, A. (2017). EB1-binding-myomegalin protein

complex promotes centrosomal microtubules functions. *Proc Natl Acad Sci U S A, 114*(50), E10687-E10696. doi:10.1073/pnas.1705682114

Broekhuis, J. R., Verhey, K. J., & Jansen, G. (2014). Regulation of cilium length and intraflagellar transport by the RCK-kinases ICK and MOK in renal epithelial cells. *PLoS ONE, 9*(9), e108470. doi:10.1371/journal.pone.0108470

Calahorro, F., & Ruiz-Rubio, M. (2013). Human alpha- and beta-NRXN1 isoforms rescue behavioral impairments of Caenorhabditis elegans neurexin-deficient mutants. *Genes Brain Behav, 12*(4), 453-464. doi:10.1111/gbb.12046

Cavalleri, G. L., Walley, N. M., Soranzo, N., Mulley, J., Doherty, C. P., Kapoor, A., . . . Sisodiya, S. M. (2007). A multicenter study of BRD2 as a risk factor for juvenile myoclonic epilepsy. *Epilepsia, 48*(4), 706-712. doi:10.1111/j.1528-1167.2007.00977.x

Cepeda, C., Chen, J. Y., Wu, J. Y., Fisher, R. S., Vinters, H. V., Mathern, G. W., & Levine, M. S. (2014). Pacemaker GABA synaptic activity may contribute to network synchronization in pediatric cortical dysplasia. *Neurobiol Dis, 62*, 208-217. doi:10.1016/j.nbd.2013.10.001

Chan, C. K., Low, J. S., Lim, K. S., Low, S. K., Tan, C. T., & Ng, C. C. (2020). Whole exome sequencing identifies a novel SCN1A mutation in genetic (idiopathic) generalized epilepsy and juvenile myoclonic epilepsy subtypes. *Neurol Sci, 41*(3), 591-598. doi:10.1007/s10072-019-04122-9

Charlier, C., Singh, N. A., Ryan, S. G., Lewis, T. B., Reus, B. E., Leach, R. J., & Leppert, M. (1998). A pore mutation in a novel KQT-like potassium channel gene in an idiopathic epilepsy family. *Nat Genet, 18*(1), 53-55. doi:10.1038/ng0198-53

Chavali, P. L., Putz, M., & Gergely, F. (2014). Small organelle, big responsibility: the role of centrosomes in development and disease. *Philos Trans R Soc Lond B Biol Sci, 369*(1650). doi:10.1098/rstb.2013.0468

Chaya, T., Omori, Y., Kuwahara, R., & Furukawa, T. (2014). ICK is essential for cell type-specific ciliogenesis and the regulation of ciliary transport. *EMBO J, 33*(11), 1227-1242. doi:10.1002/embj.201488175

Chen, J. F., Zhang, Y., Wilde, J., Hansen, K. C., Lai, F., & Niswander, L. (2014). Microcephaly disease gene Wdr62 regulates mitotic progression of embryonic neural stem cells and brain size. *Nat Commun, 5*, 3885. doi:10.1038/ncomms4885

Chowdhury, A., & Brodie, M. J. (2016). Pharmacological outcomes in juvenile myoclonic epilepsy: Support for sodium valproate. *Epilepsy Res, 119*, 62-66. doi:10.1016/j.eplepsyres.2015.11.012

Claes, L., Del-Favero, J., Ceulemans, B., Lagae, L., Van Broeckhoven, C., & De Jonghe, P. (2001). De novo mutations in the sodium-channel gene SCN1A cause severe myoclonic epilepsy of infancy. *Am J Hum Genet, 68*(6), 1327-1332. doi:10.1086/320609

Colicino, E. G., & Hehnly, H. (2018). Regulating a key mitotic regulator, polo-like kinase 1 (PLK1). *Cytoskeleton (Hoboken), 75*(11), 481-494. doi:10.1002/cm.21504

Consortium, E., Consortium, E. M., Steffens, M., Leu, C., Ruppert, A. K., Zara, F., . . . Sander, T. (2012). Genome-wide association analysis of genetic generalized epilepsies implicates susceptibility loci at 1q43, 2p16.1, 2q22.3 and 17q21.32. *Hum Mol Genet, 21*(24), 5359-5372. doi:10.1093/hmg/dds373

Conte, F. F., Ribeiro, P. A., Marchesini, R. B., Pascoal, V. D., Silva, J. M., Oliveira, A. R., . . . Lopes-Cendes, I. (2009). Expression profile and distribution of Efhc1 gene transcript during rodent brain development. *J Mol Neurosci, 39*(1-2), 69-77. doi:10.1007/s12031-009-9179-6

Cossette, P., Liu, L., Brisebois, K., Dong, H., Lortie, A., Vanasse, M., . . . Rouleau, G. A. (2002). Mutation of GABRA1 in an autosomal dominant form of juvenile myoclonic epilepsy. *Nat Genet, 31*(2), 184-189. doi:10.1038/ng885

Coste de Bagneaux, P., von Elsner, L., Bierhals, T., Campiglio, M., Johannsen, J., Obermair, G. J., . . . Kutsche, K. (2020). A homozygous missense variant in CACNB4 encoding the auxiliary calcium channel beta4 subunit causes a severe neurodevelopmental disorder and impairs channel and non-channel functions. *PLoS Genet, 16*(3), e1008625. doi:10.1371/journal.pgen.1008625

da Silva Sousa, P., Lin, K., Garzon, E., Sakamoto, A. C., & Yacubian, E. M. (2005). Self-perception of factors that precipitate or inhibit seizures in juvenile myoclonic epilepsy. *Seizure, 14*(5), 340-346. doi:10.1016/j.seizure.2005.04.007

Damiani, D., Goffinet, A. M., Alberts, A., & Tissir, F. (2016). Lack of Diaph3 relaxes the spindle checkpoint causing the loss of neural progenitors. *Nat Commun, 7*, 13509. doi:10.1038/ncomms13509

de Brouwer, A. P., Yntema, H. G., Kleefstra, T., Lugtenberg, D., Oudakker, A. R., de Vries, B. B., . . . Hamel, B. C. (2007). Mutation frequencies of X-linked mental retardation genes in families from the EuroMRX consortium. *Hum Mutat, 28*(2), 207-208. doi:10.1002/humu.9482

de Kovel, C. G. F., Syrbe, S., Brilstra, E. H., Verbeek, N., Kerr, B., Dubbs, H., . . . Koeleman, B. P. C. (2017). Neurodevelopmental Disorders Caused by De Novo Variants in KCNB1 Genotypes and Phenotypes. *JAMA Neurol, 74*(10), 1228-1236. doi:10.1001/jamaneurol.2017.1714

De Martino, S. P., Errington, F., Ashworth, A., Jowett, T., & Austin, C. A. (1999). sox30: a novel zebrafish sox gene expressed in a restricted manner at the midbrain-hindbrain boundary during neurogenesis. *Dev Genes Evol, 209*(6), 357-362. doi:10.1007/s004270050264

de Nijs, L., Leon, C., Nguyen, L., Loturco, J. J., Delgado-Escueta, A. V., Grisar, T., & Lakaye, B. (2009). EFHC1 interacts with microtubules to regulate cell division and cortical development. *Nat Neurosci, 12*(10), 1266-1274. doi:10.1038/nn.2390

de Nijs, L., Wolkoff, N., Coumans, B., Delgado-Escueta, A. V., Grisar, T., & Lakaye, B. (2012). Mutations of EFHC1, linked to juvenile myoclonic epilepsy, disrupt radial and tangential migrations during brain development. *Hum Mol Genet, 21*(23), 5106-5117. doi:10.1093/hmg/dds356

Delgado-Escueta, A. V. (2007). Advances in genetics of juvenile myoclonic epilepsies. *Epilepsy Curr, 7*(3), 61-67. doi:10.1111/j.1535-7511.2007.00171.x

Delgado-Escueta, A. V., Koeleman, B. P., Bailey, J. N., Medina, M. T., & Duron, R. M. (2013). The quest for juvenile myoclonic epilepsy genes. *Epilepsy Behav, 28 Suppl 1*, S52-57. doi:10.1016/j.yebeh.2012.06.033

Deppe, M., Kellinghaus, C., Duning, T., Moddel, G., Mohammadi, S., Deppe, K., . . . Knecht, S. (2008). Nerve fiber impairment of anterior thalamocortical circuitry in juvenile myoclonic epilepsy. *Neurology, 71*(24), 1981-1985. doi:10.1212/01.wnl.0000336969.98241.17

Devinsky, O., Vezzani, A., O'Brien, T. J., Jette, N., Scheffer, I. E., de Curtis, M., & Perucca, P. (2018). Epilepsy. *Nat Rev Dis Primers, 4*, 18024. doi:10.1038/nrdp.2018.24

Dhanuka, A. K., Jain, B. K., Daljit, S., & Maheshwari, D. (2001). Juvenile myoclonic epilepsy: a clinical and sleep EEG study. *Seizure, 10*(5), 374-378. doi:10.1053/seiz.2001.0522

Dibbens, L. M., Harkin, L. A., Richards, M., Hodgson, B. L., Clarke, A. L., Petrou, S., . . . Mulley, J. C. (2009). The role of neuronal GABA(A) receptor subunit mutations in idiopathic generalized epilepsies. *Neurosci Lett, 453*(3), 162-165. doi:10.1016/j.neulet.2009.02.038

Dibbens, L. M., Mullen, S., Helbig, I., Mefford, H. C., Bayly, M. A., Bellows, S., . . . Berkovic, S. F. (2009). Familial and sporadic 15q13.3 microdeletions in idiopathic generalized epilepsy: precedent for disorders with complex inheritance. *Hum Mol Genet, 18*(19), 3626-3631. doi:10.1093/hmg/ddp311

Doxsey, S. J., Stein, P., Evans, L., Calarco, P. D., & Kirschner, M. (1994). Pericentrin, a highly conserved centrosome protein involved in microtubule organization. *Cell, 76*(4), 639-650. doi:10.1016/0092-8674(94)90504-5

Durner, M., Shinnar, S., Resor, S. R., Moshe, S. L., Rosenbaum, D., Cohen, J., . . . Greenberg, D. A. (2000). No evidence for a major susceptibility locus for juvenile myoclonic epilepsy on chromosome 15q. *Am J Med Genet, 96*(1), 49-52. doi:10.1002/(sici)1096-8628(20000207)96:1<49::aid-ajmg10>3.0.co;2-j

El-Daher, M. T., Cagnard, N., Gil, M., Da Cruz, M. C., Leveau, C., Sepulveda, F., . . . de Saint Basile, G. (2018). Tetratricopeptide repeat domain 7A is a nuclear factor that modulates transcription and chromatin structure. *Cell Discov, 4*, 61. doi:10.1038/s41421-018-0061-y

