# Dissecting the molecular genetic basis of juvenile myoclonic epilepsy

A thesis submitted for the degree of

**Doctor of Philosophy**

By

**Pooja Barak**


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

**August 2021**

*Dedicated to*

*patients and their families*

# *Declaration*

I hereby declare that this thesis, entitled **"*Dissecting the molecular genetic basis of 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.

**Pooja Barak**

Place: Bengaluru

Date: August 23, 2021

# *Certificate*

This is to certify that the work described in this thesis **"*Dissecting the molecular genetic basis of juvenile myoclonic epilepsy*"** is the result of the investigations carried out by Ms. Pooja Barak 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: August 23, 2021

# *Acknowledgements*

First and foremost, I take this opportunity to express my appreciation and gratitude towards my advisor Professor Anuranjan Anand for his invaluable guidance, continuous support, and patience during the journey of my Ph.D. I would like to thank him for always encouraging me and helping me mature as a researcher.

I would like to thank the faculty members of MBGU, 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. Ravi Manjithaya, Dr. Kushagra Bansal, Prof. James Chelliah, Prof. Sheeba Vasu and Dr. G. R. Ramesh for their course work during my master's training and their comments and suggestions throughout the course of my PhD.

I thank our clinical collaborators, Dr. KS Mani, Dr. P Satishchandra, Dr. Sanjib Sinha, Dr. S K Shankar, Dr. K Radhakrishnan, Dr. G Kuruttukulam, Dr. Girish Gadre and Dr. Dilip KS, for their contribution towards this study.

I wish to thank all the former and present members of the human molecular genetics' lab, Dr. Manpreet Kaur, Dr. Praveen Raju, Dr. Nishtha Pandey, Dr. Kalpita Karan, Dr. Shalini Roychoudhuri, Dr. Rammurthi, Sourav, Shveta, Mariyam, Shrilaxmi, Sambhavi, Girija, Sukanya, Meenu, Varsha, Moumita, Angshumi, Yashwini, Nazia, Soumya, Jigyasa, Anjali and Chandrasekhara for making my stay in the lab pleasant and for helping troubleshoot the problems that I encountered in my work. I am indebted to Dr. Manpreet Kaur for not only guiding me during my MS but also for helping me all through my Ph.D. She was always there for me, whether it was a professional challenge or a personal crunch.

A special thank goes to all the summer trainees, Bindu, Sukanya, Natasha, Vaibhavi, Akshitha, Nivetha and Lakshmi, who I had an opportunity to train. All of them contributed to my work. This experience taught me a lot about mentoring and beyond.

My stay at JNCASR became very comforting because of the presence of these special people in my life, Anubha, Mariyam, Kirthana, Sutanuka, Shrilaxmi and Sambhavi. Standing by my side through thick and thin, helped me stay positive and remain focused throughout my journey.

I acknowledge the support of Anitha and Greeshma from the sequencing facility; Suma ma'am, Sunil, Prajwal and Keerthana from the confocal facility; Ramya, Nandini, Vidhya and Raju from the MBGU office.

I would also like to thank other key personnel of JNCASR who helped me all along some or the other way, Jaychandra Sir, Joydeep Sir, Dr. Princy, Dr. Panneer, Prof. Subi George, Prof. Sheeba Vasu, Prof. Tapas Maji, Prof. Ranjani Viswanatha, Prof. Jayanta Haldar, Dr. Kavitha and staff at administration office, academic office, computer lab, animal house, library, dhanvantari, hostel office, mess/dining facility, chandraya canteen, housekeeping & gardening team and security office.

Above all, I would like to thank my family. Words cannot express how grateful I am to my parents, my husband Pratik, my siblings Neeraj and Priya, my parents-in-law and brother-in-law, Pratyush for all the love and support that they have always provided me.

**Table of Contents**

**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 |
| CDC20B | cell division cycle 20B |
| 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 |
| DES | Desmin |
| DNA | Deoxyribonucleic acid |
| dNTP | Doxynucleotide triphosphate |
| EDTA | Ethylenediaminetetraacetic acid |
| EEG | Electroencephalogram |
| EFHC1 | EF hand domain (C-terminal)-containing 1 |
| 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 |
| KCl | Potassium chloride |

| | |
|---|---|
| LOD | Logarithm of Odds |
| MAF | Minor Allele Frequency |
| Mb | Megabases |
| mg | Milligram |
| $MgCl_2$ | Magnesium chloride |
| ml | Milliliter |
| mM | Millimolar |
| 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* |
| TMEM171 | Transmembrane protein 171 |
| 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 |

# Chapter 1

# Introduction

## 1.1 Epilepsy

Epilepsy is a common neurological disorder characterized by recurrent seizures usually unprovoked by any immediately identifiable cause. One of the first descriptions of epileptic seizures can be traced back to 2,000 B.C. in ancient Akkadian texts, a language widely used in the region of Mesopotamia. There are many writings that show that epilepsy was known in ancient times but was not understood until the mid-1950s. Modern advances in the understanding of epilepsy came during the 18th and 19th centuries (Gross, 1992). During this period, among others, John Hughlings Jackson studied epilepsy in detail and published his first paper on epilepsy in 1861. He had stated, "*A convulsion is but a symptom, and implies only that there is an occasional, an excessive, and a disorderly discharge of nerve tissue on muscles*". John Hughlings Jackson hypothesized that epilepsy was due to a paroxysmal, excessive discharge of neurons within a local area that could then affect more normal areas of the brain. He is considered the father of modern epileptology (Foyaca-Sibat, 2011).

According to the most recent clinical definition of epilepsy by International League Against Epilepsy (ILAE), a diagnosis of epilepsy is considered in the following circumstances: (i) at least two unprovoked or reflex seizures occurring twenty-four hours apart; (ii) one unprovoked or reflex seizure and a probability of further seizures similar to the general recurrence risk (at least 60%) after two unprovoked seizures, occurring over the next 10 years; (iii) diagnosis of an epilepsy syndrome (Fisher et al., 2014).

Seizures, a hallmark of epilepsy, are brief episodes of involuntary shaking which may involve a part of the body or the whole body and may be accompanied by loss of consciousness. An epileptic seizure is a

clinical manifestation of an abnormal and excessive discharge of a set of neurons in the brain. Seizures may arise in a localized area of the brain or may involve the whole brain. The signs and symptoms depend on which area of the brain is affected. The hippocampal formation and cerebral cortex are considered the most epileptogenic regions of the brain (Bozzi et al., 2012).

## 1.2 Epilepsy classification

The classification of epilepsies is an evolving process for which the goal is to establish a clinically relevant, scientifically based classification. Classification of epilepsy was first published in 1960 with official updates made in 1981 for seizures (Commission on Classification and Terminology of the International League Against Epilepsy, 1981) and in 1989 for epilepsies (Commission on Classification and Terminology of the International League Against Epilepsy, 1989). In 2010, the first major modernization of the epilepsy classification was recommended. It was based on the new insights gained in the field of epilepsy due to scientific advances in modern neuroimaging, genomic technologies, and concepts in molecular biology (Berg et al 2010).

The International League Against Epilepsy (ILAE) revised the classification of epilepsies in 2017. The classification is now organized on three levels (Figure 1) (Scheffer et al., 2017). The first level defines the seizure types into focal seizures which originate within brain networks limited to one hemisphere; generalized seizures which affect the brain in both hemispheres; and seizures of unknown onset. The second level comprises the diagnosis of the epilepsy type, which consists of four main classes: focal epilepsy, generalized epilepsy, combined focal and generalized epilepsy, and unknown epilepsy. This level of classification also includes a new category of combined focal and generalized epilepsy. This signifies the existence of shared mechanisms between the well-established groups of focal epilepsy and generalized epilepsy. The third level of classification is the delineation

of epilepsy syndrome when a precise syndromic diagnosis can be made. Due to the connotations of understanding the aetiology for clinical management, the revised classification emphasizes the importance of establishing the aetiology at each of the three levels. The etiological classes are structural, genetic, infectious, metabolic, immune, and unknown.

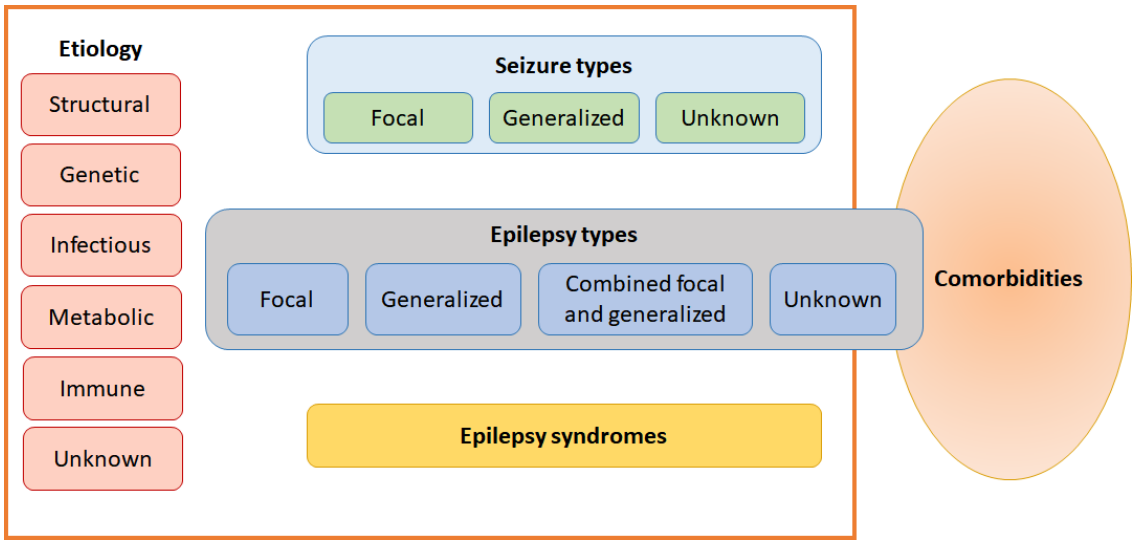



**Figure 1:** Adapted from International League Against Epilepsy 2017 classification of the epilepsies, with permission (Scheffer et al., 2017).

### 1.3 Genetic Generalized Epilepsies (GGE)

Genetic generalized epilepsies (GGE) constitute a heterogeneous group of epilepsies with generalized seizures. This syndrome is associated with bilaterally synchronous and symmetrical EEG discharges. GGE syndromes are common and account for 30-40% of all human epilepsies (Jallon and Latour, 2005a). They are usually not associated with any brain lesions, mental impairment, or neurological abnormalities other than seizures. GGE represents electro-clinical syndromes, in which a combination of clinical features such as seizure types, age of onset and electroencephalogram (EEG) features (generalized spike-wave discharges) are required for diagnosis.

GGEs are divided based on the age of onset and the type of seizure, into four common sub-syndromes: childhood absence epilepsy (CAE),

juvenile absence epilepsy (JAE), juvenile myoclonic epilepsy (JME) and epilepsy with generalized tonic-clonic seizures (EGTCS). In addition, there are rare GGE sub-syndromes that include benign myoclonic epilepsy of infancy (BMEI), early-onset absence epilepsy, myoclonic atonic epilepsy (MAE), epilepsy with myoclonic absences, eyelid myoclonia with absences and absence status epilepsy. The frequent seizure types associated with GGE are, absence seizures that involve brief, sudden lapses of consciousness, myoclonic seizures which give rise to jerking spasms in muscles or group of muscles and generalized tonic-clonic seizures (GTCS) (Mattson, 2003). GTCS are manifested as bilateral, convulsive tonic and clonic muscle contractions. The tonic phase of the seizure causes muscle stiffening in the extremities and the clonic phase involves rapid contraction and relaxation of the muscles that lead to contractions. The EEG trait of generalized spike-waves is the defining factor in the diagnosis of GGE, indicating bilateral synchronous discharges in a normal background.

### 1.4 Juvenile myoclonic epilepsy (JME)

Juvenile myoclonic epilepsy was first described by Théodore Herpin in 1867 {The Epileptology of Théodore Herpin (1799–1865) - Eadie - 2002 - Epilepsia - Wiley Online Library} and then, several patients with JME were reported and described in an article by Janz and Christian in 1957 (Janz and Christian, 1957). Therefore, JME is also known as Janz syndrome. Later, the term juvenile myoclonic epilepsy was included in the International League Against Epilepsy classification of epileptic syndromes in 1989 (ILAE 1989).

JME is an age-dependent disorder with onset typically during adolescence (12-15 years). It is characterized by irregular myoclonic jerks (MJ) occurring early in the morning, although many patients experience generalized tonic-clonic seizures (GTCS) and absence seizures as well. Myoclonus manifests itself as sudden jerks in the muscles and, usually involves bilateral and proximal upper extremities.

It can also afflict distal muscles, such as the hands or the lower limbs (Moschetta et al., 2011). Having interpreted as a manifestation of myoclonus nervousness or restlessness, it is the seizure that brings the patient to the doctor or hospital. Convulsive seizures usually come a few months after the onset of myoclonus, but in some, it may take several years (Asconapé et al 1984).