Elmslie, F. V., Rees, M., Williamson, M. P., Kerr, M., Kjeldsen, M. J., Pang, K. A., . . . Gardiner, R. M. (1997). Genetic mapping of a major susceptibility locus for juvenile myoclonic epilepsy on chromosome 15q. *Hum Mol Genet, 6*(8), 1329-1334. doi:10.1093/hmg/6.8.1329

Engel, J., Jr., & International League Against, E. (2001). A proposed diagnostic scheme for people with epileptic seizures and with epilepsy: report of the ILAE Task Force on Classification and Terminology. *Epilepsia, 42*(6), 796-803. doi:10.1046/j.1528-1157.2001.10401.x

Epi25 Collaborative. Electronic address, s. b. u. e. a., & Epi, C. (2019). Ultra-Rare Genetic Variation in the Epilepsies: A Whole-Exome Sequencing Study of 17,606 Individuals. *Am J Hum Genet, 105*(2), 267-282. doi:10.1016/j.ajhg.2019.05.020

Epi, K. C. (2012). Epi4K: gene discovery in 4,000 genomes. *Epilepsia, 53*(8), 1457-1467. doi:10.1111/j.1528-1167.2012.03511.x

Epi, K. c., & Epilepsy Phenome/Genome, P. (2017). Ultra-rare genetic variation in common epilepsies: a case-control sequencing study. *Lancet Neurol, 16*(2), 135-143. doi:10.1016/S1474-4422(16)30359-3

Escayg, A., De Waard, M., Lee, D. D., Bichet, D., Wolf, P., Mayer, T., . . . Meisler, M. H. (2000). Coding and noncoding variation of the human calcium-channel beta4-subunit gene CACNB4 in patients with idiopathic generalized epilepsy and episodic ataxia. *Am J Hum Genet, 66*(5), 1531-1539. doi:10.1086/302909

Escayg, A., & Goldin, A. L. (2010). Sodium channel SCN1A and epilepsy: mutations and mechanisms. *Epilepsia, 51*(9), 1650-1658. doi:10.1111/j.1528-1167.2010.02640.x

Escayg, A., Heils, A., MacDonald, B. T., Haug, K., Sander, T., & Meisler, M. H. (2001). A novel SCN1A mutation associated with generalized epilepsy with febrile seizures plus--and prevalence of variants in patients with epilepsy. *Am J Hum Genet, 68*(4), 866-873. doi:10.1086/319524

Escayg, A., MacDonald, B. T., Meisler, M. H., Baulac, S., Huberfeld, G., An-Gourfinkel, I., . . . Malafosse, A. (2000). Mutations of SCN1A, encoding a neuronal sodium channel, in two families with GEFS+2. *Nat Genet, 24*(4), 343-345. doi:10.1038/74159

Fagerberg, L., Hallstrom, B. M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., . . . Uhlen, M. (2014). Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics, 13*(2), 397-406. doi:10.1074/mcp.M113.035600

Feng, C. A., Spiller, C., Merriner, D. J., O'Bryan, M. K., Bowles, J., & Koopman, P. (2017). SOX30 is required for male fertility in mice. *Sci Rep, 7*(1), 17619. doi:10.1038/s41598-017-17854-5

Feng, Y., Kapornai, K., Kiss, E., Tamas, Z., Mayer, L., Baji, I., . . . Barr, C. L. (2010). Association of the GABRD gene and childhood-onset mood

disorders. *Genes Brain Behav, 9*(6), 668-672. doi:10.1111/j.1601-183X.2010.00598.x

Fisher, J. L. (2004). A mutation in the GABAA receptor alpha 1 subunit linked to human epilepsy affects channel gating properties. *Neuropharmacology, 46*(5), 629-637. doi:10.1016/j.neuropharm.2003.11.015

Fisher, R. S., Acevedo, C., Arzimanoglou, A., Bogacz, A., Cross, J. H., Elger, C. E., . . . Wiebe, S. (2014). ILAE official report: a practical clinical definition of epilepsy. *Epilepsia, 55*(4), 475-482. doi:10.1111/epi.12550

Fisher, R. S., Cross, J. H., French, J. A., Higurashi, N., Hirsch, E., Jansen, F. E., . . . Zuberi, S. M. (2017). Operational classification of seizure types by the International League Against Epilepsy: Position Paper of the ILAE Commission for Classification and Terminology. *Epilepsia, 58*(4), 522-530. doi:10.1111/epi.13670

Flex, E., Pizzuti, A., Di Bonaventura, C., Douzgou, S., Egeo, G., Fattouch, J., . . . Giallonardo, A. T. (2005). LGI1 gene mutation screening in sporadic partial epilepsy with auditory features. *J Neurol, 252*(1), 62-66. doi:10.1007/s00415-005-0599-0

Fontana, B. D., Mezzomo, N. J., Kalueff, A. V., & Rosemberg, D. B. (2018). The developing utility of zebrafish models of neurological and neuropsychiatric disorders: A critical review. *Exp Neurol, 299*(Pt A), 157-171. doi:10.1016/j.expneurol.2017.10.004

Fu, Z., Gailey, C. D., Wang, E. J., & Brautigan, D. L. (2019). Ciliogenesis associated kinase 1: targets and functions in various organ systems. *FEBS Lett, 593*(21), 2990-3002. doi:10.1002/1873-3468.13600

Fu, Z., Kim, J., Vidrich, A., Sturgill, T. W., & Cohn, S. M. (2009). Intestinal cell kinase, a MAP kinase-related kinase, regulates proliferation and G1 cell cycle progression of intestinal epithelial cells. *Am J Physiol Gastrointest Liver Physiol, 297*(4), G632-640. doi:10.1152/ajpgi.00066.2009

Gama, L., & Breitwieser, G. E. (1998). A carboxyl-terminal domain controls the cooperativity for extracellular Ca2+ activation of the human calcium sensing receptor. A study with receptor-green fluorescent protein fusions. *J Biol Chem, 273*(45), 29712-29718. doi:10.1074/jbc.273.45.29712

Ganapathy, A., Mishra, A., Soni, M. R., Kumar, P., Sadagopan, M., Kanthi, A. V., . . . Mannan, A. U. (2019). Multi-gene testing in neurological disorders showed an improved diagnostic yield: data from over 1000 Indian patients. *J Neurol, 266*(8), 1919-1926. doi:10.1007/s00415-019-09358-1

Gardiner, M. (2005). Genetics of idiopathic generalized epilepsies. *Epilepsia, 46 Suppl 9*, 15-20. doi:10.1111/j.1528-1167.2005.00310.x

Garner, J. P. (2014). The significance of meaning: why do over 90% of behavioral neuroscience results fail to translate to humans, and what can we do to fix it? *ILAR J, 55*(3), 438-456. doi:10.1093/ilar/ilu047

Gastaut, H. (1969). Classification of the epilepsies. Proposal for an international classification. *Epilepsia, 10*, Suppl:14-21. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4980563

Genomes Project, C., Abecasis, G. R., Auton, A., Brooks, L. D., DePristo, M. A., Durbin, R. M., . . . McVean, G. A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature, 491*(7422), 56-65. doi:10.1038/nature11632

Genton, P., & Gelisse, P. (2013). The history of juvenile myoclonic epilepsy. *Epilepsy Behav, 28 Suppl 1*, S2-7. doi:10.1016/j.yebeh.2013.01.002

Gilsoul, M., Grisar, T., Delgado-Escueta, A. V., de Nijs, L., & Lakaye, B. (2019). Subtle Brain Developmental Abnormalities in the Pathogenesis of Juvenile Myoclonic Epilepsy. *Front Cell Neurosci, 13*, 433. doi:10.3389/fncel.2019.00433

Giorgi, F. S., Guida, M., Caciagli, L., Pagni, C., Pizzanelli, C., Bonanni, E., . . . Bonuccelli, U. (2016). Social cognition in Juvenile Myoclonic Epilepsy. *Epilepsy Res, 128*, 61-67. doi:10.1016/j.eplepsyres.2016.10.017

Gowers, W. R. (1901). *Epilepsy and Other Chronic Convulsive Diseases: Their Causes, Symptoms, and Treatment*: Old Hickory Bookshop, 1901.

Greenberg, D. A., Cayanis, E., Strug, L., Marathe, S., Durner, M., Pal, D. K., . . . Kang, H. (2005). Malic enzyme 2 may underlie susceptibility to adolescent-onset idiopathic generalized epilepsy. *Am J Hum Genet, 76*(1), 139-146. doi:10.1086/426735

Greenberg, D. A., Durner, M., Shinnar, S., Resor, S., Rosenbaum, D., Klotz, I., . . . Altstiel, L. (1996). Association of HLA class II alleles in patients with juvenile myoclonic epilepsy compared with patients with other forms of adolescent-onset generalized epilepsy. *Neurology, 47*(3), 750-755. doi:10.1212/wnl.47.3.750

Griffin, A., Krasniak, C., & Baraban, S. C. (2016). Advancing epilepsy treatment through personalized genetic zebrafish models. *Prog Brain Res, 226*, 195-207. doi:10.1016/bs.pbr.2016.03.012

Griffith, E., Walker, S., Martin, C. A., Vagnarelli, P., Stiff, T., Vernay, B., . . . O'Driscoll, M. (2008). Mutations in pericentrin cause Seckel syndrome with defective ATR-dependent DNA damage signaling. *Nat Genet, 40*(2), 232-236. doi:10.1038/ng.2007.80

Grimm, D., Bauer, J., Wise, P., Kruger, M., Simonsen, U., Wehland, M., . . . Corydon, T. J. (2020). The role of SOX family members in solid tumours and metastasis. *Semin Cancer Biol, 67*(Pt 1), 122-153. doi:10.1016/j.semcancer.2019.03.004

Grisar, T., Lakaye, B., de Nijs, L., LoTurco, J., Daga, A., & Delgado-Escueta, A. V. (2012). Myoclonin1/EFHC1 in cell division, neuroblast migration, synapse/dendrite formation in juvenile myoclonic epilepsy. In th, J. L. Noebels, M. Avoli, M. A. Rogawski, R. W. Olsen, & A. V. Delgado-Escueta (Eds.), *Jasper's Basic Mechanisms of the Epilepsies*. Bethesda (MD).

Gu, W., Sander, T., Heils, A., Lenzen, K. P., & Steinlein, O. K. (2005). A new EF-hand containing gene EFHC2 on Xp11.4: tentative evidence for association with juvenile myoclonic epilepsy. *Epilepsy Res, 66*(1-3), 91-98. doi:S0920-1211(05)00139-7 [pii]

10.1016/j.eplepsyres.2005.07.003

Guan, Y., Shao, Y., Li, D., & Liu, M. (2014). Generation of site-specific mutations in the rat genome via CRISPR/Cas9. *Methods Enzymol, 546*, 297-317. doi:10.1016/B978-0-12-801185-0.00014-3

Guissart, C., Li, X., Leheup, B., Drouot, N., Montaut-Verient, B., Raffo, E., . . . Koenig, M. (2015). Mutation of SLC9A1, encoding the major Na(+)/H(+) exchanger, causes ataxia-deafness Lichtenstein-Knorr syndrome. *Hum Mol Genet, 24*(2), 463-470. doi:10.1093/hmg/ddu461

Gurkaslar, H. K., Culfa, E., Arslanhan, M. D., Lince-Faria, M., & Firat-Karalar, E. N. (2020). CCDC57 Cooperates with Microtubules and

Microcephaly Protein CEP63 and Regulates Centriole Duplication and Mitotic Progression. *Cell Rep, 31*(6), 107630. doi:10.1016/j.celrep.2020.107630

Hamilton, N. B., Clarke, L. E., Arancibia-Carcamo, I. L., Kougioumtzidou, E., Matthey, M., Karadottir, R., . . . Attwell, D. (2017). Endogenous GABA controls oligodendrocyte lineage cell number, myelination, and CNS internode length. *Glia, 65*(2), 309-321. doi:10.1002/glia.23093

Han, F., Dong, Y., Liu, W., Ma, X., Shi, R., Chen, H., . . . Liu, J. (2014). Epigenetic regulation of sox30 is associated with testis development in mice. *PLoS ONE, 9*(5), e97203. doi:10.1371/journal.pone.0097203

Han, F., Yin, L., Jiang, X., Zhang, X., Zhang, N., Yang, J. T., . . . Liu, J. Y. (2021). Identification of SRY-box 30 as an age-related essential gatekeeper for male germ-cell meiosis and differentiation. *Aging Cell, 20*(5), e13343. doi:10.1111/acel.13343

Hao, X., Han, F., Ma, B., Zhang, N., Chen, H., Jiang, X., . . . Liu, J. (2018). SOX30 is a key regulator of desmosomal gene suppressing tumor growth and metastasis in lung adenocarcinoma. *J Exp Clin Cancer Res, 37*(1), 111. doi:10.1186/s13046-018-0778-3

Harley, V. R., Lovell-Badge, R., & Goodfellow, P. N. (1994). Definition of a consensus DNA binding site for SRY. *Nucleic Acids Res, 22*(8), 1500-1501. doi:10.1093/nar/22.8.1500

Hauser, W. A., & Beghi, E. (2008). First seizure definitions and worldwide incidence and mortality. *Epilepsia, 49 Suppl 1*, 8-12. doi:10.1111/j.1528-1167.2008.01443.x

Heinzen, E. L., Depondt, C., Cavalleri, G. L., Ruzzo, E. K., Walley, N. M., Need, A. C., . . . Goldstein, D. B. (2012). Exome sequencing followed by large-scale genotyping fails to identify single rare variants of large effect in idiopathic generalized epilepsy. *Am J Hum Genet, 91*(2), 293-302. doi:10.1016/j.ajhg.2012.06.016

Helbig, I., Hartmann, C., & Mefford, H. C. (2013). The unexpected role of copy number variations in juvenile myoclonic epilepsy. *Epilepsy Behav, 28 Suppl 1*, S66-68. doi:10.1016/j.yebeh.2012.07.005

Helbig, I., Mefford, H. C., Sharp, A. J., Guipponi, M., Fichera, M., Franke, A., . . . Sander, T. (2009). 15q13.3 microdeletions increase risk of

idiopathic generalized epilepsy. *Nat Genet, 41*(2), 160-162. doi:10.1038/ng.292

Helsmoortel, C., Vandeweyer, G., Ordoukhanian, P., Van Nieuwerburgh, F., Van der Aa, N., & Kooy, R. F. (2015). Challenges and opportunities in the investigation of unexplained intellectual disability using family-based whole-exome sequencing. *Clin Genet, 88*(2), 140-148. doi:10.1111/cge.12470

Hempel, A., Pagnamenta, A. T., Blyth, M., Mansour, S., McConnell, V., Kou, I., . . . McNeill, A. (2016). Deletions and de novo mutations of SOX11 are associated with a neurodevelopmental disorder with features of Coffin-Siris syndrome. *J Med Genet, 53*(3), 152-162. doi:10.1136/jmedgenet-2015-103393

Heron, S. E., Khosravani, H., Varela, D., Bladen, C., Williams, T. C., Newman, M. R., . . . Zamponi, G. W. (2007). Extended spectrum of idiopathic generalized epilepsies associated with CACNA1H functional variants. *Ann Neurol, 62*(6), 560-568. doi:10.1002/ana.21169

Ho, C., Conner, D. A., Pollak, M. R., Ladd, D. J., Kifor, O., Warren, H. B., . . . Seidman, C. E. (1995). A mouse model of human familial hypocalciuric hypercalcemia and neonatal severe hyperparathyroidism. *Nat Genet, 11*(4), 389-394. doi:10.1038/ng1295-389

Hopp, T. P., & Woods, K. R. (1981). Prediction of protein antigenic determinants from amino acid sequences. *Proc Natl Acad Sci U S A, 78*(6), 3824-3828. doi:10.1073/pnas.78.6.3824

Huang, H., Winter, E. E., Wang, H., Weinstock, K. G., Xing, H., Goodstadt, L., . . . Fechtel, K. (2004). Evolutionary conservation and selection of human disease gene orthologs in the rat and mouse genomes. *Genome Biol, 5*(7), R47. doi:10.1186/gb-2004-5-7-r47

Ikeda, T. (2010). NDP Kinase 7 Is a Conserved Microtubule-Binding Protein Preferentially Expressed in Ciliated Cells. *Cell Structure and Function, 35*(1), 23-30. doi:DOI 10.1247/csf.09016

Ikeda, T., Ikeda, K., Enomoto, M., Park, M. K., Hirono, M., & Kamiya, R. (2005). The mouse ortholog of EFHC1 implicated in juvenile myoclonic epilepsy is an axonemal protein widely conserved among organisms with motile cilia and flagella. *FEBS Lett, 579*(3), 819-822. doi:10.1016/j.febslet.2004.12.070

Inan, M., Zhao, M., Manuszak, M., Karakaya, C., Rajadhyaksha, A. M., Pickel, V. M., . . . Manfredi, G. (2016). Energy deficit in parvalbumin neurons leads to circuit dysfunction, impaired sensory gating and social disability. *Neurobiol Dis, 93*, 35-46. doi:10.1016/j.nbd.2016.04.004

Iwamoto, K., Bundo, M., Yamada, K., Takao, H., Iwayama-Shigeno, Y., Yoshikawa, T., & Kato, T. (2005). DNA methylation status of SOX10 correlates with its downregulation and oligodendrocyte dysfunction in schizophrenia. *J Neurosci, 25*(22), 5376-5381. doi:10.1523/JNEUROSCI.0766-05.2005

Izsak, J., Seth, H., Andersson, M., Vizlin-Hodzic, D., Theiss, S., Hanse, E., . . . Illes, S. (2019). Robust Generation of Person-Specific, Synchronously Active Neuronal Networks Using Purely Isogenic Human iPSC-3D Neural Aggregate Cultures. *Front Neurosci, 13*, 351. doi:10.3389/fnins.2019.00351

J, H. J. (2013). *ELUCIDATING THE ROLE OF INSULIN-LIKE GROWTH FACTOR 2 MRNA BINDING PROTEIN 2 IN THE PATHOGENESIS OF TYPE 2 DIABETES.* (Doctorate of Philosophy in Pharmacology), Georgetown University, Washington, DC.

Jaishankar, S. (2021). *On molecular genetic aspects of juvenile myoclonic epilepsy.* (PhD), JNCASR.

Jallon, P., & Latour, P. (2005). Epidemiology of idiopathic generalized epilepsies. *Epilepsia, 46 Suppl 9*, 10-14. doi:10.1111/j.1528-1167.2005.00309.x

Janz, D. (1985). Epilepsy with impulsive petit mal (juvenile myoclonic epilepsy). *Acta Neurol Scand, 72*(5), 449-459. doi:10.1111/j.1600-0404.1985.tb00900.x

Janz, D., & Matthes, A. (1955). Propulsive petit mal epilepsy; clinical aspects and pathogenesis of the so-called akinetic seizures, nodding convulsions and Salaam convulsions. *Bibl Paediatr*(60), 5-60. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14389220

Jardine, S., Dhingani, N., & Muise, A. M. (2019). TTC7A: Steward of Intestinal Health. *Cell Mol Gastroenterol Hepatol, 7*(3), 555-570. doi:10.1016/j.jcmgh.2018.12.001

Jayalakshmi, S. S., Mohandas, S., Sailaja, S., & Borgohain, R. (2006). Clinical and electroencephalographic study of first-degree relatives

and probands with juvenile myoclonic epilepsy. *Seizure, 15*(3), 177-183. doi:10.1016/j.seizure.2005.12.011

Jiang, T., Hou, C. C., She, Z. Y., & Yang, W. X. (2013). The SOX gene family: function and regulation in testis determination and male fertility maintenance. *Mol Biol Rep, 40*(3), 2187-2194. doi:10.1007/s11033-012-2279-3

Jobe, P. C., Mishra, P. K., Ludvig, N., & Dailey, J. W. (1991). Scope and contribution of genetic models to an understanding of the epilepsies. *Crit Rev Neurobiol, 6*(3), 183-220. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1773452

Jones, B. L., & Smith, S. M. (2016). Calcium-Sensing Receptor: A Key Target for Extracellular Calcium Signaling in Neurons. *Front Physiol, 7*, 116. doi:10.3389/fphys.2016.00116

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., . . . Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature, 596*(7873), 583-589. doi:10.1038/s41586-021-03819-2

Jurczyk, A., Gromley, A., Redick, S., San Agustin, J., Witman, G., Pazour, G. J., . . . Doxsey, S. (2004). Pericentrin forms a complex with intraflagellar transport proteins and polycystin-2 and is required for primary cilia assembly. *J Cell Biol, 166*(5), 637-643. doi:10.1083/jcb.200405023

Kan, L., Jalali, A., Zhao, L. R., Zhou, X., McGuire, T., Kazanis, I., . . . Kessler, J. A. (2007). Dual function of Sox1 in telencephalic progenitor cells. *Dev Biol, 310*(1), 85-98. doi:10.1016/j.ydbio.2007.07.026