### 1.4.1 Epidemiology

The incidence of JME has been estimated to be 1 per 100,000 individuals. The prevalence of JME has been estimated to be 5-10% of all epilepsies and 18% of idiopathic generalized epilepsies (IGE) (Jallon and Latour, 2005b). JME typically appears in the second decade of life. However, the age of onset of JME spans a wide range from about 8-36 years, with peak onset between 12 and 18 years (Canevini et al., 1992; Vijai et al., 2003a). Those with the onset of JME outside the 8–36 years of age bracket are uncommon and should be carefully evaluated for other diagnoses.

### 1.4.2 Clinical features

JME is characterized by the occurrence of absence-, myoclonic-, and generalized tonic-clonic (GTC)- seizures. All patients with JME exhibit myoclonic seizures with 85%-90% of patients having GTC seizures, and about 20-40% of patients showing absence seizures (Vijai et al., 2003a; Jayalakshmi et al., 2010).

Myoclonic seizures, which are the hallmark of this condition, are essential for the diagnosis of JME. Very rarely do these manifest as the only seizure type in the patients. They are depicted as brief, typically bilateral, jerking movements of the arms and occasionally legs. The patients retain consciousness. Myoclonic seizures sometimes get misdiagnosed as twitches or anxiety/nervousness. Increasing intensity and frequency of myoclonic jerks in clusters can evolve into GTS. GTS seizures generally appear a few months after the onset of myoclonic

seizures. The first seizures to manifest in JME are generally absence seizures that tend to appear 3 to 5 years before the onset of myoclonic or GTS seizures. Sometimes absence seizures can occur as early as 5 to 6 years of age. A diagnosis of childhood or juvenile absence epilepsy is made in cases with an early and isolated occurrence of absence seizures. In JME patients, seizures, especially myoclonic seizures, occur 30 minutes to an hour after waking in the morning. Typical EEG features of JME consist of generalized discharges of single or multiple spikes and slow waves of 3–5 Hz frequency. Rare incidences of complexes as slow as 2 Hz or as fast as 7 Hz may be evident in some cases. Precipitating factors of epileptic seizures are wide-ranging. The most common one, sleep deprivation, is reported in 58.3-89.5% of cases. Other seizure triggering factors that have been ascertained include fatigue (73.7%), photosensitivity (36.8%), menses (24.1%), mental concentration (22.8%), and stress, excitement, or frustration (12.3%). Another precipitating factor found in 51.2% of cases is alcohol ingestion (Alfradique and Vasconcelos, 2007). About 30% to 40% of JME patients exhibit photosensitivity. Flashing lights, sunlight, TV, and computer can trigger seizures in many patients with JME. Photosensitive patients have a propensity to have an earlier onset of seizures.

Physical examination of JME patients shows that there are no abnormalities, and cognition is well preserved. Studies suggest an increased prevalence of psychiatric disorders, including anxiety, mood disorder, and personality disorders in patients with JME (Moschetta et al., 2011).

### 1.4.3 JME genetics

JME is primarily genetic in origin and follows both Mendelian and complex modes of inheritance. In JME families, the clinical and EEG traits are largely inherited in an autosomal dominant inheritance pattern. However, the inherited mutant alleles often show incomplete

penetrance and lead to phenotypic heterogeneity within families. Several studies indicate complex polygenic genetic architecture of JME suggesting the implicated genes may be modified by additional genes or possibly by environmental factors. To date, a few genes with disease-segregating mutations or susceptibility-conferring alleles have been identified in a rather small number of JME patients, indicating that its genetic determinants are largely undiscovered (Table 1).

**Table 1: Genes and loci associated with JME/ GGE**

| Loci/Gene | Type of study | Reference |
|---|---|---|
| **Loci** | | |
| 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) |
| | | |
| **Genes** | | |
| *BRD2 (RING3)* | Linkage and association analysis | Greenberg et al., 2005, Pal et al., 2003 |
| *EFHC1* | Linkage and Mutational analysis | Liu et al., 1995, Liu et al., 1996, Bai et al., 2002, Suzuki et al., 2004, Raju et al., 2017 |
| *GABRA1* | Linkage and mutational analysis | Cossette et al., 2002; Lachance-Touchette et al., 2011 |
| *CLCN2* | Linkage and Mutational analysis | Stogmann et al., 2006 |
| *CASR* | Linkage and Mutational analysis | Kapoor et al., 2008 |
| *ICK* | Linkage and Mutational analysis | Bailey et al., 2018 |
| *CACNB4* | Mutational analysis | Escayg et al., 2000 |

| | | |
|---|---|---|
| *GABRD* | Mutational analysis | Dibbens et al., 2004 |
| *CACNA1H* | Mutational analysis | Heron et al., 2004, Heron et al., 2007 |
| *SLC2A1* | Mutational analysis | Striano et al., 2012 |
| *SLC12A5* | Mutational analysis | Kahle et al., 2014, Duy et al., 2019 |
| *RORB* | Mutational analysis | Rudolf et al., 2016 |
| *KCNMA1* | Mutational analysis | Li et al., 2018 |
| *TAP1* | Linkage and association analysis | Layouni et al., 2010 |
| *KCNQ3* | Association study | Vijai et al., 2003b |
| *CX36* | Association study | Mas et al., 2004, Hempelmann et al., 2006 |
| *KCNJ10* | Association study | Lenzen et al., 2005 |
| *KCNQ2* | Association study and sequencing | Neubauer et al., 2008 |
| *CHRNA4* | Association study | Rozycka et al., 2009 |
| *CHRNA7* | case-control and familial segregation analysis, array CGH analysis | Helbig et al., 2009, Mefford et al., 2010, Szafranski et al., 2010 |
| *GRM4* | Genome-wide association study | Muhle et al., 2010, Parihar et al., 2014 |
| *CHRM3* | Genome-wide association study | Steffens et al., 2012 (EPICURE Consortium and EMINet Consortium) |
| VRK2 | Genome-wide association study | Steffens et al., 2012 (EPICURE Consortium and EMINet Consortium) |
| ZEB2 | Genome-wide association study | Steffens et al., 2012 (EPICURE Consortium and EMINet Consortium), ILAE Consortium on Complex Epilepsies, 2018 |
| SCN1A | Genome-wide association study | Steffens et al., 2012 (EPICURE Consortium and EMINet Consortium), ILAE Consortium on Complex Epilepsies, 2018 |
| PNPO | Genome-wide association study | Steffens et al., 2012 (EPICURE Consortium and EMINet Consortium), ILAE Consortium on Complex Epilepsies, 2018 |
| FANCL | Genome-wide association study | ILAE Consortium on Complex Epilepsies, 2018 |
| BCL11A | Genome-wide association study | ILAE Consortium on Complex Epilepsies, 2018 |
| SCN3A | Genome-wide association study | ILAE Consortium on Complex Epilepsies, 2018 |
| SCN2A | Genome-wide association study | ILAE Consortium on Complex Epilepsies, 2018 |
| TTC21B | Genome-wide association study | ILAE Consortium on Complex Epilepsies, 2018 |
| STAT4 | Genome-wide association study | ILAE Consortium on Complex Epilepsies, 2018 |
| PCDH7 | Genome-wide association study | ILAE Consortium on Complex Epilepsies, 2018 |
| GABRA2 | Genome-wide association study | ILAE Consortium on Complex Epilepsies, 2018 |
| KCNN2 | Genome-wide association study | ILAE Consortium on Complex Epilepsies, 2018 |
| ATXN1 | Genome-wide association study | ILAE Consortium on Complex Epilepsies, 2018 |
| GRIK1 | Genome-wide association study | ILAE Consortium on Complex Epilepsies, 2018 |
| STX1B | Genome-wide association study | ILAE Consortium on Complex Epilepsies, 2018 |
| *CACNA1G* | Genome-wide association study | Feng et al., 2019 (Epi25 Collaborative) |
| *EEF1A2* | Genome-wide association study | Feng et al., 2019 (Epi25 Collaborative) |
| *GABRG2* | Genome-wide association study | Feng et al., 2019 (Epi25 Collaborative) |

## 1.5 Molecular genetics

Over the past two decades, JME has been a subject of many genetic studies, which have led to identification of several loci and genes (Delgado-Escueta, 2007; Delgado-Escueta et al., 2013; Thakran et al., 2020). These studies further support and establish the view of JME as a complex trait that is associated with high genetic and phenotypic heterogeneity. Genetic linkage- and association- studies have unravelled genes, loci as well as several susceptibility alleles for JME, reflecting its monogenic, oligogenic and polygenic genetic architecture. There are many different genetic approaches employed to identify the loci or genes associated with JME. Some of the most common techniques employed are genome-wide linkage analysis, candidate gene sequencing, association studies and massively parallel exome/genome sequencing studies.

### 1.5.1 Linkage studies and mutational analysis of candidate genes

JME, an IGE sub-syndrome manifests a complex mode of inheritance and is inherited in a Mendelian or near-Mendelian manner in multi-generational families. Though, the underlying genetic factor can be a variant of a large effect responsible for the disease or a low-frequency allele with moderate effect - both, along with modifier alleles result in phenotypic and clinical heterogeneity among the affected relatives in families as well as in populations. So far, linkage studies on such relatively uncommon large multiplex JME families have identified a few loci and genes with disease-segregating mutations in affected family members. However, the mutations in these genes do not account for a majority of the JME patients from across different populations. This indicates a complex genetic and polygenic basis of JME, wherein only a small fraction of causative determinants has been deciphered and validated. The family-based linkage studies help by focusing the search to JME-linked specific chromosome loci and filtering the candidate

genes to be screened, excluding major part of the exome or genome for involvement in JME.

Screening of candidate genes for mutations is based on a prior hypothesis and aims at identifying more rare variants in epilepsy patients as compared to control subjects. These genes are normally chosen due to their association with an epilepsy sub-syndrome or seizure phenotype in either humans or animal models or both (Escayg et al., 2001; Arsov et al., 2012). So far, both these approaches have led to the identification of genetic determinants for JME in a small number of families, and risk alleles remain undefined in about 90% of the patients. Linkage analysis has revealed various genes associated with JME: *EFHC1* at 6p12 (EJM1), *GABRA1* at 5q34 (EJM5), *CACNB4* at 2q23 (EJM6), *GABRD* at 1p36, *CASR* at 3q13.3-q21.1, *CLCN2* at 3q27 (EJM8), and *ICK* at 6p12 (EJM10). In addition, EJM loci, EJM2 at 15q14, EJM3 at 6p21, EJM4 at 5q12-q14, and EJM9 at 2q33-q36, have been identified by linkage analysis.

Variations in *EFHC1* are the most commonly observed in families with JME (Ma et al., 2006; Medina et al., 2008). *EFHC1* encodes a microtubule-associated protein that is involved in cell division and neuronal migration. In vitro studies have shown that EFHC1 variants cause impairment in radial migration by disruption of radial glial scaffolds which are progenitor cells for cortical development (Noctor et al., 2001; de Nijs et al., 2012). Therefore, these defects during corticogenesis may damage the normal brain circuitry during development which in turn can cause epilepsy (de Nijs et al., 2009).

### 1.5.2 Association studies

Like many other complex phenotypes, JME has been subjected to genetic association studies in various populations (Tan et al., 2004). Many candidate genes, chosen based on their physiological roles or their localization in and around JME-linked loci, have been analyzed for association with JME. A few genome-wide association studies (GWAS)

for epilepsy patients have been performed and results from these studies suggest that common genetic variations make a modest contribution towards epilepsy susceptibility.

Since the first locus for JME was found to be in the proximity of the HLA region on chromosome 6, initial association studies in JME addressed whether JME-locus is inside or outside the HLA region. Preliminary evidence for an association between JME and alleles in the HLA region was reported in German (Durner et al., 1992) and Saudi Arabian (Obeid et al., 1994) populations. Greenberg and colleagues (1996) reported that the frequency of HLA-DR13 and HLA-DQB1 alleles were significantly higher in JME patients compared to non-JME patients, indicating that the JME locus probably lies within the HLA region and that JME is genetically different from certain other forms of IGEs. However, other groups found no significant evidence of association of HLA alleles with JME in various populations (Moen et al., 1995; Sander et al., 1997; Le Hellard et al., 1999).

Pal et al., 2003 identified highly significant linkage disequilibrium (LD) between JME and a core haplotype of five SNP and microsatellite markers with the LD peaking at *BRD2* at 6p21.3. Two JME-associated SNP variants in the promoter region of *BRD2* were detected; however, no other potentially causative variants were identified in 20 probands from families which had shown positive linkage to the locus. Subsequently, a multi-centre 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 Ireland, but showed a lack of replication in the German, Australian and Indian populations studied (Cavalleri et al., 2007). Similarly, association analysis in a Dutch population identified no significant difference between the allele frequencies of SNP markers located in *BRD2* in cases and controls (de Kovel et al., 2007).

In a group of 46 JME patients, including 29 individuals from families used to identify susceptibility locus at 15q14 (Elmslie et al., 1997), a significant association between JME and the c.588C>T variant, located in exon 2 of the *CX36* gene was observed (Mas et al 2004). Bioinformatics analysis of the associated SNP suggested that it perturbed the expression of *CX36* by affecting exonic splicing enhancers. For the IGE-linked locus at 18q21, case-control and family-based association methods identified *ME2* (malic enzyme 2) as a gene predisposing to IGE. An increased risk to IGE was found when the *ME2*-centered nine-SNP haplotype was present in the homozygous state (Greenberg et al 2005). However, a population-based association study involving 660 IGE patients and 666 controls of German descent did not exhibit the association of genetic variations in *ME2* to common IGE syndromes (Lenzen et al., 2005b).