Kapoor, A., Satishchandra, P., Ratnapriya, R., Reddy, R., Kadandale, J., Shankar, S. K., & Anand, A. (2008). An idiopathic epilepsy syndrome linked to 3q13.3-q21 and missense mutations in the extracellular calcium sensing receptor gene. *Ann Neurol, 64*(2), 158-167. doi:10.1002/ana.21428

Kasteleijn-Nolst Trenite, D. G., Schmitz, B., Janz, D., Delgado-Escueta, A. V., Thomas, P., Hirsch, E., . . . Genton, P. (2013). Consensus on diagnosis and management of JME: From founder's observations to current trends. *Epilepsy Behav, 28 Suppl 1*, S87-90. doi:10.1016/j.yebeh.2012.11.051

Kazis, D., Petridis, F., Chatzikonstantinou, S., Karantali, E., Jamali, R., Chowdhury, R., . . . Mavroudis, I. (2021). Gray Matter Changes in Juvenile Myoclonic Epilepsy. A Voxel-Wise Meta-Analysis. *Medicina (Kaunas), 57*(11). doi:10.3390/medicina57111136

Kilburn, C. L., Pearson, C. G., Romijn, E. P., Meehl, J. B., Giddings, T. H., Jr., Culver, B. P., . . . Winey, M. (2007). New Tetrahymena basal body protein components identify basal body domain structure. *J Cell Biol, 178*(6), 905-912. doi:10.1083/jcb.200703109

Kim, J. H. (2017). Grey and White Matter Alterations in Juvenile Myoclonic Epilepsy: A Comprehensive Review. *J Epilepsy Res, 7*(2), 77-88. doi:10.14581/jer.17013

Kim, J. H., Lee, J. K., Koh, S. B., Lee, S. A., Lee, J. M., Kim, S. I., & Kang, J. K. (2007). Regional grey matter abnormalities in juvenile myoclonic epilepsy: a voxel-based morphometry study. *Neuroimage, 37*(4), 1132-1137. doi:10.1016/j.neuroimage.2007.06.025

Kim, W. K., Yun, S., Kwon, Y., You, K. T., Shin, N., Kim, J., & Kim, H. (2017). mRNAs containing NMD-competent premature termination codons are stabilized and translated under UPF1 depletion. *Sci Rep, 7*(1), 15833. doi:10.1038/s41598-017-16177-9

King, S. M. (2006). Axonemal protofilament ribbons, DM10 domains, and the link to juvenile myoclonic epilepsy. *Cell Motil Cytoskeleton, 63*(5), 245-253. doi:10.1002/cm.20129

Klassen, T., Davis, C., Goldman, A., Burgess, D., Chen, T., Wheeler, D., . . . Noebels, J. (2011). Exome sequencing of ion channel genes reveals complex profiles confounding personal risk assessment in epilepsy. *Cell, 145*(7), 1036-1048. doi:10.1016/j.cell.2011.05.025

Kokubu, H., & Lim, J. (2014). X-gal Staining on Adult Mouse Brain Sections. *Bio Protoc, 4*(5). doi:10.21769/bioprotoc.1064

Kolaskar, A. S., & Tongaonkar, P. C. (1990). A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett, 276*(1-2), 172-174. doi:10.1016/0014-5793(90)80535-q

Konno, D., Shioi, G., Shitamukai, A., Mori, A., Kiyonari, H., Miyata, T., & Matsuzaki, F. (2008). Neuroepithelial progenitors undergo LGN-dependent planar divisions to maintain self-renewability during mammalian neurogenesis. *Nat Cell Biol, 10*(1), 93-101. doi:10.1038/ncb1673

Korkolopoulou, P., Levidou, G., El-Habr, E. A., Adamopoulos, C., Fragkou, P., Boviatsis, E., . . . Piperi, C. (2013). Sox11 expression in astrocytic gliomas: correlation with nestin/c-Met/IDH1-R132H expression phenotypes, p-Stat-3 and survival. *Br J Cancer, 108*(10), 2142-2152. doi:10.1038/bjc.2013.176

Kuebler, D., & Tanouye, M. A. (2000). Modifications of seizure susceptibility in Drosophila. *J Neurophysiol, 83*(2), 998-1009. doi:10.1152/jn.2000.83.2.998

Kuebler, D., Zhang, H., Ren, X., & Tanouye, M. A. (2001). Genetic suppression of seizure susceptibility in Drosophila. *J Neurophysiol, 86*(3), 1211-1225. doi:10.1152/jn.2001.86.3.1211

Kuijpers, M., & Hoogenraad, C. C. (2011). Centrosomes, microtubules and neuronal development. *Mol Cell Neurosci, 48*(4), 349-358. doi:10.1016/j.mcn.2011.05.004

Kumar, M., Gouw, M., Michael, S., Samano-Sanchez, H., Pancsa, R., Glavina, J., . . . Gibson, T. J. (2020). ELM-the eukaryotic linear motif resource in 2020. *Nucleic Acids Res, 48*(D1), D296-D306. doi:10.1093/nar/gkz1030

Kumar, P., Henikoff, S., & Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc, 4*(7), 1073-1081. doi:nprot.2009.86 [pii]

10.1038/nprot.2009.86

Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J Mol Biol, 157*(1), 105-132. doi:10.1016/0022-2836(82)90515-0

Kyttala, A., Moraghebi, R., Valensisi, C., Kettunen, J., Andrus, C., Pasumarthy, K. K., . . . Trokovic, R. (2016). Genetic Variability Overrides the Impact of Parental Cell Type and Determines iPSC Differentiation Potential. *Stem Cell Reports, 6*(2), 200-212. doi:10.1016/j.stemcr.2015.12.009

Lachance-Touchette, P., Choudhury, M., Stoica, A., Di Cristo, G., & Cossette, P. (2014). Single-cell genetic expression of mutant GABAA receptors causing Human genetic epilepsy alters dendritic spine and GABAergic bouton formation in a mutation-specific manner. *Front Cell Neurosci, 8*, 317. doi:10.3389/fncel.2014.00317

Lahiry, P., Wang, J., Robinson, J. F., Turowec, J. P., Litchfield, D. W., Lanktree, M. B., . . . Hegele, R. A. (2009). A multiplex human

syndrome implicates a key role for intestinal cell kinase in development of central nervous, skeletal, and endocrine systems. *Am J Hum Genet, 84*(2), 134-147. doi:10.1016/j.ajhg.2008.12.017

Lai, C. H., Chou, C. Y., Ch'ang, L. Y., Liu, C. S., & Lin, W. (2000). Identification of novel human genes evolutionarily conserved in Caenorhabditis elegans by comparative proteomics. *Genome Res, 10*(5), 703-713. doi:10.1101/gr.10.5.703

Lal, D., Reinthaler, E. M., Dejanovic, B., May, P., Thiele, H., Lehesjoki, A. E., . . . Neubauer, B. A. (2016). Evaluation of Presumably Disease Causing SCN1A Variants in a Cohort of Common Epilepsy Syndromes. *PLoS ONE, 11*(3), e0150426. doi:10.1371/journal.pone.0150426

Lau, E. O., Damiani, D., Chehade, G., Ruiz-Reig, N., Saade, R., Jossin, Y., . . . Tissir, F. (2021). DIAPH3 deficiency links microtubules to mitotic errors, defective neurogenesis, and brain dysfunction. *Elife, 10*. doi:10.7554/eLife.61974

Le Hellard, S., Neidhart, E., Thomas, P., Feingold, J., Malafosse, A., & Tafti, M. (1999). Lack of association between juvenile myoclonic epilepsy and HLA-DR13. *Epilepsia, 40*(1), 117-119. doi:10.1111/j.1528-1157.1999.tb01999.x

Lee, J., & Wu, C. F. (2006). Genetic modifications of seizure susceptibility and expression by altered excitability in Drosophila Na(+) and K(+) channel mutants. *J Neurophysiol, 96*(5), 2465-2478. doi:10.1152/jn.00499.2006

Lee, Y., Lee, K. J., Jang, J. W., Lee, S. I., & Kim, S. (2020). An EEG system to detect brain signals from multiple adult zebrafish. *Biosens Bioelectron, 164*, 112315. doi:10.1016/j.bios.2020.112315

Lee, Y., Seo, H. W., Lee, K. J., Jang, J. W., & Kim, S. (2020). A Microfluidic System for Stable and Continuous EEG Monitoring from Multiple Larval Zebrafish. *Sensors (Basel), 20*(20). doi:10.3390/s20205903

Lenzen, K. P., Heils, A., Lorenz, S., Hempelmann, A., Hofels, S., Lohoff, F. W., . . . Sander, T. (2005). Supportive evidence for an allelic association of the human KCNJ10 potassium channel gene with idiopathic generalized epilepsy. *Epilepsy Res, 63*(2-3), 113-118. doi:10.1016/j.eplepsyres.2005.01.002

Lerche, H., Berkovic, S. F., Lowenstein, D. H., Euro, E.-C. C., Epi, P. G. X. C., & Epi, K. C. E. P. G. P. (2019). Intestinal-Cell Kinase and

Juvenile Myoclonic Epilepsy. *N Engl J Med, 380*(16), e24. doi:10.1056/NEJMc1805100

Leu, C., Stevelink, R., Smith, A. W., Goleva, S. B., Kanai, M., Ferguson, L., . . . Lal, D. (2019). Polygenic burden in focal and generalized epilepsies. *Brain, 142*(11), 3473-3481. doi:10.1093/brain/awz292

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics, 25*(14), 1754-1760. doi:10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics, 25*(16), 2078-2079. doi:10.1093/bioinformatics/btp352

Li, L., Yang, W. T., Zheng, P. S., & Liu, X. F. (2018). SOX17 restrains proliferation and tumor formation by down-regulating activity of the Wnt/beta-catenin signaling pathway via trans-suppressing beta-catenin in cervical cancer. *Cell Death Dis, 9*(7), 741. doi:10.1038/s41419-018-0782-8

Lin, D., Fu, Z., Yang, G., Gao, D., Wang, T., Liu, Z., . . . Wang, Y. (2020). Exportin-5 SUMOylation promotes hepatocellular carcinoma progression. *Exp Cell Res, 395*(2), 112219. doi:10.1016/j.yexcr.2020.112219

Lin, K., Carrete, H., Jr., Lin, J., Peruchi, M. M., de Araujo Filho, G. M., Guaranha, M. S., . . . Yacubian, E. M. (2009). Magnetic resonance spectroscopy reveals an epileptic network in juvenile myoclonic epilepsy. *Epilepsia, 50*(5), 1191-1200. doi:10.1111/j.1528-1167.2008.01948.x

Linck, R., Fu, X., Lin, J., Ouch, C., Schefter, A., Steffen, W., . . . Nicastro, D. (2014). Insights into the structure and function of ciliary and flagellar doublet microtubules: tektins, Ca2+-binding proteins, and stable protofilaments. *J Biol Chem, 289*(25), 17427-17444. doi:M114.568949 [pii]

10.1074/jbc.M114.568949

Loscher, W. (2017). Animal Models of Seizures and Epilepsy: Past, Present, and Future Role for the Discovery of Antiseizure Drugs. *Neurochem Res, 42*(7), 1873-1888. doi:10.1007/s11064-017-2222-z

Loucks, C. M., Park, K., Walker, D. S., McEwan, A. H., Timbers, T. A., Ardiel, E. L., . . . Leroux, M. R. (2019). EFHC1, implicated in juvenile

myoclonic epilepsy, functions at the cilium and synapse to modulate dopamine signaling. *Elife, 8*. doi:10.7554/eLife.37271

Loughman, A., Bowden, S. C., & D'Souza, W. (2014). Cognitive functioning in idiopathic generalised epilepsies: a systematic review and meta-analysis. *Neurosci Biobehav Rev, 43*, 20-34. doi:10.1016/j.neubiorev.2014.02.012

Macdonald, R. L., Gallagher, M. J., Feng, H. J., & Kang, J. (2004). GABA(A) receptor epilepsy mutations. *Biochem Pharmacol, 68*(8), 1497-1506. doi:10.1016/j.bcp.2004.07.029

Maguire, J., & Mody, I. (2008). GABA(A)R plasticity during pregnancy: relevance to postpartum depression. *Neuron, 59*(2), 207-213. doi:10.1016/j.neuron.2008.06.019

Manganotti, P., Bongiovanni, L. G., Fuggetta, G., Zanette, G., & Fiaschi, A. (2006). Effects of sleep deprivation on cortical excitability in patients affected by juvenile myoclonic epilepsy: a combined transcranial magnetic stimulation and EEG study. *J Neurol Neurosurg Psychiatry, 77*(1), 56-60. doi:10.1136/jnnp.2004.041137

Marchler-Bauer, A., Zheng, C., Chitsaz, F., Derbyshire, M. K., Geer, L. Y., Geer, R. C., . . . Bryant, S. H. (2013). CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res, 41*(Database issue), D348-352. doi:10.1093/nar/gks1243

Marini, C., Scheffer, I. E., Nabbout, R., Suls, A., De Jonghe, P., Zara, F., & Guerrini, R. (2011). The genetics of Dravet syndrome. *Epilepsia, 52 Suppl 2*, 24-29. doi:10.1111/j.1528-1167.2011.02997.x

Marthiens, V., & Basto, R. (2020). Centrosomes: The good and the bad for brain development. *Biol Cell, 112*(6), 153-172. doi:10.1111/boc.201900090

Mas, C., Taske, N., Deutsch, S., Guipponi, M., Thomas, P., Covanis, A., . . . Meda, P. (2004). Association of the connexin36 gene with juvenile myoclonic epilepsy. *J Med Genet, 41*(7), e93. doi:10.1136/jmg.2003.017954

Matsuoka, H., Nakamura, M., Ohno, T., Shimabukuro, J., Suzuki, T., Numachi, Y., & Awata, S. (2005). The role of cognitive-motor function in precipitation and inhibition of epileptic seizures. *Epilepsia, 46 Suppl 1*, 17-20. doi:10.1111/j.0013-9580.2005.461006.x

Mattheisen, G. B., Tsintsadze, T., & Smith, S. M. (2018). Strong G-Protein-Mediated Inhibition of Sodium Channels. *Cell Rep, 23*(9), 2770-2781. doi:10.1016/j.celrep.2018.04.109

May, P., Girard, S., Harrer, M., Bobbili, D. R., Schubert, J., Wolking, S., . . . Epi, P. G. X. C. (2018). Rare coding variants in genes encoding GABAA receptors in genetic generalised epilepsies: an exome-based case-control study. *Lancet Neurol, 17*(8), 699-708. doi:10.1016/S1474-4422(18)30215-1

McClure-Begley, T. D., & Klymkowsky, M. W. (2017). Nuclear roles for cilia-associated proteins. *Cilia, 6*, 8. doi:10.1186/s13630-017-0052-x

McConnell, S. K., & Kaznowski, C. E. (1991). Cell cycle dependence of laminar determination in developing neocortex. *Science, 254*(5029), 282-285. doi:10.1126/science.254.5029.282

McGuire, M. J., Gertz, S. M., McCutcheon, J. D., Richardson, C. R., & Poulsen, D. J. (2019). Use of a Wireless Video-EEG System to Monitor Epileptiform Discharges Following Lateral Fluid-Percussion Induced Traumatic Brain Injury. *J Vis Exp*(148). doi:10.3791/59637

Medlej, Y., Asdikian, R., Wadi, L., Salah, H., Dosh, L., Hashash, R., . . . Obeid, M. (2019). Enhanced setup for wired continuous long-term EEG monitoring in juvenile and adult rats: application for epilepsy and other disorders. *BMC Neurosci, 20*(1), 8. doi:10.1186/s12868-019-0490-z

Mefford, H. C. (2015). Copy Number Matters in Epilepsy. *Epilepsy Curr, 15*(4), 180-182. doi:10.5698/1535-7511-15.4.180

Mefford, H. C., Muhle, H., Ostertag, P., von Spiczak, S., Buysse, K., Baker, C., . . . Eichler, E. E. (2010). Genome-wide copy number variation in epilepsy: novel susceptibility loci in idiopathic generalized and focal epilepsies. *PLoS Genet, 6*(5), e1000962. doi:10.1371/journal.pgen.1000962

Mehta, G. A., Khanna, P., & Gatza, M. L. (2019). Emerging Role of SOX Proteins in Breast Cancer Development and Maintenance. *J Mammary Gland Biol Neoplasia, 24*(3), 213-230. doi:10.1007/s10911-019-09430-6

Michaelis, M., Sobczak, A., & Weitzel, J. M. (2014). In vivo microinjection and electroporation of mouse testis. *J Vis Exp*(90). doi:10.3791/51802

Mijimolle, N., Velasco, J., Dubus, P., Guerra, C., Weinbaum, C. A., Casey, P. J., . . . Barbacid, M. (2005). Protein farnesyltransferase in embryogenesis, adult homeostasis, and tumor development. *Cancer Cell, 7*(4), 313-324. doi:10.1016/j.ccr.2005.03.004

Moon, H., Song, J., Shin, J. O., Lee, H., Kim, H. K., Eggenschwiller, J. T., . . . Ko, H. W. (2014). Intestinal cell kinase, a protein associated with endocrine-cerebro-osteodysplasia syndrome, is a key regulator of cilia length and Hedgehog signaling. *Proc Natl Acad Sci U S A, 111*(23), 8541-8546. doi:10.1073/pnas.1323161111

Morris, G., Rowell, R., & Cunningham, M. (2021). Limitations of animal epilepsy research models: Can epileptic human tissue provide translational benefit? *ALTEX, 38*(3), 451-462. doi:10.14573/altex.2007082

Moulard, B., Guipponi, M., Chaigne, D., Mouthon, D., Buresi, C., & Malafosse, A. (1999). Identification of a new locus for generalized epilepsy with febrile seizures plus (GEFS+) on chromosome 2q24-q33. *Am J Hum Genet, 65*(5), 1396-1400. doi:10.1086/302621

Mouse Genome Sequencing, C., Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., . . . Lander, E. S. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature, 420*(6915), 520-562. doi:10.1038/nature01262

Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., . . . Yamanaka, S. (2008). Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat Biotechnol, 26*(1), 101-106. doi:10.1038/nbt1374

Nakamura, K., Noguchi, T., Takahara, M., Omori, Y., Furukawa, T., Katoh, Y., & Nakayama, K. (2020). Anterograde trafficking of ciliary MAP kinase-like ICK/CILK1 by the intraflagellar transport machinery is required for intraciliary retrograde protein trafficking. *J Biol Chem, 295*(38), 13363-13376. doi:10.1074/jbc.RA120.014142

Nicastro, D., Fu, X., Heuser, T., Tso, A., Porter, M. E., & Linck, R. W. (2011). Cryo-electron tomography reveals conserved features of doublet microtubules in flagella. *Proc Natl Acad Sci U S A, 108*(42), E845-853. doi:10.1073/pnas.1106178108

Niu, W., & Parent, J. M. (2020). Modeling genetic epilepsies in a dish. *Dev Dyn, 249*(1), 56-75. doi:10.1002/dvdy.79

Noebels, J. L. (2002). A Gene for JME at Last: The alpha1 GABA Receptor Subunit. *Epilepsy Curr, 2*(4), 131-132. doi:10.1046/j.1535-7597.2002.t01-1-00050.x

Oakley, J. C., Kalume, F., Yu, F. H., Scheuer, T., & Catterall, W. A. (2009). Temperature- and age-dependent seizures in a mouse model of severe myoclonic epilepsy in infancy. *Proc Natl Acad Sci U S A, 106*(10), 3994-3999. doi:10.1073/pnas.0813330106

Obeid, T., el Rab, M. O., Daif, A. K., Panayiotopoulos, C. P., Halim, K., Bahakim, H., & Bamgboye, E. (1994). Is HLA-DRW13 (W6) associated with juvenile myoclonic epilepsy in Arab patients? *Epilepsia, 35*(2), 319-321. doi:10.1111/j.1528-1157.1994.tb02437.x

Ohmori, I., Ouchida, M., Miki, T., Mimaki, N., Kiyonaka, S., Nishiki, T., . . . Matsui, H. (2008). A CACNB4 mutation shows that altered Ca(v)2.1 function may be a genetic modifier of severe myoclonic epilepsy in infancy. *Neurobiol Dis, 32*(3), 349-354. doi:10.1016/j.nbd.2008.07.017

Okamoto, N., Miya, F., Tsunoda, T., Kato, M., Saitoh, S., Yamasaki, M., . . . Kosaki, K. (2015). Targeted next-generation sequencing in the diagnosis of neurodevelopmental disorders. *Clin Genet, 88*(3), 288-292. doi:10.1111/cge.12492

Osaki, E., Nishina, Y., Inazawa, J., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., . . . Semba, K. (1999). Identification of a novel Sry-related gene and its germ cell-specific expression. *Nucleic Acids Res, 27*(12), 2503-2510. doi:10.1093/nar/27.12.2503

Oud, M. M., Bonnard, C., Mans, D. A., Altunoglu, U., Tohari, S., Ng, A. Y. J., . . . Arts, H. H. (2016). A novel ICK mutation causes ciliary disruption and lethal endocrine-cerebro-osteodysplasia syndrome. *Cilia, 5*, 8. doi:10.1186/s13630-016-0029-1