A two-staged genome-wide association study, including 3020 patients with GGE (1434 GAE and 1134 JME patients) and 3914 controls of European ancestry was carried out by EPICURE Consortium and EMINet Consortium to identify susceptibility alleles for common generalized epilepsies (Steffens et al., 2012). Joint analysis for stage-1 and -2 of the study revealed a significant association at 2p16.1 (rs13026414) and 17q21.32 (rs72823592) for genetic generalized epilepsies. The search for syndrome-related susceptibility genes discovered a significant association at 2q22.3 (rs10496964) for GAE (absence epilepsies) and at 1q43 (rs12059546) for JME. The JME-associated region at 1q43 encompasses the gene encoding the M3 muscarinic acetylcholine receptor (*CHRM3*) (rs12059546).

A meta-analysis of genome-wide association studies from 12 cohorts of epilepsy patients and controls was conducted to detect risk variants for common epilepsies (ILAE Consortium on Complex Epilepsies 2014). The data obtained from published and unpublished genetic cohort studies from EPICURE, EPIGEN, Philadelphia (PA, USA), the Imperial-Liverpool-Melbourne Collaboration, GenEpa, and Hong Kong (China)

were used for the meta-analysis. This study included 8696 cases and 26157 controls, belonging to European, Asian or African ancestry. Analysis in the all-epilepsy category revealed significant association at 2q24.3 (*SCN1A*) and 4p15.1 (*PCDH7*). However, for meta-analysis of genetic generalized epilepsy (GGE), 2606 cases and 18990 controls from eight cohorts were used. The analysis detected the genome-wide significant signal at 2p16.1 (rs2947349), located near the genes encoding vaccinia-related kinase 2 (*VRK2*) and Fanconi anaemia, complementation group L (*FANCL*). The association of 2p16.1 region in the GGE cohort had been earlier reported by EPICURE Consortium (Steffens et al 2012, EPICURE Consortium and EMINet Consortium). Additionally, suggestive evidence for association with GGE was found at regions 4p15.1 (*PCDH7*), 5q22.3 (intergenic) and 11q22.2 (*MMP8*).

A systematic review of genetic association studies was done to evaluate the findings of existing genetic association studies that have examined the genetic variants underlying JME. Data from fifty studies, encompassing 229 polymorphisms on or near 55 genes were used for the analysis. Only three variants, rs2029461 SNP in *GRM4*, rs3743123 in *CX36* and rs3918149 in *BRD2,* showed a significant association with JME in at least two different background populations. The lack of replication of results has been attributed to various factors, such as limitations of experimental design, endophenotypes, channelopathy issues and genetic heterogeneity (dos Santos et al., 2017).

With an aim to fill the missing genetic heritability component of epilepsies, an expanded genome-wide association study was conducted on 15,212 cases and 29,677 controls (International League Against Epilepsy Consortium on Complex Epilepsies, 2018). Broadly, the cases were classified into focal epilepsies, genetic generalized epilepsies, and unclassified epilepsies. This study led to the identification of 11 novel genome-wide significant loci, 16q12.1, 2p24.1, 2q32.3, 4p12, 5q22.3, 6p22.3, 6q22.33, 21q22.11, 16p11.2, 3q25.31 and 6q22.31, and five previously reported loci, 2p16.1, 2q24.3, 4p15.1, 17q21.32 and 2q22.3,

were replicated. Eleven loci, 2p24.1, 2p16.1, 2q24.3, 2q32.3, 4p15.1, 4p12, 5q22.3, 6p22.3, 6q22.33, 17q21.32 and 21q22.11, were found to be associated with genetic generalized epilepsies, the group of epilepsies where despite having the highest heritability, least genetic progress have been made. A novel genome-wide significant association at 16p11.2 was found to be associated with JME. Gene mapping and biological prioritization implicate *STX1B* as a potential gene underlying the association of JME at the 16p11.2 locus (Table 1).

### 1.5.3 Massive parallel sequencing-based consortium studies

Given that linkage-based candidate gene analysis and association studies have not had much success in identifying large-effect variants in a majority of patients with common generalized epilepsies, comprehensive characterization of variants across the genome by next-generation sequencing (NGS) may help reveal common risk factors for epilepsies. The advancement in next-generation sequencing technologies and the reduced costs have made whole-exome sequencing (WES) or whole-genome sequencing (WGS) of individual epilepsy patients feasible.

The first reported NGS study for IGE had analyzed the exome sequences of 118 IGE patients [juvenile myoclonic epilepsy (n = 93) and absence epilepsy (n = 25)] and 242 controls of European ancestry (Heinzen et al., 2012). The study employed a two-stage approach involving exome sequencing of 118 IGE-affected individuals and a subsequent large-scale follow-up genotyping of identified candidate variants in a larger cohort of 878 IGE-affected subjects and 1830 controls. In total, 3897 candidate variants were subsequently genotyped in the larger follow-up cohort. No variant reached the statistical significance level of association in the study. However, several variants were enriched in patients than controls, notably, in genes *GREM1, OR10S1, PPEF2* and *CHD1*. A total of 1289 variants were reported exclusive to IGE patients and among these, a missense

mutation (c.620C>T, p.Ala207Val) in *PSME2* was observed in five unrelated cases. Findings from this study failed to reveal a single variant that could account for more than 1% of the cases, indicating high genetic heterogeneity in generalized epilepsies. This study pointed out that identification of disease-causing rare variants that are present at very low frequencies needs to be carried out through genome-/exome-based sequencing studies in very large sample sizes or co-segregation analyses in multiplex families (Heinzen et al., 2012).

Additionally, the massively parallel sequencing technology has provided diagnostic screening methods for genetically complex and heterogenous disorders such as epilepsies. One such study reported the use of targeted panels comprising 265 most relevant epilepsy genes or genes related to epilepsy phenotypes (37 IGE associated genes) to identify mutations for epilepsy. Disease-causing mutations were detected in 16 out of 33 patients with different epileptic syndromes, belonging to a cohort from Germany and Switzerland (Lemke et al., 2012).

The National Institute of Neurological Disorders and Stroke (NINDS) initiated a project in the year 2012 wherein the goal was to sequence 4000 epilepsy patients to uncover the genetics of human epilepsies (Epi4K: Gene discovery in 4,000 genomes, 2012). This Epi4K project focused on two areas of epilepsy. The first one aimed at addressing the genetics of epileptic encephalopathies, conditions defined by severe epilepsy that are generally refractory and accompanied by comorbid cognitive and behavioural dysfunctions. While epileptic encephalopathies are extremely diverse, the study focused on two of the more common types: Infantile Spasms (IS) and Lennox-Gastaut syndrome (LGS). They analysed 165 patients with IS and LGS collected by the Epilepsy Phenome/Genome Project (EPGP) for de novo mutations that were not present in the parents using exome sequencing. Using this approach, the Epi4K researchers identified possibly pathogenic de novo mutations in at least 15% of patients. Nine genes were found to have de novo mutations in two or more

probands, of which five genes had an already known association with epileptic encephalopathy. The known epileptic encephalopathy genes with mutations in multiple Epi4K subjects were: *SCN1A, STXBP1, SCN8A, SCN2A*, and *CDKL5*. Additional novel genes identified with mutations in multiple individuals included: *GABRB3, ALG13, CACNA1A, CHD2, FLNA, GABRA1*, *GRIN1*, *GRIN2B, HNRNPU*, *IQSEC2, MTOR,* and *NEDD4L* (Kearney, 2014).

The second category of epilepsy patients that the Epi4K initiative wanted to use in the whole genome sequencing (WGS) was from the common forms of epilepsy such as IGE and NFE found in families with two or more affected members. Sequence data from 640 individuals with familial genetic generalised epilepsy and 525 individuals with familial non-acquired focal epilepsy were compared to 3877 controls. Ultra-rare deleterious variants in genes established as causative for dominant epilepsy disorders (familial genetic generalised epilepsy: odd ratio [OR] 2·3, 95% CI 1·7–3·2, p=9·1 × 10−8; familial non-acquired focal epilepsy 3·6, 2·7–4·9, p=1·1 × 10−17) were found at significantly higher rates. Five known epilepsy genes ranked as the top five genes enriched for ultra-rare deleterious variation in individuals with familial non-acquired focal epilepsy. No individual gene was found to be significantly associated with familial genetic generalised epilepsy (Allen et al., 2017).

As part of an ongoing Epi25 Collaboration, a whole-exome sequencing analysis of 13,487 epilepsy-affected individuals and 15,678 control individuals has been performed. Compared to controls, individuals with any type of epilepsy carried an excess of ultra-rare, deleterious variants in constrained genes and genes previously associated with epilepsy. Genetic architectures of severe developmental and epileptic encephalopathies (DEEs) and two less severe epilepsies, genetic generalized epilepsy, and non-acquired focal epilepsy (NAFE) were compared. A gene-based rare variant collapsing analysis found that variants in DEE affected individuals are in significantly more intolerant

genic sub-regions than those in NAFE-affected individuals. Analysis of variants in genes without a prior disease association revealed a significant burden of loss-of-function variants in the genes most intolerant to such variants, indicating additional epilepsy-risk genes yet to be discovered. Compared to controls, individuals with epilepsy carried an excess of ultra-rare, deleterious variants in constrained genes and genes previously associated with epilepsy. The strongest enrichment of ultra-rare, deleterious variants was seen in individuals with DEEs and the least strong in individuals with NAFE. Inhibitory GABAA receptor genes showed enrichment for missense variants across all three classes of epilepsy. No enrichment for variants was seen in excitatory receptor genes. Even though no single gene surpassed exome-wide significance among individuals with GGE or NAFE, highly constrained genes and genes encoding ion channels, such as *CACNA1G, EEF1A2*, and *GABRG2* for GGE and *LGI1*, *TRIM3*, and *GABRG2* for NAFE, were among the top associations identified (Feng et al., 2019).

### 1.5.4 Copy number variations (CNV)

Apart from SNPs, CNVs have been demonstrated as risk or causal genetic markers in GGE. CNVs are a form of structural variation, which results in the alteration of the number of copies of one or more sections of DNA. They correspond to large segments of DNA being deleted (fewer than normal number) or duplicated (more than normal number), and may range from a kilobase to several mega-bases in size. CNVs are found on all chromosomes without any clinical significance; however, they can be enriched in certain genomic regions known as “hotspots”. The development of chromosome microarray and SNP genotyping arrays facilitated genome-wide screening for de-novo CNVs in large cohorts (Sharp et al., 2008). CNVs are increasingly found in epilepsy patients when often associated with developmental delay, autism spectrum disorder (ASD), intellectual disability and dysmorphic features (Sharp et al., 2008; Shinawi et al., 2010; Moreira et al., 2014).

Besides these, CNVs have also been reported in epilepsy syndromes associated with normal intellect. Helbig and colleagues identified microdeletions at 15q13.3 in 12 out of 1,223 individuals with IGE from Europe and North America (Helbig et al., 2009). This region was earlier implicated in 15q13.3 microdeletion syndrome in individuals with mental retardation and epilepsy, schizophrenia, autism, and other neuropsychiatric features (Sharp et al., 2008; Stefansson et al., 2008; Miller et al., 2009). The deletions in this critical region encompass several genes such as the *CHRNA7* (Cholinergic Receptor Nicotinic Alpha 7 Subunit), which is considered a candidate gene for epilepsy (Elmslie et al., 1997). In another independent cohort of 539 IGE familial and sporadic IGE cases (JME, n=183) from Australia and Europe, 15q13.3 microdeletions were observed in 7 cases (3 individuals with JME) (Dibbens et al., 2009). The 15q13.3 microdeletions were found *de novo* in three probands, two individuals had inherited from the unaffected parent and the transmission could not be determined in the other two probands. Though 15q13.3 microdeletions exhibit a strong association with IGE in case-control studies, it does not segregate as a dominant Mendelian trait in multiplex IGE families, indicating it as a susceptibility genetic factor for epilepsies. In a group of 246 cases with 15q13.3 microdeletion syndrome, 28% of the individuals were reported to have epilepsy/seizures (Lowther et al., 2015).

de Kovel et al., 2010 reported a study aimed to investigate the genomic hotspot regions, previously implicated in various neuropsychiatric disorders to identify the genetic factors conferring risk to IGE syndromes. A group of 1234 IGE patients from North-western Europe and 3022 controls from Germany were assessed using high-density SNP arrays, and recurrent microdeletions were found at 15q11.2 and 16p13.11. In an additional study, by using whole-genome oligonucleotide array CGH to a cohort of 517 patients with IGE or non-lesional epilepsies (JME, n=189), one or more rare CNVs were detected

in 8.9% of the affected individuals. Of these, 2.9% of the probands were found to harbour deletions in regions 15q11.2, 15q13.3, or 16p13.11 (Mefford et al., 2010).

Several studies report CNVs associated with epilepsies in the absence of dysmorphic features in the same hotspot regions which were previously identified in patients with the neurocognitive phenotypes, suggesting that these common etiological genetic factors have highly variable phenotypic expressivity (Mefford, 2014; Scheffer and Mefford, 2014). However, the microdeletions at three genomic hotspots 15q11.2, 15q13.3 and 16p13.11 are notably associated with IGE than any other epilepsy type, though with incomplete penetrance. It indicates that the microdeletion itself is not sufficient to cause epilepsy and may act as a risk factor in the complex genetic architecture known for IGE.

**Table 2: Copy number variations associated with epilepsies**

| Locus | Size | CNVs | Gene | Phenotype | References |
|---|---|---|---|---|---|
| 2p16.3 | ~287 and ~79 kb | Exon-disrupting deletion | *NRXN1* | Severe early-onset epilepsy | Harrison et al., 2011 |
| 2q24.2-q24.3 | 11Mb | Duplication/ deletion | *SCN1A*, *SCN2A*, *SLC4A10* | Idiopathic epilepsy | Krepischi et al., 2010 |
| 6p12.1 | 99.9kb | Micro-duplication | *BMP5* | Epilepsy | Naseer et al., 2015 |
| 7q11.22 | 78.7kb | Deletion | *AUTS2* | JME | Mefford et al., 2010 |
| 7q32.3 | 63.9kb | Microdeletion | *PODXL* | Epilepsy | Naseer et al., 2015 |
| 7q35 | 785.8kb | Deletion/ hemizygous deletions | *CNTNAP2* | GTCS | Mefford et al., 2010 |
| 15q11.2 | | Microdeletion | *NIPA2, CYFIP1* | GGE | de Kovel et al., 2010 |
| 15q13.3 | 1.4Mb | Microdeletion | *CHRNA7* | GGE (JME) | Helbig et al., 2009; Mefford et al., 2010 |
| 16p13.11 | 800Kb | Deletion | *NDE1* | MTLE | Liu et al., 2012 |
| 16p13.2 | 47Kb | Microdeletion | *GRIN2A* | RE (childhood epilepsy) | Dimassi et al., 2014 |
| 22q11.2 | 3Mb | Microdeletion | *DGCR6, DGCR6L* | GGE (JME) | Strehlow et al., 2016 |

JME: Juvenile myoclonic epilepsy, GTCS: Generalised tonic-clonic seizure, GGE: Genetic generalized epilepsy, MTLE: Mesial temporal lobe epilepsy, RE: Rolandic epilepsy

## 1.6 Modelling Epilepsy

Researchers require animal models that capture the relevant human pathology, in order to logically and efficiently design and test novel therapies for epilepsy. Animal models should recapitulate the aetiology of the human neurobiological disorder, must display the same or similar phenotypes as seen in patients. The animal models should eventually only act in response to treatments that are effective in patients with the modelled disorder (Garner, 2014). Such models provide an indirect opportunity to better understand human pathology, allowing us to find novel drug targets and to test the usefulness of therapies through preclinical trials.

Even though there are important differences between human and mouse neurobiology (Hodge et al., 2019), mice however represent strong candidate organisms for modelling human genetic disorders. Greater than 99% of human genes have mouse homologues (Chinwalla et al., 2002), and, within the brain, the three-dimensional pattern of gene expression is grossly conserved between humans and mice (Strand et al., 2007). In terms of epilepsy, mouse models of temporal lobe epilepsy mimic patients, with regard to the appearance of ictal seizures from the hippocampal formation and then circulating and spreading to the rest of the brain (Riban et al., 2002). Similarly, cortical and thalamic structures in both human and mouse models are responsible for generalized seizures (Cao et al., 2020) (Table 3).

**Table 3: Animal Models of epilepsy**

| Etiology | Rodent model | Epilepsy type |
|---|---|---|
| Genetic or presumed genetic epilepsy models | | |
| Tuberous sclerosis complex | Cell-specific conditional Tsc1-knockout or Tsc2-knockout mice | Tuberous sclerosis |

| | | |
|---|---|---|
| | | complex |
| Spontaneous mutations | Genetic absence epilepsy rat from Strasbourg | Rat absence epilepsy |
| | Wistar Albino Glaxo Rijswijk rats | |
| | High-voltage spike and wave spindles in rats | |
| | Tottering mice | Absence epilepsy |
| | Lethargic mice | |
| | Stargazer mice | |
| | Mocha 2j mice | |
| | Slow-wave mice | |
| | Ent mice | |
| | Ducky mice | |
| | Gabrg2 conditional knock-in mutation in mice | |
| Induced monogenic mutations | Knockout or knock-in mutations of voltage-gated ion channel subunits (sodium, potassium and calcium) in mice | Multiple |
| | Knockout or knock-in mutations of neurotransmitter receptor subunits (GABAA and nicotinic) and transporters | |
| | Knockout or knock-in mutations of accessory synaptic proteins | |
| | Cstb-knockout mice | |
| Developmental epileptic encephalopathies | SCN1A or SCN1B knock-in of human mutations in mice or constitutive or conditional knockout mice | Dravet syndrome |
| | ARX knock-in of human mutations in mice or constitutive or conditional knockout mice | Infantile spams |
| | Apc conditional knockout mouse | |
| | SCN8A knock-in of human mutations in mice Infantile spams | |

Models are both mice and rats unless mentioned otherwise. (Devinsky et al., 2018)

## 1.7 iPSC as a model system for studying epilepsy

Human-induced pluripotent stem cells (iPSCs) were first generated from fibroblasts in the year 2007. They can now be made from patients with neurological disorders and healthy controls (Takahashi et al., 2007; Kogut et al., 2018). Patient-derived iPSC lines are valuable for modelling disorders with complex genetic architecture, which would be impossible to model in whole animals due to the differences in the genetic background. If patient-derived cells are not available, control iPSC lines can also be engineered using tools such as CRISPR/Cas9 to recapitulate mutations seen in patients with monogenic disorders (Eggenschwiler et al., 2016).

**Table 4: Examples of iPSC model cell lines generated for certain epilepsy types**

| Epilepsy | Gene | Mutation | Source | Reference |
|---|---|---|---|---|
| Autosomal dominant temporal lobe epilepsy (ADTLE) | LGI1 | c.1418C>T (p.Ser473Leu) | Peripheral blood mononuclear cells | Tan et al., 2017 |
| Benign familial infantile epilepsy (BFIE) | PRRT2 | c.649dupC (p.Arg217Profs*8) | Fibroblast | Fruscione et al., 2018 |
| Glucose transporter 1 (GLUT1) deficiency syndrome | SLC2A1 | c.1454C>T (p.Pro485Leu) | Fibroblast | Meyer et al., 2018 |
| Chromosome 15q13.3 microdeletion syndrome | 15q11q13 | Microduplication | Fibroblast | Turco et al., 2018 |
| Focal cortical dysplasia type II | – | – | Fibroblast | Majolo et al., 2019 |
| Rett syndrome | MECP2 | Microdeletion (exon3–4) ,c.763C>T (p.Arg255*) | Fibroblast | Hinz et al., 2019 |
| Tuberous sclerosis complex | *TSC2* | c.1563dupA (p.His522Thrfs*67) | Fibroblast | Nadadhur et al., 2019 |
| Angelman syndrome | Chromosome deletion | 4.8 Mb 15q11-q13 deletion | Fibroblast | Pólvora-Brandão et al., 2018 |

| | | | | |
|---|---|---|---|---|
| CDKL5 encephalopathy | CDKL5 | c.1039C>T (p.Gln347*), c.863C>T (p.Thr288Ile) | Fibroblast | Amenduni et al., 2011 |
| Dravet syndrome | SCN1A | c.4933C>T (p.Arg1645*) | Fibroblast | Higurashi et al., 2013 |
| Type I Sialidosis (neuraminidase deficiency) | NEU1 | c.544A>G (p.Ser182Gly), c.667_679del (p.Leu223Glufs*76) | Peripheral blood mononuclear cells | Liu et al., 2018 |
| Myoclonic epilepsy with ragged red fibers (MERRF) | MT-TK | m.8344A>G | Fibroblast | Wu et al., 2018 |

Different types of neuronal and non-neuronal cell types can be differentiated from iPSCs (Hong and Do, 2019; Wiegand and Banerjee, 2019). This enables the opportunity to decipher different mechanistic aspects of the disorder as well testing different drug targets. Multi-electrode array systems can be used to assess the effect of candidate drugs on spontaneous neural activity (Tidball and Parent, 2016). There are some shortcomings of using iPSCs such as variations in the differentiation potential of iPSCs from generated from different donors (Kyttälä et al., 2016), neurons derived from human iPSCs are slow to fully differentiate (Nicholas et al., 2013) and to develop mature electrophysiology (Prè et al., 2014). Human neuronal cultures can not be used for modelling network disturbances in epilepsy because they lack the ability to reliably develop synchronous network activity unless they are co-cultured with astrocytes (Kuijlaars et al., 2016).

Researchers are now able to generate cerebral organoids from single iPSC lines due to advances in three-dimensional culture techniques and differentiation methods. These organoids are comprised of varied neuronal and glial cell types. These cells have the ability to show some degree of self-organization and exhibit spontaneous, synchronized neural activity (Izsak et al., 2019; Trujillo et al., 2019; Zafeiriou et al., 2020). These enhanced features could make organoids a better model system with the possibility of modelling both network-level and cell-autonomous defects in epilepsy.

## 1.8 Objectives of my work

To date, only a few genes for juvenile myoclonic epilepsy have been mapped and identified in large families. In the present study, I aimed to identify potentially disease-causing genes at three genomic loci linked to juvenile myoclonic epilepsy in families from the southern parts of India. These loci were identified by genome-wide linkage studies and mapped to chromosomes 5p15-q12, 2q33-q36 (EJM9, Ratnapriya et al., 2010) and 5q12–q14 (EJM4, Kapoor et al., 2007) regions. The objectives of my study were: (i) identification of the potentially causative variants/genes in the disease-linked intervals mentioned above employing whole exome/genome sequencing experiments. (ii) To evaluate cell biological correlates of the variants in the genes proposed to be responsible for epilepsy.

Chapters 2, 3 and 4 of this thesis present the results of these studies.

# Chapter 2

## Identification and characterization of *CDC20B*, a potential JME gene

### Summary

With an aim to identify a locus for JME, genome-wide linkage analysis was conducted in a 3-generation south Indian family, GLH 35, with several of its members affected by the disorder. In this analysis, a maximum two-point LOD score of 1.7 at θ=0 was obtained for the marker D5S426 on chromosome 5p13.2. Manual haplotyping helped define a 64-megabase critical region between markers D5S1981 and D5S407 at 5p15-q12. Whole-exome and whole-genome sequencing experiments revealed a rare missense variant, c.1525G>A (p.Ala509Thr) in the gene, *CDC20B* (cell division cycle 20B). Four additional rare, missense variants, p.Gln94His, p.Arg200Gly, p.Ile212Lys and p.Trp363Arg were identified among 552 JME patients examined. Interestingly, *CDC20B* is among the top 200 genes identified with the burden of ultra-rare variants in the Epi25 collaborative, with the GLH35 family variant, p.Ala509Thr, identified in six epilepsy patients. CDC20B encodes a 591-amino acid long, WD repeat domain-containing protein. During late telophase and cytokinesis, the protein localizes to the midbody, during other cell cycle stages, it is present in the cytoplasm. We found that CDC20B co-immunoprecipitates with TUBG1, EFHC1 and PLK1. Cytokinesis defects were observed on over-expression of the CDC20B variants, p.Arg200Gly, p.Trp363Arg and p.Ala509Thr, suggesting a cell cycle associated role for the protein. Further, two variants, p.Trp363Arg and p.Ala509Thr expressed in mouse embryos at E14.5 by in-utero electroporation, showed longer projections in the neuronal progenitors than the ones observed for the wildtype CDC20B allele. The projection lengths of the neuronal cells may influence micro-developmental aspects of neural circuits which in turn can predispose to epilepsy.

## 2.1 Materials and methods

### 2.1.1 Family ascertainment

A three-generation multi-affected family, GLH35 (Figure 1), was ascertained at Lourdes hospital, Cochin, through a 24-year-old proband affected with JME. The proband started to exhibit absence seizures (ABS) at the age of 11, followed by myoclonic seizures (MS) at 13, and generalized tonic-clonic seizures (GTCS) at 14. The major precipitating factors were sleep deprivation and emotional stress with the time of seizure occurrence within about two hours of waking up. EEG recordings showed frequent generalized epileptiform abnormalities, consistent with the diagnosis of JME. The proband was administered carbamazepine (CBZ), phenobarbitone (PB) and valproic acid (VPA) with VPA showing a good response. General physical and neurological examinations found no additional clinical manifestation in the proband.

All members of the family were followed up for 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 written consent. The diagnosis of JME was established as per the International League Against Epilepsy guidelines (Scheffer et al., 2017b). Proband's father (II:3), aunt (II:8), first cousin (III:1) and grandfather's sister (I:2) had MJ, ABS and GTCS. Proband's brother (III:3) had febrile seizures. Proband's uncle (II:5), aunt (II:9), grandfather (I:4) and grandfather's sister (I:3) had GTCS. Three more first cousins of the proband had febrile seizures, but their samples were not available for analysis. For individuals from the first generation (I:2, I:3, I:4) only clinical information could be obtained from the family, as the members had deceased. Additional five hundred and fifty-two unrelated JME patients participated in this study. For controls, 480

healthy individuals, apparently not related to each other and without any history of neurological disorder, were examined.