Oud, M. M., Lamers, I. J., & Arts, H. H. (2017). Ciliopathies: Genetics in Pediatric Medicine. *J Pediatr Genet, 6*(1), 18-29. doi:10.1055/s-0036-1593841

Ougland, R., Jonson, I., Moen, M. N., Nesse, G., Asker, G., Klungland, A., & Larsen, E. (2016). Role of ALKBH1 in the Core Transcriptional Network of Embryonic Stem Cells. *Cell Physiol Biochem, 38*(1), 173-184. doi:10.1159/000438619

Ozdas, S., Canatar, I., & Ozdas, T. (2021). Effects of Knockdown of XPO5 by siRNA on the Biological Behavior of Head and Neck Cancer Cells. *Laryngoscope*. doi:10.1002/lary.29787

Paige Taylor, S., Kunova Bosakova, M., Varecha, M., Balek, L., Barta, T., Trantirek, L., . . . Krejci, P. (2016). An inactivating mutation in intestinal cell kinase, ICK, impairs hedgehog signalling and causes short rib-polydactyly syndrome. *Hum Mol Genet, 25*(18), 3998-4011. doi:10.1093/hmg/ddw240

Pal, D. K., Evgrafov, O. V., Tabares, P., Zhang, F., Durner, M., & Greenberg, D. A. (2003). BRD2 (RING3) is a probable major susceptibility gene for common juvenile myoclonic epilepsy. *Am J Hum Genet, 73*(2), 261-270. doi:10.1086/377006

Panayiotopoulos, C. P., Obeid, T., & Tahan, A. R. (1994). Juvenile myoclonic epilepsy: a 5-year prospective study. *Epilepsia, 35*(2), 285-296. doi:10.1111/j.1528-1157.1994.tb02432.x

Papatheodorou, I., Moreno, P., Manning, J., Fuentes, A. M., George, N., Fexova, S., . . . Brazma, A. (2020). Expression Atlas update: from tissues to single cells. *Nucleic Acids Res, 48*(D1), D77-D83. doi:10.1093/nar/gkz947

Parker, L., Padilla, M., Du, Y., Dong, K., & Tanouye, M. A. (2011). Drosophila as a model for epilepsy: bss is a gain-of-function mutation in the para sodium channel gene that leads to seizures. *Genetics, 187*(2), 523-534. doi:10.1534/genetics.110.123299

Pascalicchio, T. F., de Araujo Filho, G. M., da Silva Noffs, M. H., Lin, K., Caboclo, L. O., Vidal-Dourado, M., . . . Yacubian, E. M. (2007). Neuropsychological profile of patients with juvenile myoclonic epilepsy: a controlled study of 50 patients. *Epilepsy Behav, 10*(2), 263-267. doi:10.1016/j.yebeh.2006.11.012

Pejaver, V., Urresti, J., Lugo-Martinez, J., Pagel, K. A., Lin, G. N., Nam, H. J., . . . Radivojac, P. (2020). Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nat Commun, 11*(1), 5918. doi:10.1038/s41467-020-19669-x

Peljto, A. L., Barker-Cummings, C., Vasoli, V. M., Leibson, C. L., Hauser, W. A., Buchhalter, J. R., & Ottman, R. (2014). Familial risk of epilepsy: a population-based study. *Brain, 137*(Pt 3), 795-805. doi:10.1093/brain/awt368

Petersen, B. S., Fredrich, B., Hoeppner, M. P., Ellinghaus, D., & Franke, A. (2017). Opportunities and challenges of whole-genome and -exome sequencing. *BMC Genet, 18*(1), 14. doi:10.1186/s12863-017-0479-5

Pingault, V., Bodereau, V., Baral, V., Marcos, S., Watanabe, Y., Chaoui, A., . . . Bondurand, N. (2013). Loss-of-function mutations in SOX10 cause Kallmann syndrome with deafness. *Am J Hum Genet, 92*(5), 707-724. doi:10.1016/j.ajhg.2013.03.024

Pitkanen, A., Loscher, W., Vezzani, A., Becker, A. J., Simonato, M., Lukasiuk, K., . . . Beck, H. (2016). Advances in the development of biomarkers for epilepsy. *Lancet Neurol, 15*(8), 843-856. doi:10.1016/S1474-4422(16)00112-5

Portnoi, M. F., Dumargne, M. C., Rojo, S., Witchel, S. F., Duncan, A. J., Eozenou, C., . . . Bashamboo, A. (2018). Mutations involving the SRY-related gene SOX8 are associated with a spectrum of human reproductive anomalies. *Hum Mol Genet, 27*(7), 1228-1240. doi:10.1093/hmg/ddy037

Quadrato, G., Nguyen, T., Macosko, E. Z., Sherwood, J. L., Min Yang, S., Berger, D. R., . . . Arlotta, P. (2017). Cell diversity and network dynamics in photosensitive human brain organoids. *Nature, 545*(7652), 48-53. doi:10.1038/nature22047

Quraishi, I. H., Stern, S., Mangan, K. P., Zhang, Y., Ali, S. R., Mercier, M. R., . . . Kaczmarek, L. K. (2019). An Epilepsy-Associated KCNT1 Mutation Enhances Excitability of Human iPSC-Derived Neurons by Increasing Slack KNa Currents. *J Neurosci, 39*(37), 7438-7449. doi:10.1523/JNEUROSCI.1628-18.2019

Raju, P. (2014). *Connecting the paralogs: Contribution of EFHC1 and EFHC2 in juvenile myoclonic epilepsy.* (Ph.D.), Jawaharlal Nether Centre for Advanced Scientific Research, Bangalore.

Ratnapriya, R. (2009). *Identification of genetic loci for human epilepsies.* (Ph.D.), JNCASR, Bangalore.

Rauch, A., Thiel, C. T., Schindler, D., Wick, U., Crow, Y. J., Ekici, A. B., . . . Reis, A. (2008). Mutations in the pericentrin (PCNT) gene cause primordial dwarfism. *Science, 319*(5864), 816-819. doi:10.1126/science.1151174

Renganathan, R., & Delanty, N. (2003). Juvenile myoclonic epilepsy: under-appreciated and under-diagnosed. *Postgrad Med J, 79*(928), 78-80. doi:10.1136/pmj.79.928.78

Revinski, D. R., Zaragosi, L. E., Boutin, C., Ruiz-Garcia, S., Deprez, M., Thome, V., . . . Barbry, P. (2018). CDC20B is required for deuterosome-mediated centriole production in multiciliated cells. *Nat Commun, 9*(1), 4668. doi:10.1038/s41467-018-06768-z

Reynolds, J. R. (1862). Epilepsy: its symptoms, treatment, and relation to other chronic, convulsive diseases. *American Journal of Psychiatry, 19(2)*, 198-209.

Riban, V., Bouilleret, V., Pham-Le, B. T., Fritschy, J. M., Marescaux, C., & Depaulis, A. (2002). Evolution of hippocampal epileptic activity during the development of hippocampal sclerosis in a mouse model of temporal lobe epilepsy. *Neuroscience, 112*(1), 101-111. doi:10.1016/s0306-4522(02)00064-7

Roebling, R., Scheerer, N., Uttner, I., Gruber, O., Kraft, E., & Lerche, H. (2009). Evaluation of cognition, structural, and functional MRI in juvenile myoclonic epilepsy. *Epilepsia, 50*(11), 2456-2465. doi:10.1111/j.1528-1167.2009.02127.x

Rossetto, M. G., Zanarella, E., Orso, G., Scorzeto, M., Megighian, A., Kumar, V., . . . Daga, A. (2011). Defhc1.1, a homologue of the juvenile myoclonic gene EFHC1, modulates architecture and basal activity of the neuromuscular junction in Drosophila. *Hum Mol Genet, 20*(21), 4248-4257. doi:10.1093/hmg/ddr352

Roubin, R., Acquaviva, C., Chevrier, V., Sedjai, F., Zyss, D., Birnbaum, D., & Rosnet, O. (2013). Myomegalin is necessary for the formation of centrosomal and Golgi-derived microtubules. *Biol Open, 2*(2), 238-250. doi:10.1242/bio.20123392

Ruden, J. B., Dugan, L. L., & Konradi, C. (2021). Parvalbumin interneuron vulnerability and brain disorders. *Neuropsychopharmacology, 46*(2), 279-287. doi:10.1038/s41386-020-0778-9

Salisbury, J. L., Suino, K. M., Busby, R., & Springett, M. (2002). Centrin-2 is required for centriole duplication in mammalian cells. *Curr Biol, 12*(15), 1287-1292. doi:10.1016/s0960-9822(02)01019-9

Salvati, K. A., Mason, A. J., Gailey, C. D., Wang, E. J., Fu, Z., & Beenhakker, M. P. (2021). Mice Harboring a Non-Functional CILK1/ICK Allele Fail to Model the Epileptic Phenotype in Patients

Carrying Variant CILK1/ICK. *Int J Mol Sci, 22*(16). doi:10.3390/ijms22168875

Samarut, E., Dona, M., & Yogananthar jah, P. (2018). Zebrafish Ride the Danube's Wave for the 10th European Zebrafish Meeting. *Zebrafish, 15*(1), 79-83. doi:10.1089/zeb.2017.1518

Samarut, E., Swaminathan, A., Riche, R., Liao, M., Hassan-Abdi, R., Renault, S., . . . Drapeau, P. (2018). gamma-Aminobutyric acid receptor alpha 1 subunit loss of function causes genetic generalized epilepsy by impairing inhibitory network neurodevelopment. *Epilepsia, 59*(11), 2061-2074. doi:10.1111/epi.14576

Sander, T., Bockenkamp, B., Hildmann, T., Blasczyk, R., Kretz, R., Wienker, T. F., . . . Ziegler, A. (1997). Refined mapping of the epilepsy susceptibility locus EJM1 on chromosome 6. *Neurology, 49*(3), 842-847. doi:10.1212/wnl.49.3.842

Schanze, I., Schanze, D., Bacino, C. A., Douzgou, S., Kerr, B., & Zenker, M. (2013). Haploinsufficiency of SOX5, a member of the SOX (SRY-related HMG-box) family of transcription factors is a cause of intellectual disability. *Eur J Med Genet, 56*(2), 108-113. doi:10.1016/j.ejmg.2012.11.001

Schauwecker, P. E. (2011). The relevance of individual genetic background and its role in animal models of epilepsy. *Epilepsy Res, 97*(1-2), 1-11. doi:10.1016/j.eplepsyres.2011.09.005

Scheffer, I. E., Berkovic, S., Capovilla, G., Connolly, M. B., French, J., Guilhoto, L., . . . Zuberi, S. M. (2017). ILAE classification of the epilepsies: Position paper of the ILAE Commission for Classification and Terminology. *Epilepsia, 58*(4), 512-521. doi:10.1111/epi.13709