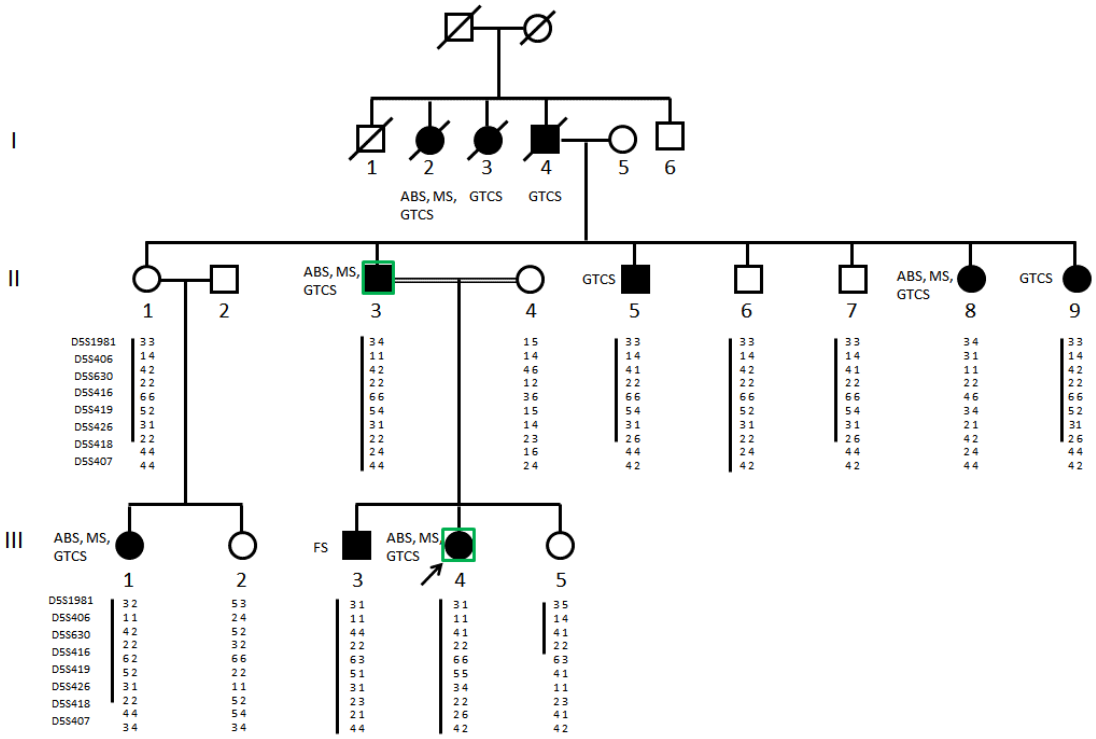



**Figure 1: GLH35 family pedigree:** Pedigree of GLH35 depicting the disease-linked haplotype (5p15-q12) shared among its six affected members (II:3, II:5, II:8, II:9) and three unaffected members (II:1, II:6 and II:7). Males are denoted by squares and females are denoted by circles. Filled symbols represent affected individuals and empty, unaffected ones. Haplotypes are shown below the symbols. Clinical features of affected subjects are indicated along with the symbols (AB: Absences, MS: Myoclonic seizures, GTCS: Generalized tonic-clonic seizures). The green boxes indicate the individuals, II:3 and III:4, whole exome or genome sequenced.

### 2.1.2 Genotyping and linkage analysis

Genomic DNA was isolated from the whole blood of the participating individuals using the phenol-chloroform method (Green and Sambrook, 2017). Thirteen participating individuals from the family were genotyped (Figure 1). Using 382 microsatellite markers, spaced over the genome with an average resolution of 10 centimorgans (cM), from the ABI Prism Linkage Mapping Set $V_{2.5}$ (Applied Biosystems, MA, USA), genome-wide linkage analysis was performed. Twelve additional markers (D5S417, D5S635, D5S1953, D5S2004, D5S1991, D5S1997,

D5S2074, D5S502, D5S1993, D5S1964, D5S1969, D5S664) were used for fine mapping at an average resolution of about 2 cM for chromosome 5. Following the manufacturer's protocol (Applied Biosystems), the microsatellite markers were amplified for the family members by individual PCR under the following conditions: 50 ng of genomic DNA, 5 µM of each primer, 9 µl of True allele PCR premix (Ampli*Taq*Gold DNA polymerase, buffer, magnesium chloride and dNTPs) in the reaction volume of 15 µl. The reactions were set up in a GeneAmp PCR System 9700 (Applied Biosystems) for 12 minutes at 95ºC (to activate the Ampli*Taq*Gold DNA polymerase); then for 15 seconds at 94 ºC, 15 seconds at 55ºC, and 30 seconds at 72ºC, for 10 cycles; and then, for 15 seconds at 89ºC, 15 seconds at 55ºC, and 30 seconds at 72ºC, for 25 cycles; followed by a final extension for 10 minutes at 72ºC. After amplification, PCR products from each mapping set were pooled with the GeneScan 500LIZ size standard (Thermo Fisher Scientific) and denatured along with formamide for 5 minutes at 95ºC followed by electrophoresis on the DNA Analyzer 3730 (Applied Biosystems). DNA from Centre d'E´ tude du Polymorphisme Human (CEPH) individual, 1347-02 was used as an internal control for each marker. The output data were analyzed for fragment sizing using Genescan software 3.7 (Applied Biosystems) followed by allele calling using Genotyper software 3.7 (Applied Biosystems). The Genethon and deCODE human linkage maps (Dib et al., 1996; Kong et al., 2002) were used to obtain the marker order and inter-marker distances. Each marker was analyzed manually for segregation in the family. Genotypes were checked for Mendelian inconsistencies if any, and haplotypes were generated manually, allowing for a minimum number of recombination events.

Parametric two-point LOD scores were calculated using the MLINK program of LINKAGE package5.2 (Lathrop et al., 1984). LOD score was calculated at different recombination fractions, using model parameters of an autosomal dominant mode of inheritance, 90% penetrance value,

1% phenocopy, mutant allele frequency of 0.0001, equal marker allele frequencies, and no difference in male and female recombination rates. Haplotypes were assigned manually based on genotyping data. Haplotyping was done to identify the shared genomic regions among affected individuals in the family, define the recombination boundaries and search for genotyping errors.

### 2.1.3 Whole-exome sequencing (WES)

#### 2.1.3.1 Library preparation and sequencing

Five micrograms of genomic DNA was fragmented (sonication at 55 pulses, 30s ON and 30s OFF) (Bioruptor-Diagenode, NJ, USA) and purified using Agencourt AMPure XP beads (Beckman Coulter, CA, USA). The target peak for base-pair size was 150-400 bp. The sheared DNA was analyzed for size distribution using Agilent DNA 1000 Bioanalyzer (Agilent Technologies, CA, USA). Successively, the sheared DNA fragments were used to construct DNA libraries using Agilent's SureSelectXT Target Enrichment System for illumina Paired-End Sequencing Library. The constructed library was modified by a series of steps using different enzymes to repair ends and make blunt-ended 5' phosphorylated fragments, add a single nucleotide A overhang and ligate 60bp sequence adaptors to fragment ends. Each step was followed by a purification step using Agencourt AMPure XP beads. After ligation, the adapter-ligated fragments were enriched by PCR and concentrated using a vacuum concentrator (Eppendorf, HB, Germany). The library was then hybridized to SureSelect biotinylated RNA baits at 65°C for 24-72 hours. Hybridized library fragments were isolated by magnetic capture using Dyna M-280 streptavidin-coated beads (Invitrogen, CA, USA) followed by purification of the capture library-bead solution using AMPure XP beads. PCR amplification was carried out to enrich the captured library and the amplified products were purified using AMPure XP beads. The Amplified Capture DNA was analyzed using the high sensitivity bioanalyzer chip which shows a

peak in the size range of 300-400 nucleotides (Figure 2). The SureSelect Human exome kit is designed to enrich 51Mb region in the genome. The sequencing was carried out for the captured libraries with illumine Genome Analyzer IIx platform (illumina, CA, USA) obtaining the 72 bp paired-end reads.

#### 2.1.3.2 Sequence data quality check (QC), processing and alignment

The whole-exome FASTAQ sequencing reads were aligned to human genome reference (hg19/GRch37) using BWA v-0.6.0 (Li and Durbin, 2009a). The reads showing at least 70% of bases with a minimum Phred score of 20, obtained by SeqQC v-2.0 (http:// genotypic.co.in/Products/7/Seq-QC.aspx), were used for alignment. Using SAMtools v-0.1.7a (Li et al., 2009), duplicate reads arising possibly from PCR artefacts, were removed. The variant calling was done using SAMtools at a Phred-like SNP quality score of 20. The variants identified were annotated by SNPeff and filtered against the dbSNP131. Novel variants were further examined in updated databases such as dbSNP139 (http:// www.ncbi.nlm.nih.gov/SNP/), 1000 Genomes (http://browser.1000genomes.org/index.html), Ensemble (http://asia.ensembl.org/index.html) and EVS datasets (http:// evs.gs.washington.edu/EVS/). To obtain potential variants, which may have been missed at high coverage, variants up to 3X read depth were manually examined. Those transcript regions which remained uncovered by the whole-exome sequencing were manually identified and examined by Sanger sequencing.

### 2.1.4 Whole-genome sequencing (WGS)

#### 2.1.4.1 Library preparation and sequencing

Three micrograms of genomic DNA was fragmented using Covaris to generate 300-400 base pair fragments. 100 ng of fragmented DNA was used to generate a sequencing library using NEBNext Ultra II DNA

Library Prep Kit for Illumina (New England Biolabs, MA, USA). In brief, the fragmented DNA was subjected to end repair followed by A – tailing and adapter ligation. Ampurebead-based size selection was performed to obtain the library of the desired size. The size selected DNA was enriched by PCR amplification using illumina index adapter primers (*P7 adapter 5'* – AGATCGGAAGAGCACACGTCTGAACTCCAGTCA – 3', *P5 adapter 5'* – AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT – 3'). The amplified product was purified using Ampure beads to remove unused primers. The libraries were quantitated using Qubit DNA HS quantitation assay (Thermo Fisher Scientific) which specifically quantitates dsDNA. The library quality was checked using Agilent Bioanalyzer DNA 1000 kit (Figure 2). The QC passed libraries were sequenced using Illumina HiseqX sequencer (illumina).

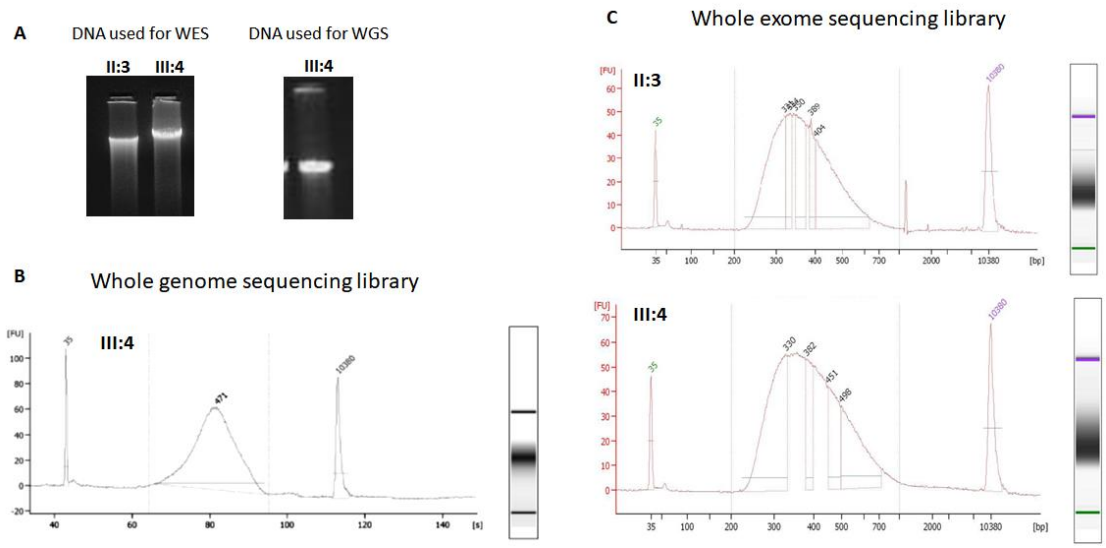



**Figure 2 – DNA and library quality check: A.** The genomic DNA was checked for its integrity and quality by electrophoresis on 0.8% TAE-agarose gel. This genomic DNA was used for library preparation. The libraries were checked for their quality and size distribution using Agilent Bioanalyzer DNA 1000 kit. **B.** Size distribution of the whole genome sequencing library made using the NEBNext Ultra II DNA Library Prep Kit for illumina. **C.** Size distribution of the whole-exome sequencing library made using Agilent's SureSelectXT Target Enrichment System for illumina Paired-End Sequencing Library.

### 2.1.4.2 Sequence data quality check (QC), processing and alignment

The sequence data quality was checked using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and MultiQC (Ewels et al., 2016). The data was checked for base call quality distribution,% bases above Q20, Q30,% GC, and sequencing adapter contamination. The sequence data was processed using Trimgalore (https://github.com/FelixKrueger/TrimGalore) to remove adapter sequences and low-quality reads. QC passed reads were mapped to human reference genome build GRCh38 that was provided in GATK Resource Bundle using BWA mem (Li and Durbin, 2009b) algorithm with default parameters. The alignments were sorted, indexed and PCR duplicates were marked and removed using sambamba (Tarasov et al., 2015).

### 2.1.4.3 Variant calling, annotation and filtering

GATK (McKenna et al., 2010) best practice workflow for germ line short variant discovery (SNPs + INDELs) using GATK v4 was followed for variant calling (https://software.broadinstitute.org/gatk/best-practices/). The process, in brief, base qualities in the bam file were recalibrated and GVCF was generated using HaplotypeCaller. GVCF was genotyped to generate vcf. SNPs and INDELs were recalibrated separately by providing variants from dbSNP, 1000 Genome phase1, hapmap, omni SNPs and Mills gold standard INDELs as known and training sets. The filter threshold was set to retain 99% of true variants, which provides the highest sensitivity while still being acceptably specific.

Functional annotation for the variants identified in the genome sequencing experiment was done using Ensembl Variant Effect Predictor (VEP) build 94 (McLaren et al., 2016). The input was provided in the form of a tab-delimited vcf file (Variant Call Format). Transcript database from both GENCODE and NCBI was used. Allele frequencies

were taken from dbSNP151, 1000 Genomes, Exome Aggregation Consortium (ExAC) and Genome Aggregation Database (gnomADv2, exomes and genomes; gnomADv3, genomes). dbSNFPv3.5 (Liu et al., 2011), a database developed for functional prediction and annotation of all potential non-synonymous single-nucleotide variants in the human genome, was also used for annotation. Samtools (Li et al., 2009) was used to view the whole genome sequence data aligned with the human reference sequence (Grch38) and each exon in the region of interest on chromosome 5 was manually checked for its coverage and read depth.

The identified variants were filtered for minor allele frequency in the databases, dbSNP151, 1000 Genomes project (Auton et al., 2015), Exome Aggregation Consortium (ExAC) (Lek et al., 2016) and the Genome Aggregation Database (gnomAD) (Karczewski et al., 2020) for their minor allele frequency (MAF) to be less than 0.005.

### 2.1.5 Sanger-based sequence validation of WGS rare variants

All novel/rare variants identified in the WGS dataset were validated by Sanger sequencing. Primers were designed spanning the variant-carrying exons/regions. The Sanger confirmed variants were further analysed in the family for their segregation with the disease phenotype. The segregating variants were examined in 488 ethnically matched control sets of normal individuals for their MAF.

### 2.1.6 Polymerase chain reaction (PCR)

Primers were designed for the exons/regions carrying the variants identified in the WGS analysis. The PCR conditions were standardized for each primer set and amplification was performed on a thermal cycler GeneAmpPCR System 9700 (Applied Biosystems). The PCR reaction mixture contained 10 mM Tris-HCl, 50 mM KCl, 1.5 mM $MgCl_2$, 0.8 mM dNTPs, 0.25 µM of each primer, 0.05 U/µl Taq Polymerase (New England Biolabs), deionized water and 10 ng/µl DNA in a 20 µl volume. The standardized amplification conditions were:

Initial denaturation at 95°C for 5 minutes, followed by 40 cycles of denaturation at 95°C for 30 seconds, annealing for 30 seconds and elongation at 72°C for 30 seconds, and a final extension step for 10 minutes at 72°C. The amplified products were electrophoresed on 1% agarose gel containing 2 µl Ethidium Bromide (2 µg/ml) and purified using Multiscreen 96 well PCR purification plates (Merck Millipore, MA, USA). The purified products were eluted in 20 µl of deionized water.

### 2.1.7 Sanger-based sequencing

PCR-amplified products were single strand amplified by cycle sequencing using 1 µl of BigDye Terminator v3.1 Cycle Sequencing reaction mix (Thermo Fisher Scientific), 1X sequencing buffer, 0.25 µM primer and 3 µl of purified PCR product in a 20 µl 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. It is followed by alcohol precipitation of the amplified product 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, followed by centrifugation at 2500 g for 30 minutes. The precipitated DNA was washed with 150 µl of 70% alcohol followed by 10-minute centrifugation at 2000 g. The plate was air-dried to remove all residual alcohol and the DNA denatured at 95°C in 10 µl formamide per well. These denatured single-stranded amplified products were Sanger sequenced using an automated DNA sequencer, DNA Analyzer 3730 (Applied Biosystems). The sequences thus obtained were aligned to the respective reference gene sequences obtained from the Genbank database and the variants were identified using SeqMan 5.01 (DNASTAR, WI, USA).

### 2.1.8 Minigene assay

To study the effect of the intronic variant, c.538-30C>T in *IL7R,* on splicing, the exon close to the intronic variant along with 500 base pairs upstream and downstream of the exon-intron boundary was cloned in the exon trapping vector pSPL3 (Figure4) (the vector was a gift from Stuart W. Tompson; Tompson and Young, 2017) using the primer pairs 5' – gtgtgaattcttaaagataatacactgtg – 3' and 5' – tcagtcagggatcctagggaggagatgtctcag – 3'. PCR was done in a 50 µl reaction volume containing 200µM dNTP, 0.4 µM of each primer, 100 ng of genomic DNA and 1 Unit Phusion high-fidelity DNA polymerase (New England Biolabs). Amplification was performed using a GeneAmp 9700 at following conditions: initial denaturation at 94°C (5 minutes), followed by 25 cycles of denaturation at 94°C (30 seconds), annealing at 58°C (30 seconds) and extension at 72°C (1 minute) followed by a final extension at 72°C (10 minutes). The amplified products were electrophoresed and visualized on a 0.8% agarose/TAE/EtBr gel (Figure4). The amplified minigene-insert was purified using Multiscreen 96 well PCR purification plates (Millipore). The 1080 bp minigene-insert was cloned into a pSPL3 mammalian expression vector, using EcoRI and BamHI restriction enzyme sites. Site-directed mutagenesis was performed using QuikChange II XL site-directed mutagenesis reagents (Agilent Technologies) to generate the patient-specific variants. These wild-type and mutant constructs were used for transfection in HEK293 cells. Total RNA was isolated from the cells using TRIzol (Thermo Fisher Scientific). cDNA synthesis was performed using the superscript III first-strand synthesis system (Thermo Fisher Scientific, 18080051). Primers V1 (V1-F: 5'-TCTGAGTCACCTGGACAACC-3') and V2 (V2-R: 5'-ATCTCAGTGGTATTTGTGAGC-3') were used to amplify the minigene. The amplified products were electrophoresed and visualized on a 0.8% agarose/TAE/EtBr gel (Figure 4). The products were Sanger sequenced to confirm any splicing errors.

### 2.1.9 Luciferase assay

To investigate the effect of the rare 3' UTR variants, c.*1664A>G in *PRLR*, c.*1770A>G in *LMBRD2*, c.*620G>A in *SNX18*, c.*2812_*2813del in *SNX18*, c.*3364A>G in *SNX18*, on predicted miRNA binding sites and gene expression, 400 base pairs of the region harbouring each of the variants was cloned in the 3'UTR of the luciferase reporter gene in the pMIR-report luciferase vector (Figure5). PCR amplification of the region of interest was done in a 50 µl reaction volume containing 200 µM dNTP, 0.4 µM of each primer, 100 ng of genomic DNA and 1 unit Phusion high-fidelity DNA polymerase (New England Biolabs). Amplification was done using a GeneAmp 9700 at following conditions: initial denaturation at 94°C (1 minute), followed by 25 cycles of denaturation at 94°C (30 seconds), annealing at 55°C (30 seconds) and extension at 68°C (3 minutes) followed by a final extension at 72°C (7 minutes). The amplified products were electrophoresed and visualized on a 0.8% agarose/TAE/EtBr gel (Figure 5). The 3' UTR insert was purified using Multiscreen 96 well PCR purification plates (Millipore, LSKMPCR50). These 3' UTR insert fragments were cloned into a pMIR-report luciferase vector, using SpeI and MluI restriction enzyme sites. Site-directed mutagenesis was done to generate the variants. These constructs, along with pMIR-report β-gal control plasmid, were used to transfect HEK293 cells grown in a 12-well dish.

48 hours post-transfection, luciferase activity was measured using the luciferase reporter assay system (Promega, WI, USA). The cells were washed with 1X PBS and lysed in 100ul 1X reporter lysis buffer. The cells were incubated on ice for 1 hour with intermittent vortexing. The lysed cells were centrifuged at 15,000 rpm, 4$^{o}$C for 30 minutes. The supernatant is transferred to a fresh tube. 5 ul of lysate is added to 10 ul luciferase assay reagent, mixed by vortexing and readings taken using Sirius single tube illuminometer (Berthold Detection Systems GmbH, Germany).

β-galactosidase enzyme activity was measured to normalize for transfection efficiency. 20 ul of 5 times diluted lysate was mixed with 20 ul of 2X β-galactosidase assay buffer (Promega) that contains the substrate ONPG (ortho-Nitrophenyl-β-galactoside) in a microplate reader. The plate was incubated at 37°C for 30-40 minutes. The reaction was stopped by adding 50 ul 1M sodium bicarbonate and readings were taken using an ELISA plate reader at 420 nm.

The luciferase readings and β-galactosidase readings for each sample were divided by the average of all samples to get normalized values. To normalize for transfection efficiency, the normalized luciferase readings were then divided by the respective normalized β-galactosidase readings. These values were plotted using GraphPad Prism5. The student's unpaired t-test was used to calculate significance for each set of wild type and respective variant constructs. For SNX18-3'UTR-b,c construct, one-way-ANOVA followed by Dunnett's multiple comparison test was used as we were comparing three groups.

### 2.1.10 CDC20B mutation analysis

The gene *CDC20B*, which carried the disease co-segregating rare variant, was further analysed in a set of 552 JME affected individuals, and 488 ethnically matched control individuals. The gene sequence comprising all exons, flanking intronic boundaries, and the 5'- and 3'-untranslated regions (Figure 7) were Sanger sequenced to identify additional rare variants.

### 2.1.11 Bioinformatics analysis

A list of all genes present at the 5p15-q12 locus was downloaded from Mapviewer-NCBI (now discontinued). For each gene in this locus, details of the genomic position of all exons and introns were also retrieved. These genomic positions were used to manually check the coverage of each exon in the locus, using SAMtools-tview.

The complete gene sequence of *CDC20B* and sequences for genes harbouring patient variants were downloaded from NCBI (NC_000005.10:c55173177-55112971, Homo sapiens chromosome 5, GRCh38.p13 Primary Assembly). Primers were designed using the online tools Primer3 (Untergasser et al., 2012) and OligoCalc (Kibbe, 2007). Protein sequences for CDC20B across different species were downloaded from NCBI to perform multiple sequence alignments using the tool Clustal Omega (Madeira et al., 2019).

### 2.1.12 Plasmids and antibodies

The CDC20B cDNA (Genscript, NJ, USA, OHu14900) was sub-cloned into pEGFPN1 (enhanced GFP tag fused to C-term of the protein), p3X-FLAG-CMV10 (3X FLAG tag fused to the N-term of the protein) and pcDNA3.1+ (untagged) vectors. Site-directed mutagenesis was performed using QuikChange II XL Mutagenesis reagents (Agilent Technologies) to generate patient-specific variants {c.282G>T (p.Gln94His), c.598A>G (p.Arg200Gly), c.635T>A (p.Ile212Lys) and c.1087T>C (p.Trp363Arg)}. To study CDC20B domains, different regions were either sub-cloned or premature stop sites were introduced in the cDNA (Figure 12). pEYFP-Plk1 (39843) was obtained from Addgene (Watertown, USA). p3X-FLAG-CMV10-EFHC1 was available in the laboratory.

The antibodies used in this study were: anti-CDC20B (polyclonal, rabbit raised; Abcam, Cambridge, UK, ab185481 and Proteintech, IL, US, 13376-1-AP), anti-CDC20B (monoclonal, mouse raised; Sigma, MO, US, SAB1402036), 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 Fluor 488 (Thermo Fisher Scientific, A11008), goat anti-mouse Alexa Fluor 568 (Thermo Fisher Scientific, A11001),

goat anti-rabbit IgG – HRP (Genei, KA, India, HPO3), goat anti-rabbit IgG – HRP (Genei, HPO5).

### 2.1.13 Cell culture and transfection

HeLa cells, HEK-293 cells, U-87 MG cells, SH-SY5Y cells and NIH/3T3 cells were maintained in DMEM (high glucose, Sigma) supplemented with 10% heat-inactivated fetal bovine serum (MP Biomedicals, CA, USA), 2 mM L-glutamine (Sigma) and antibiotics (100 U/ml Penicillin and 10 mg/ml streptomycin; Sigma) 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) coated 18 mm glass coverslips and transiently transfected with 1 µg plasmid per well using Lipofectamine 2000 transfection reagent (Thermo Fischer Scientific). HeLa cells were transfected using jetPRIME transfection reagent (Polyplus-transfection, France). Protein expression was checked 24 to 48 hours post-transfection.