Schepers, G. E., Teasdale, R. D., & Koopman, P. (2002). Twenty pairs of sox: extent, homology, and nomenclature of the mouse and human sox transcription factor gene families. *Dev Cell, 3*(2), 167-170. doi:10.1016/s1534-5807(02)00223-x

Schwarz, J. M., Rodelsperger, C., Schuelke, M., & Seelow, D. (2010). MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods, 7*(8), 575-576. doi:nmeth0810-575 [pii]

10.1038/nmeth0810-575

Seneviratne, U., Cook, M., & D'Souza, W. (2012). The prognosis of idiopathic generalized epilepsy. *Epilepsia, 53*(12), 2079-2090. doi:10.1111/j.1528-1167.2012.03723.x

Sesarini, C. V., Costa, L., Granana, N., Coto, M. G., Pallia, R. C., & Argibay, P. F. (2015). Association between GABA(A) receptor subunit polymorphisms and autism spectrum disorder (ASD). *Psychiatry Res, 229*(1-2), 580-582. doi:10.1016/j.psychres.2015.07.077

Sharafian, S., Alimadadi, H., Shahrooei, M., Gharagozlou, M., Aghamohammadi, A., & Parvaneh, N. (2018). A Novel TTC7A Deficiency Presenting With Combined Immunodeficiency and Chronic Gastrointestinal Problems. *J Investig Allergol Clin Immunol, 28*(5), 358-360. doi:10.18176/jiaci.0290

Sharp, A. J., Mefford, H. C., Li, K., Baker, C., Skinner, C., Stevenson, R. E., . . . Eichler, E. E. (2008). A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. *Nat Genet, 40*(3), 322-328. doi:10.1038/ng.93

Shorvon, S. D. (2011). The causes of epilepsy: changing concepts of etiology of epilepsy over the past 150 years. *Epilepsia, 52*(6), 1033-1044. doi:10.1111/j.1528-1167.2011.03051.x

Singh, B., Monteil, A., Bidaud, I., Sugimoto, Y., Suzuki, T., Hamano, S., . . . Yamakawa, K. (2007). Mutational analysis of CACNA1G in idiopathic generalized epilepsy. Mutation in brief #962. Online. *Hum Mutat, 28*(5), 524-525. doi:10.1002/humu.9491

Singh, N. A., Charlier, C., Stauffer, D., DuPont, B. R., Leach, R. J., Melis, R., . . . Leppert, M. (1998). A novel potassium channel gene, KCNQ2, is mutated in an inherited epilepsy of newborns. *Nat Genet, 18*(1), 25-29. doi:10.1038/ng0198-25

Smart, S. L., Lopantsev, V., Zhang, C. L., Robbins, C. A., Wang, H., Chiu, S. Y., . . . Tempel, B. L. (1998). Deletion of the K(V)1.1 potassium channel causes epilepsy in mice. *Neuron, 20*(4), 809-819. doi:10.1016/s0896-6273(00)81018-1

Somayajula, S., Vooturi, S., & Jayalakshmi, S. (2015). Psychiatric disorders among 165 patients with juvenile myoclonic epilepsy in India and association with clinical and sociodemographic variables. *Epilepsy Behav, 53*, 37-42. doi:10.1016/j.yebeh.2015.09.024

Song, J., Hu, J., & Tanouye, M. (2007). Seizure suppression by top1 mutations in Drosophila. *J Neurosci, 27*(11), 2927-2937. doi:10.1523/JNEUROSCI.3944-06.2007

Song, L., Liu, D., He, J., Wang, X., Dai, Z., Zhao, Y., . . . Wang, B. (2016). SOX1 inhibits breast cancer cell growth and invasion through

suppressing the Wnt/beta-catenin signaling pathway. *APMIS, 124*(7), 547-555. doi:10.1111/apm.12543

Sousa, A. M. M., Meyer, K. A., Santpere, G., Gulden, F. O., & Sestan, N. (2017). Evolution of the Human Nervous System Function, Structure, and Development. *Cell, 170*(2), 226-247. doi:10.1016/j.cell.2017.06.036

Startin, C. M., Fiorentini, C., de Haan, M., & Skuse, D. H. (2015). Variation in the X-linked EFHC2 gene is associated with social cognitive abilities in males. *PLoS ONE, 10*(6), e0131604. doi:10.1371/journal.pone.0131604

Stephen, L. J., & Brodie, M. J. (2020). Pharmacological Management of the Genetic Generalised Epilepsies in Adolescents and Adults. *CNS Drugs, 34*(2), 147-161. doi:10.1007/s40263-020-00698-5

Stilwell, G. E., Saraswati, S., Littleton, J. T., & Chouinard, S. W. (2006). Development of a Drosophila seizure model for in vivo high-throughput drug screening. *Eur J Neurosci, 24*(8), 2211-2222. doi:10.1111/j.1460-9568.2006.05075.x

Stoddard, D., Zhao, Y., Bayless, B. A., Gui, L., Louka, P., Dave, D., . . . Nicastro, D. (2018). Tetrahymena RIB72A and RIB72B are microtubule inner proteins in the ciliary doublet microtubules. *Mol Biol Cell, 29*(21), 2566-2577. doi:10.1091/mbc.E18-06-0405

Stovall, D. B., Wan, M., Miller, L. D., Cao, P., Maglic, D., Zhang, Q., . . . Sui, G. (2013). The regulation of SOX7 and its tumor suppressive role in breast cancer. *Am J Pathol, 183*(5), 1645-1653. doi:10.1016/j.ajpath.2013.07.025

Strand, A. D., Aragaki, A. K., Baquet, Z. C., Hodges, A., Cunningham, P., Holmans, P., . . . Olson, J. M. (2007). Conservation of regional gene expression in mouse and human brain. *PLoS Genet, 3*(4), e59. doi:10.1371/journal.pgen.0030059

Sun, X. (2009). *Transcriptional networks of lung airway epithelial ciliogenesis.* (Master of Science), The university of New Mexico, New Mexico.

Sun, Y., Pasca, S. P., Portmann, T., Goold, C., Worringer, K. A., Guan, W., . . . Dolmetsch, R. E. (2016). A deleterious Nav1.1 mutation selectively impairs telencephalic inhibitory neurons derived from Dravet Syndrome patients. *Elife, 5*. doi:10.7554/eLife.13073

Suzuki, T., Delgado-Escueta, A. V., Aguan, K., Alonso, M. E., Shi, J., Hara, Y., . . . Yamakawa, K. (2004). Mutations in EFHC1 cause juvenile myoclonic epilepsy. *Nat Genet, 36*(8), 842-849. doi:10.1038/ng1393

ng1393 [pii]

Suzuki, T., Inoue, I., Yamagata, T., Morita, N., Furuichi, T., & Yamakawa, K. (2008). Sequential expression of Efhc1/myoclonin1 in choroid plexus and ependymal cell cilia. *Biochem Biophys Res Commun, 367*(1), 226-233. doi:10.1016/j.bbrc.2007.12.126

Suzuki, T., Miyamoto, H., Nakahari, T., Inoue, I., Suemoto, T., Jiang, B., . . . Yamakawa, K. (2009). Efhc1 deficiency causes spontaneous myoclonus and increased seizure susceptibility. *Hum Mol Genet, 18*(6), 1099-1109. doi:10.1093/hmg/ddp006

Swartz, B. E., Simpkins, F., Halgren, E., Mandelkern, M., Brown, C., Krisdakumtorn, T., & Gee, M. (1996). Visual working memory in primary generalized epilepsy: an 18FDG-PET study. *Neurology, 47*(5), 1203-1212. doi:10.1212/wnl.47.5.1203

Tae, W. S., Hong, S. B., Joo, E. Y., Han, S. J., Cho, J. W., Seo, D. W., . . . Kim, S. I. (2006). Structural brain abnormalities in juvenile myoclonic epilepsy patients: volumetry and voxel-based morphometry. *Korean J Radiol, 7*(3), 162-172. doi:10.3348/kjr.2006.7.3.162

Tahara, N., Kawakami, H., Chen, K. Q., Anderson, A., Yamashita Peterson, M., Gong, W., . . . Kawakami, Y. (2019). Sall4 regulates neuromesodermal progenitors and their descendants during body elongation in mouse embryos. *Development, 146*(14). doi:10.1242/dev.177659

Tao, H., Manak, J. R., Sowers, L., Mei, X., Kiyonari, H., Abe, T., . . . Bassuk, A. G. (2011). Mutations in prickle orthologs cause seizures in flies, mice, and humans. *Am J Hum Genet, 88*(2), 138-149. doi:10.1016/j.ajhg.2010.12.012

Taverna, E., Gotz, M., & Huttner, W. B. (2014). The cell biology of neurogenesis: toward an understanding of the development and evolution of the neocortex. *Annu Rev Cell Dev Biol, 30*, 465-502. doi:10.1146/annurev-cellbio-101011-155801

Thakran, S., Guin, D., Singh, P., Singh, P., Kukal, S., Rawat, C., . . . Kukreti, R. (2020). Genetic Landscape of Common Epilepsies:

Advancing towards Precision in Treatment. *Int J Mol Sci, 21*(20). doi:10.3390/ijms21207784

Thompson, C. H., Hawkins, N. A., Kearney, J. A., & George, A. L., Jr. (2017). CaMKII modulates sodium current in neurons from epileptic Scn2a mutant mice. *Proc Natl Acad Sci U S A, 114*(7), 1696-1701. doi:10.1073/pnas.1615774114

Togawa, K., Yan, Y. X., Inomoto, T., Slaugenhaupt, S., & Rustgi, A. K. (2000). Intestinal cell kinase (ICK) localizes to the crypt region and requires a dual phosphorylation site found in map kinases. *J Cell Physiol, 183*(1), 129-139. doi:10.1002/(SICI)1097-4652(200004)183:1<129::AID-JCP15>3.0.CO;2-S

Tomson, T., Battino, D., & Perucca, E. (2016). Valproic acid after five decades of use in epilepsy: time to reconsider the indications of a time-honoured drug. *Lancet Neurol, 15*(2), 210-218. doi:10.1016/S1474-4422(15)00314-2

Tong, Y., Park, S. H., Wu, D., Xu, W., Guillot, S. J., Jin, L., . . . Fu, Z. (2017). An essential role of intestinal cell kinase in lung development is linked to the perinatal lethality of human ECO syndrome. *FEBS Lett, 591*(9), 1247-1257. doi:10.1002/1873-3468.12644

Torkamani, A., Bersell, K., Jorge, B. S., Bjork, R. L., Jr., Friedman, J. R., Bloss, C. S., . . . Kearney, J. A. (2014). De novo KCNB1 mutations in epileptic encephalopathy. *Ann Neurol, 76*(4), 529-540. doi:10.1002/ana.24263

Toshinobu Miyamoto, M. I., Takeshi Shin, Masafumi Ikezawa, Yasuhiro Utsuno, Yasuaki Saijo, Hiroshi Okada, Kazuo Sengoku. (2020). SOX30 might not be associated with Sertoli cell-only syndrome in azoospermic Japanese men. *Integr Mol Med*.