### 2.1.14 RNA isolation

Glassware to be used during RNA isolation and MilliQ water to be used for making solutions/reagents were Diethyl Pyrocarbonate (DEPC) treated and double autoclaved. TRIzol reagent was added to the cells and incubated 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 12,000 g at 4°C. The mixture separated into a lower red phenol-chloroform, interphase, and a colourless 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 for 10 minutes at 12,000 g at 4°C. Total RNA precipitate forms 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 7500 g 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 water bath set at 55–60°C for 10–15 minutes for dissolution.

### 2.1.15 cDNA synthesis

The SuperScript III First-Strand Synthesis System (Invitrogen, MA, USA) for RT-PCR (reverse transcriptase-polymerase chain reaction) was used to synthesize first-strand cDNA from total RNA. 1 µg total RNA, 5 µM oligo$(dT)_{20}$, 1 mM dNTP mix and DEPC treated water to makeup volume to 10 µl were combined in a 0.2 ml PCR tube and incubated at 65$^{o}$C for 5 minutes followed by transferring to ice. The cDNA synthesis mix comprising, 2-X RT buffer, 10 mM $MgCl_2$, 0.02 M DTT, 40 units of RNase OUT enzyme and 200 units of SuperScript III RT enzyme, was prepared and 10 µl 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, followed by chilling on ice. 1 µl RNase H was added to each tube and the reaction was incubated at 37$^{o}$C for 20 minutes. The synthesized cDNA was stored at -20$^{o}$C.

### 2.1.16 Western analysis

For isolating whole-cell protein, the cells/tissues were lysed using lysis buffer (150 mM NaCl, 10 mM Tris-pH 7.5, 0.1% SDS, 1% Triton X-100, 1% deoxycholate and 5 mM EDTA). The total protein content of the cells was collected by centrifuging the lysed cells at 13,000 rpm for 20 minutes at 4$^{o}$C. Total protein content was quantified using the Pierce BCA protein assay kit (Thermo Fischer Scientific). 50 µg of total protein was separated on a 10% SDS-polyacrylamide gel (Laemmli, 1970) followed by transfer to nitrocellulose membrane (Pall, NY, USA) at

20 volts for 1 hour. The membrane was blocked with 5% skimmed milk for 4 hours at 4°C followed by incubation with primary antibodies, anti-CDC20B (1:500) or anti-gamma tubulin (1:5000) or anti-alpha tubulin (1:5000) or anti-FLAG (1:500) or anti-GFP (1:2500), at 4°C for 16 hours. HRP-conjugated secondary antibodies (1:5000) were used to probe the membrane at 4°C for 4 hours. The membrane was washed between each incubation step with 0.05% PBST (0.05% tween-20 in 1X PBS). Clarity western ECL chemiluminescence substrate (Biorad, CA, USA) was used to detect the protein bands on photographic films.

### 2.1.17 Immunocytochemistry

To perform immunocytochemistry analysis, cells were grown on 18mm poly-L-lysine coated glass coverslips. The cells were fixed with 2% paraformaldehyde (PFA) for 15 minutes at room temperature or with ice-cold methanol (for gamma-tubulin staining) at -20°C for 5 minutes. The fixed cells were permeabilized using 0.1% Triton X-100 for 10 minutes at room temperature and then blocked with 3% BSA (bovine serum albumin) for 1 hour at room temperature. Incubations with primary antibodies, anti-CDC20B (1:500) and anti-gamma tubulin (1:5000), and fluorescently labelled secondary antibodies (1:500), diluted in 1% BSA, were done for 1 hour each at room temperature. Nuclear staining was performed using 1µg/ml 4, 6-diamidino-2-phenylindole (DAPI) for 15 minutes at room temperature. The coverslips were mounted on glass slides using Polyvinyl alcohol mounting medium (Sigma, 10981) and sealed with transparent nail polish. Imaging was done using Zeiss LSM 510-meta/-880 confocal microscope. Zeiss LSM 880 with airyscan mode was used for super-resolution microscopy.

### 2.1.18 Immuno-pulldown assay

Total protein was isolated from cells at 4°C using mild lysis buffer (150 mM KCl, 25 mM tris-pH 7.4, 5 mM EDTA, 0.5% Np40, protease inhibitor cocktail (Roche, 05892791001), 1 mM PMSF). 20 µl

Dynabead protein G mix (Invitrogen) was allowed to bind with 2 µg anti-FLAG antibody, 2 µg anti-GFP anti-body or 2 µg anti-CDC20B monoclonal antibody for 4 hours with rotation (10 rpm) at 4°C. The bead-antibody complex was then incubated with 2 mg protein lysate for 12 hours at 4°C. Post-incubation, the bead-antibody-antigen complex was washed 4 times for 1 minute each, on ice, with 200 µl cold 1 X PBS. Elution was done by boiling the beads in 50 µl 2X dye for 10 minutes. The eluted protein fraction was separated from the magnetic beads and stored in a fresh tube at -20°C or loaded on an SDS-PAGE gel for further analysis.

### 2.1.19 In-utero electroporation

C57BL/6 pregnant mice were anaesthetized using isoflurane. The abdominal area of the mice was sterilized with 70% alcohol and shaved using a blade. About a 1 cm long incision was made in the abdomen of the pregnant mice and the embryos were pulled out using ring forceps. The embryos were kept hydrated in warm 1X PBS. A solution containing 2 µg/µl of EGFPN1-CDC20B wildtype, Trp363Arg mutant, Ala509Thr mutant or EGFPN1-empty vector with 1ul of fastgreen (Sigma) was injected into the lateral ventricle of embryonic brains at E14.5 and electroporated with a 5mm forceps electrode, which transmitted five electric pulses at 50 V for 50 ms at 1-second interval through the uterine wall. The embryos were pushed back into the abdomen and the incision was stitched using a stainless-steel surgical suture curved needle and silk suture. The mice were kept in a recovery chamber for 1-2 hours before shifting to their cage. The embryos were harvested 24-48 hours post electroporation. The brains were removed and transferred to tubes containing ice-cold 4% PFA for tissue fixation. These brains were stored in PFA at 4°C for 24 hours.

### 2.1.20 Cryosectioning

The PFA fixed brains were transferred to 15% sucrose for 24 hours at 4°C followed by 30% sucrose for another 24 hours at 4°C. The brains

were mounted on the chuck of a cryosectioning machine (Leica Biosystems, IL, USA) in a freezing medium and allowed to freeze for 20 minutes at -20°C. 20-micron thick sections were collected on gelatin-coated glass slides and stored at -20°C until staining.

### 2.1.21 Immunohistochemistry

The slides were removed from the -20°C freezer and allowed to thaw at room temperature for 30 minutes. The freeze media was washed off by dipping the slides in 1X PBS for 15 minutes. The sections were permeabilized in 0.05% PBS-tween20 solution for 30 minutes at room temperature. Blocking was done with 10% NGS for 1 hour at room temperature. This was followed by staining with anti-GFP antibody (Thermo Fischer Scientific, 1:1000 made in 1% NGS) at 4°C for 16 hours and fluorescently tagged secondary antibody (1:400 in 1%NGS) at 4°C for 4 hours. The slides were then incubated with 1ug/ml DAPI solution for 30 minutes at room temperature. The slides were dipped in 1X PBS and allowed to dry completely. The dried slides were then dipped in water and allowed to dry before mounting in 70% glycerol. Imaging was done using Zeiss LSM 880 microscope.

## 2.2 Results

### 2.2.1 Linkage analysis

In the genome-wide linkage analysis, parametric two-point LOD scores were calculated at 60-90% penetrance values. The highest two-point LOD score of 1.7 was obtained at recombination fraction0, for D5S426 at 5p13.2, at 90% penetrance value and 1% phenocopy (Table 1). No suggestive evidence of linkage was found for any marker elsewhere in the genome on searching for regions where LOD score was more than 1.0 and the marker co-segregating among all affected individuals.

Haplotype analysis showed a set of 12- marker disease haplotype block shared among all affected members, except II:8 and asymptomatic

carriers. The proximal and distal boundaries of the linked region were marked by D5S1981 and D5S407, respectively, spanning a 64Mb genomic region (Figure 1).

**Table1: Two-point LOD analysis for the 5p15-q12 markers at different θ values in Family GLH35**

| Marker | Genetic Position (cM) | Physical Position (Mb) | LOD at the theta value | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | 0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
| D5S416 | 27.9 | 16.7 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| D5s1997 | 29.7 | 16.9 | -1.55 | -0.72 | -0.36 | -0.15 | -0.04 | 0.00 |
| D5s2074 | 35.6 | 21 | -0.02 | -0.01 | -0.01 | 0.00 | 0.00 | 0.00 |
| D5s502 | 38.9 | 25.5 | 0.01 | 0.13 | 0.13 | 0.08 | 0.03 | 0.00 |
| D5S419 | 39.5 | 26.5 | 0.88 | 0.70 | 0.50 | 0.29 | 0.09 | 0.00 |
| D5s1993 | 44.4 | 31.6 | 0.01 | 0.13 | 0.13 | 0.08 | 0.03 | 0.00 |
| **D5S426** | **51.6** | **34.7** | **1.70** | **1.39** | **1.05** | **0.68** | **0.30** | **0.00** |
| D5s1964 | 54.4 | 37.8 | -0.07 | 0.04 | 0.07 | 0.05 | 0.01 | 0.00 |
| D5S418 | 58.1 | 39.9 | 1.69 | 1.38 | 1.04 | 0.67 | 0.29 | 0.00 |
| D5s1969 | 61.1 | 50.2 | 0.89 | 0.70 | 0.50 | 0.28 | 0.08 | 0.00 |
| D5s664 | 63.8 | 51.9 | 0.95 | 0.76 | 0.55 | 0.32 | 0.10 | 0.00 |
| D5S407 | 65 | 52.9 | 0.88 | 0.70 | 0.50 | 0.28 | 0.09 | 0.00 |

Marker order is as per the human genome physical map (Human Genome Map Viewer Build 36.3 database, NCBI, NIH, USA)

### 2.2.2 Whole-exome sequencing

To examine the 5p15-q12 region in detail, whole-exome-based sequencing was undertaken in two affected members of the GLH35 family, II:3 and III:4. This region spans 64 Mb of the sequence length and encodes 177 protein-coding genes. In the whole-exome sequencing experiment, a total of 9.01 GB sequence, as 41.58 million paired-end reads of 72 bp, was generated from sequencing on the GAIIx machine, where more than 98% of bases were of high quality and total coverage of 99% was obtained. The average read depth was >60X. The region of our interest has 2255 coding exons and 570 non-coding exons out of which 136 coding exons and 415 non-coding exons were not covered by the whole-exome sequencing. These uncovered segments were examined by Sanger sequencing.

**Table 2: Sequence coverage summary for the whole-exome sequencing experiment (WES)**

| Sample | II:3 | | III:4 | |
|---|---|---|---|---|
| **Target** | **Whole Exome Probes Targeted** | **chr5 – Exome Probes Targeted** | **Whole Exome Probes Targeted** | **chr5 – Exome Probes Targeted** |
| Total number of reads | 83.15 Million | | 76.91 Million | |
| Reads aligned to complete genome | 82.95 Million | | 76.75 Million | |
| %Reads aligned to complete genome | 99.76 | | 99.78 | |
| Reads aligned to the target region | 67.09 Mb | 688 Kb | 63.83 Mb | 711 Kb |
| %Reads aligned to target region | 80.69 | 0.83 | 82.99 | 0.93 |
| Target Sequence Length | 51.54 Mb | 529 Kb | 51.54 Mb | 529 Kb |
| Total target covered | 50.68 Mb | 521 Kb | 50.60 Mb | 521 Kb |
| **%Total Target covered** | **98.34** | **98.45** | **98.19** | **98.52** |
| %Total Target covered with at least 5X Read Depth | 94.22 | 95.43 | 94.01 | 95.62 |
| %Total Target covered with at least 10X Read Depth | 89.68 | 91.93 | 89.86 | 93.57 |
| %Total Target covered with at least 15X Read Depth | 85.33 | 88.02 | 85.97 | 91.50 |
| %Total Target covered with at least 20X Read Depth | 80.94 | 83.85 | 81.91 | 88.69 |
| **Average Read Depth** | **67.94** | **66.42** | **65.41** | **69.27** |

Percentage of region coverage after sequences aligned to Human genome reference sequence (hg19/Grch37)

### 2.2.3 Whole-genome sequencing

Whole-exome sequencing does not cover untranslated regions (UTR) and some of the protein-coding exons too. To exhaustively cover all important regions of the genome, whole-genome sequencing was carried out on DNA from one member of the family GLH35, III:4. In the whole-genome sequencing experiment, a total of 180 GB sequence, as 1.2 billion paired-end reads of 150 bp, was generated from sequencing on the illumine Hiseq X (illumina). More than 97% of bases were of high-quality (Phred score > 20) and an average read depth of 52X was obtained (Table 3). A total of 90,928 variants were identified out of which 1,497 variants were either absent or present at a MAF<0.005 across the databases dbSNP151, 1000 Genomes, ExAC and the gnomAD.