Trcek, T., Sato, H., Singer, R. H., & Maquat, L. E. (2013). Temporal and spatial characterization of nonsense-mediated mRNA decay. *Genes Dev, 27*(5), 541-551. doi:10.1101/gad.209635.112

Truett, G. E., Heeger, P., Mynatt, R. L., Truett, A. A., Walker, J. A., & Warman, M. L. (2000). Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide and tris (HotSHOT). *Biotechniques, 29*(1), 52, 54. doi:10.2144/00291bm09

Tu, F., Sedzinski, J., Ma, Y., Marcotte, E. M., & Wallingford, J. B. (2018). Protein localization screening in vivo reveals novel regulators of

multiciliated cell development and function. *J Cell Sci, 131*(3). doi:10.1242/jcs.206565

Tu, Q., Pi, M., Karsenty, G., Simpson, L., Liu, S., & Quarles, L. D. (2003). Rescue of the skeletal phenotype in CasR-deficient mice by transfer onto the Gcm2 null background. *J Clin Invest, 111*(7), 1029-1037. doi:10.1172/JCI17054

Ugur, B., Chen, K., & Bellen, H. J. (2016). Drosophila tools and assays for the study of human diseases. *Dis Model Mech, 9*(3), 235-244. doi:10.1242/dmm.023762

Uhlen, M., Fagerberg, L., Hallstrom, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., . . . Ponten, F. (2015). Proteomics. Tissue-based map of the human proteome. *Science, 347*(6220), 1260419. doi:10.1126/science.1260419

Vadlamudi, L., Milne, R. L., Lawrence, K., Heron, S. E., Eckhaus, J., Keay, D., . . . Berkovic, S. F. (2014). Genetics of epilepsy: The testimony of twins in the molecular era. *Neurology, 83*(12), 1042-1048. doi:10.1212/WNL.0000000000000790

van Dam TJP, K. J., van der Lee R, de Vrieze E, Wunderlich KA, Rix S, et al. . (2019). CiliaCarta: An integrated and validated compendium of ciliary genes. *PLoS ONE, 14*((5)). Retrieved from https://doi.org/10.1371/journal.pone.0216705

Vargas-Hurtado, D., Brault, J. B., Piolot, T., Leconte, L., Da Silva, N., Pennetier, C., . . . Basto, R. (2019). Differences in Mitotic Spindle Architecture in Mammalian Neural Stem Cells Influence Mitotic Accuracy during Brain Development. *Curr Biol, 29*(18), 2993-3005 e2999. doi:10.1016/j.cub.2019.07.061

Veerappa, A. M., Saldanha, M., Padakannaya, P., & Ramachandra, N. B. (2013). Genome-wide copy number scan identifies disruption of PCDH11X in developmental dyslexia. *Am J Med Genet B Neuropsychiatr Genet, 162B*(8), 889-897. doi:10.1002/ajmg.b.32199

Vijai, J., Kapoor, A., Ravishankar, H. M., Cherian, P. J., Kuruttukulam, G., Rajendran, B., . . . Anand, A. (2005). Protective and susceptibility effects of hSKCa3 allelic variants on juvenile myoclonic epilepsy. *J Med Genet, 42*(5), 439-442. doi:10.1136/jmg.2004.023812

Vizard, T. N., O'Keeffe, G. W., Gutierrez, H., Kos, C. H., Riccardi, D., & Davies, A. M. (2008). Regulation of axonal and dendritic growth by

the extracellular calcium-sensing receptor. *Nat Neurosci, 11*(3), 285-291. doi:10.1038/nn2044

Vollmar, C. (2015). *Neuroimaging of functional and structural alterations in Juvenile Myoclonic Epilepsy and Frontal Lobe Epilepsy.* (Ph.D.), UCL Institute of Neurology London.

Wagner, T., Wirth, J., Meyer, J., Zabel, B., Held, M., Zimmer, J., . . . Scherer, G. (1994). Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene SOX9. *Cell, 79*(6), 1111-1120. doi:10.1016/0092-8674(94)90041-8

Wang, E. J., Gailey, C. D., Brautigan, D. L., & Fu, Z. (2020). Functional Alterations in Ciliogenesis-Associated Kinase 1 (CILK1) that Result from Mutations Linked to Juvenile Myoclonic Epilepsy. *Cells, 9*(3). doi:10.3390/cells9030694

Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res, 38*(16), e164. doi:10.1093/nar/gkq603

Wang, T. W., Stromberg, G. P., Whitney, J. T., Brower, N. W., Klymkowsky, M. W., & Parent, J. M. (2006). Sox3 expression identifies neural progenitors in persistent neonatal and adult mouse forebrain germinative zones. *J Comp Neurol, 497*(1), 88-100. doi:10.1002/cne.20984

Wegner, M. (1999). From head to toes: the multiple facets of Sox proteins. *Nucleic Acids Res, 27*(6), 1409-1420. doi:10.1093/nar/27.6.1409

Weiss, L. A., Purcell, S., Waggoner, S., Lawrence, K., Spektor, D., Daly, M. J., . . . Skuse, D. (2007). Identification of EFHC2 as a quantitative trait locus for fear recognition in Turner syndrome. *Hum Mol Genet, 16*(1), 107-113. doi:10.1093/hmg/ddl445

Welty, T. E. (2006). Juvenile myoclonic epilepsy: epidemiology, pathophysiology, and management. *Paediatr Drugs, 8*(5), 303-310. doi:10.2165/00148581-200608050-00003

Wen, J., Gao, Q., Wang, N., Zhang, W., Cao, K., Zhang, Q., . . . Shi, L. (2017). Association of microRNA-related gene XPO5 rs11077 polymorphism with susceptibility to thyroid cancer. *Medicine (Baltimore), 96*(14), e6351. doi:10.1097/MD.0000000000006351

Whissell, P. D., Eng, D., Lecker, I., Martin, L. J., Wang, D. S., & Orser, B. A. (2013). Acutely increasing deltaGABA(A) receptor activity impairs

memory and inhibits synaptic plasticity in the hippocampus. *Front Neural Circuits, 7*, 146. doi:10.3389/fncir.2013.00146

Whittington, N., Cunningham, D., Le, T. K., De Maria, D., & Silva, E. M. (2015). Sox21 regulates the progression of neuronal differentiation in a dose-dependent manner. *Dev Biol, 397*(2), 237-247. doi:10.1016/j.ydbio.2014.11.012

Williams, S. N., Locke, C. J., Braden, A. L., Caldwell, K. A., & Caldwell, G. A. (2004). Epileptic-like convulsions associated with LIS-1 in the cytoskeletal control of neurotransmitter signaling in Caenorhabditis elegans. *Hum Mol Genet, 13*(18), 2043-2059. doi:10.1093/hmg/ddh209

Willner, P. (1984). The validity of animal models of depression. *Psychopharmacology (Berl), 83*(1), 1-16. doi:10.1007/BF00427414

Woermann, F. G., Free, S. L., Koepp, M. J., Sisodiya, S. M., & Duncan, J. S. (1999). Abnormal cerebral structure in juvenile myoclonic epilepsy demonstrated with voxel-based analysis of MRI. *Brain, 122 (Pt 11)*, 2101-2108. doi:10.1093/brain/122.11.2101

Wolf, P., Yacubian, E. M., Avanzini, G., Sander, T., Schmitz, B., Wandschneider, B., & Koepp, M. (2015). Juvenile myoclonic epilepsy: A system disorder of the brain. *Epilepsy Res, 114*, 2-12. doi:10.1016/j.eplepsyres.2015.04.008

Xie, Y., Juschke, C., Esk, C., Hirotsune, S., & Knoblich, J. A. (2013). The phosphatase PP4c controls spindle orientation to maintain proliferative symmetric divisions in the developing neocortex. *Neuron, 79*(2), 254-265. doi:10.1016/j.neuron.2013.05.027

Yang, S. H., Bergo, M. O., Farber, E., Qiao, X., Fong, L. G., & Young, S. G. (2009). Caution! Analyze transcripts from conditional knockout alleles. *Transgenic Res, 18*(3), 483-489. doi:10.1007/s11248-008-9237-9

Yang, Y., Paivinen, P., Xie, C., Krup, A. L., Makela, T. P., Mostov, K. E., & Reiter, J. F. (2021). Ciliary Hedgehog signaling patterns the digestive system to generate mechanical forces driving elongation. *Nat Commun, 12*(1), 7186. doi:10.1038/s41467-021-27319-z

Yi, R., Qin, Y., Macara, I. G., & Cullen, B. R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev, 17*(24), 3011-3016. doi:10.1101/gad.1158803

Yingling, J., Youn, Y. H., Darling, D., Toyo-Oka, K., Pramparo, T., Hirotsune, S., & Wynshaw-Boris, A. (2008). Neuroepithelial stem cell proliferation requires LIS1 for precise spindle orientation and symmetric division. *Cell, 132*(3), 474-486. doi:10.1016/j.cell.2008.01.026

Zhang, D., Xie, D., Lin, X., Ma, L., Chen, J., Zhang, D., . . . Han, C. (2018). The transcription factor SOX30 is a key regulator of mouse spermiogenesis. *Development, 145*(11). doi:10.1242/dev.164723

Zhang, S., Zhu, X., Gui, X., Croteau, C., Song, L., Xu, J., . . . Guo, F. (2018). Sox2 Is Essential for Oligodendroglial Proliferation and Differentiation during Postnatal Brain Myelination and CNS Remyelination. *J Neurosci, 38*(7), 1802-1820. doi:10.1523/JNEUROSCI.1291-17.2018

Zhou, Y., Yang, S., Mao, T., & Zhang, Z. (2015). MAPanalyzer: a novel online tool for analyzing microtubule-associated proteins. *Database (Oxford), 2015*. doi:10.1093/database/bav108

Zuberi, S. M., Eunson, L. H., Spauschus, A., De Silva, R., Tolmie, J., Wood, N. W., . . . Hanna, M. G. (1999). A novel mutation in the human voltage-gated potassium channel gene (Kv1.1) associates with episodic ataxia type 1 and sometimes with partial epilepsy. *Brain, 122 (Pt 5)*, 817-825. doi:10.1093/brain/122.5.817