**Table 3: Sequence coverage summary for the whole-genome sequencing experiment (WGS)**

| **Sample** | **III:4** | |
|---|---|---|
| **Fields** | **Genome** | **5p15-q12 Region** |
| Total number of reads | 1.21 Billion | |
| Reads aligned to complete genome | 1.20 Billion | |
| **%Reads aligned to complete genome** | 98.77 | 98.84 |
| Target Sequence Length | 3.2 Gb | 61 Mb |
| Total target covered | 3.0 Gb | 58 Mb |
| **% Target Covered** | **94.19** | **95.78** |
| % Target Covered with at least 5X read depth | 93.62 | 95.45 |
| % Target Covered with at least 10X read depth | 92.82 | 95.1 |
| % Target Covered with at least 15X read depth | 91.75 | 94.63 |
| % Target Covered with at least 20X read depth | 90.05 | 93.75 |
| % Target Covered with at least 30X read depth | 83.72 | 88.96 |
| **Average read depth** | **52.53** | **52.77** |

Percentage of genome/ region coverage calculated after sequences were aligned to the human genome reference sequence (GRCh38).

### 2.2.4 Rare variant analysis and prioritization

Of the 1497 rare variants identified at the 5p15-q12 locus, 480 lie in protein-coding genes and 1423 in non-protein-coding regions such as micro-RNAs, promoters, enhancers and transcription factor binding sites. A variation at the same genomic position can have multiple annotations, which is the reason for the total number of rare variants being 1497 and not 1903 (480 + 1423). The variants in the non-protein-coding regions were first checked for their conservation using the CADD score. Variants with CADD<20 were removed from our analysis. The variants lying in conserved regions were then analysed for their pathogenicity using the FATHMM-MKL prediction tool that combines sequence conservation scores with ENCODE-based functional annotations (Shihab et al., 2015). The variants were also manually checked in the UCSC genome browser to see if they lie in repeat regions that are prone to sequencing errors. None of the variants in the non-protein-coding regions were found to have any pathogenic effects and were hence not taken forward for further analysis.

The variants in protein-coding transcripts (449) were further classified based on their consequences. There were 424 variants in intron regions, 19 in UTR regions, 27 in regions upstream or downstream of the genes, 4 variants gave rise to missense changes and 1 variant to a synonymous change in the protein sequence. Variants in low complexity regions, which are rich in polymorphisms, regions beyond 100 base pairs from the exon-intron boundary and variants in regions upstream or downstream of genes were excluded from further analysis (Figure 3). Some variants can lie in non-protein-coding regions, such as micro-RNAs, promoters, enhancers and transcription factor binding sites. These variants were checked for their conservation using the CADD score, with variants with CADD<20 removed from further analysis. The remaining variants were manually checked in the UCSC genome browser for their conservation, and we found that none of the

variants lie in evolutionarily conserved regions of the genome. Hence, these were not taken forward for further analysis.

Segregation status was checked for the rare variants identified in the protein-coding genes in the WES/WGS analysis. Ten variants, c.*2135G>A in *C1QTNF3*, c.*1664A>G in *PRLR*, c.538-30C>T in *IL7R*, c.*1770A>G in *LMBRD2*, c.8044G>C in *CPLANE1*, c.2413A>G in *NUP155*, c.*620G>A in *SNX18*, c.*2812_*2813del in *SNX18*, c.*3364A>G in *SNX18*, c.1513G>A in *CDC20B*, co-segregated with the JME phenotype in the family. The variant, c.*2135G>A in *C1QTNF3*, was present in ethnically matched normal control individuals and was therefore not taken forward for analysis. Conservation and in-silico predictions were performed for the remaining nine variants (Table 4). Minigene assay was performed for the intronic variant c.538-30C>T in *IL7R*. 3'UTR variants, c.*1664A>G in *PRLR*, c.*1770A>G in *LMBRD2*, c.*620G>A in *SNX18*, c.*2812_*2813del in *SNX18*, and c.*3364A>G in *SNX18* were examined by performing luciferase assays.

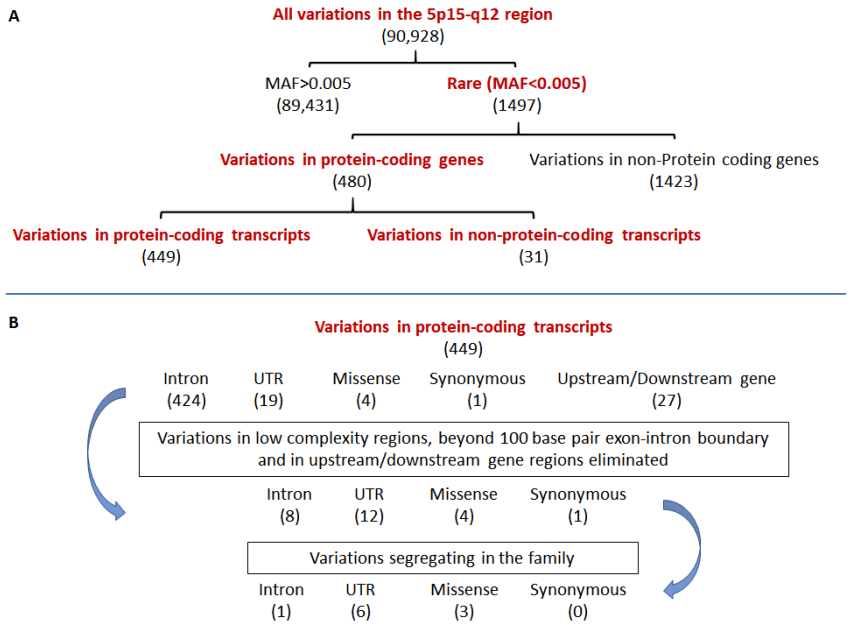



**Figure 3: Whole-exome/genome sequencing variants filtration and prioritization flow chart.** The variants identified in the 5p15-q12 locus are first filtered based on their minor allele frequency across databases. Variants with frequency less than 0.005 are further filtered based on their annotation and their segregation with the disease phenotype in the family.

**Table 4: Rare variants identified by whole-genome sequencing in the 5p15-q12 region in GLH35 family**

| Gene[a] | Consequence[b] | HGVSc[c] | HGVSp[d] | rs ID[e] | In-house controls[f] | MAF gnomAD v2.1[g] | SIFT[h] | Polyphen2[i] | Mutation Taster[j] | HSF (splice site)[k] |
|---|---|---|---|---|---|---|---|---|---|---|
| *PRLR* | 3' UTR variant | c.*1664A>G | - | Novel | 0/192 | _ | - | - | Polymorphism (0.99) | - |
| *IL7R* | intron variant | c.538-30C>T | - | rs1165054639 | 0/192 | 0.000004 | - | - | Polymorphism (0.99) | Potential alteration of splicing |
| *LMBRD2* | 3' UTR variant | c.*1770A>G | - | Novel | 0/192 | _ | - | - | Disease-causing (0.99) | - |
| *CPLANE1* | missense variant | c.8044G>C | p.Gly2682Arg | rs768671286 | 0/480 | 0.000003 | Tolerated (0.148) | Benign (0.002) | Neutral (1) | - |
| *NUP155* | missense variant | c.2413A>G | p.Ile805Val | rs376806446 | 0/480 | 0.00003 | Tolerated (1) | Benign (0) | Neutral (0.99) | - |
| *SNX18* | 3' UTR variant | c.*620G>A | - | Novel | 0/192 | _ | - | - | Disease-causing (0.99) | - |
| *SNX18* | 3' UTR variant | c.*2812_*2813del | - | Novel | 0/192 | _ | - | - | Polymorphism (0.99) | - |
| *SNX18* | 3' UTR variant | c.*3364A>G | - | rs186817959 | 0/192 | _ | - | - | Disease-causing (0.77) | - |
| *CDC20B* | missense variant | c.1513G>A | p.Ala509Thr | rs200898156 | 0/480 | 0.0002 | Damaging (0.026) | Probably damaging (1) | Damaging (0.93) | - |

[a]Gene harbouring the variant, [b]Consequence of the variant, [c]Nomenclature of the variants with respect to the 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 the missense variants, [f]These variants were checked in in-house, ethnically matched control individuals (192 or 480 individuals) by Sanger sequencing, [g]Minor allele frequency of these variants in gnomAD version 2.1, [h,i,j,k]predictions for pathogenicity of the variant.

### 2.2.5 Minigene splicing assay to analyse the intronic variant

Minigene assay was undertaken to study the effect of the intronic variant c.538-30C>T in *IL7R* on splicing. The exon close to the intronic variant along with 500 base pairs upstream and downstream of the exon-intron boundary was cloned in the exon trapping vector pSPL3 (Figure 4). The wild type and variant minigenes were transfected in HEK293 cells. Total RNA isolation was done 48 hours post-transfection and cDNA synthesis was done using a superscript III first-strand synthesis system. The primers V1 and V2, which lie in the vector-specific exons V1 and V2, were used to amplify the minigene. If splicing takes place, the intron between the exons V1 and V2 gets spliced out. In the minigene that we have constructed, under wild-type conditions, the *IL7R* exon will also be present between the vector-specific exons V1 and V2. We find that the variant minigene behaves the same as the wild type and there is no indication of a splicing error (Figure 4)

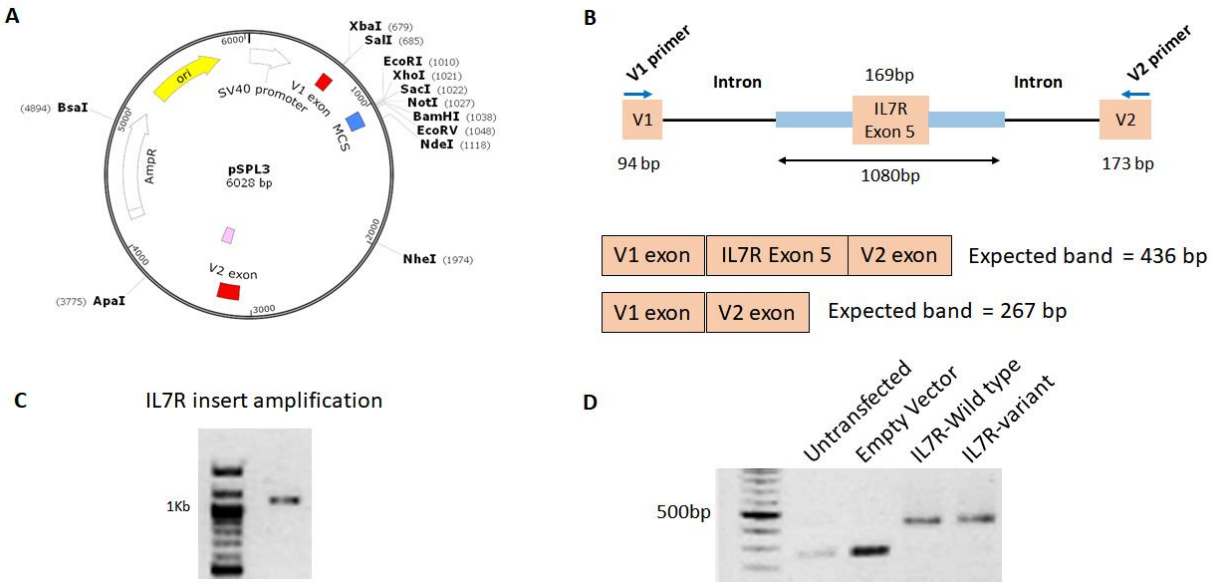



**Figure 4: Schematic of IL7R mini-gene assay. A.** Plasmid map of the pSPL3 vector. **B.** Schematic of the IL7R minigene. V1 and V2 are the exons present in the vector. IL7R exon 5 with the spanning intronic regions are cloned between V1 and V2. A product of 436 base pairs is expected when splicing takes place. **C.**

PCR amplification of the IL7R region of interest harbouring the intronic variant from genomic DNA. **D.** RT-PCR analysis of cDNA from wildtype (WT) and variant-harbouring minigene transfected cells.

### 2.2.5 Luciferase assay to study the effect of the 3'UTR variants on miRNA regulation

To investigate the effect of the rare 3'UTR variants, c.*1664A>G in *PRLR*, c.*1770A>G in *LMBRD2*, c.*620G>A in *SNX18*, c.*2812_*2813del in *SNX18*, and c.*3364A>G in *SNX18*, on predicted miRNA binding sites and gene expression, 400 base pairs of the region harbouring the variant was cloned into the 3'UTR of the luciferase reporter gene in the vector pMIR-report luciferase vector (Figure 5). The patient variants were generated by site-directed mutagenesis. The wild type and variant constructs were transfected in HEK293 cells and luciferase readings were taken 48 hours post-transfection. The luciferase readings normalized against β-galactosidase enzyme activity were plotted to compare the wildtype and variant readings. We found no significant difference in the luciferase readings between the wildtype and variant carrying luciferase constructs (Figure 5).

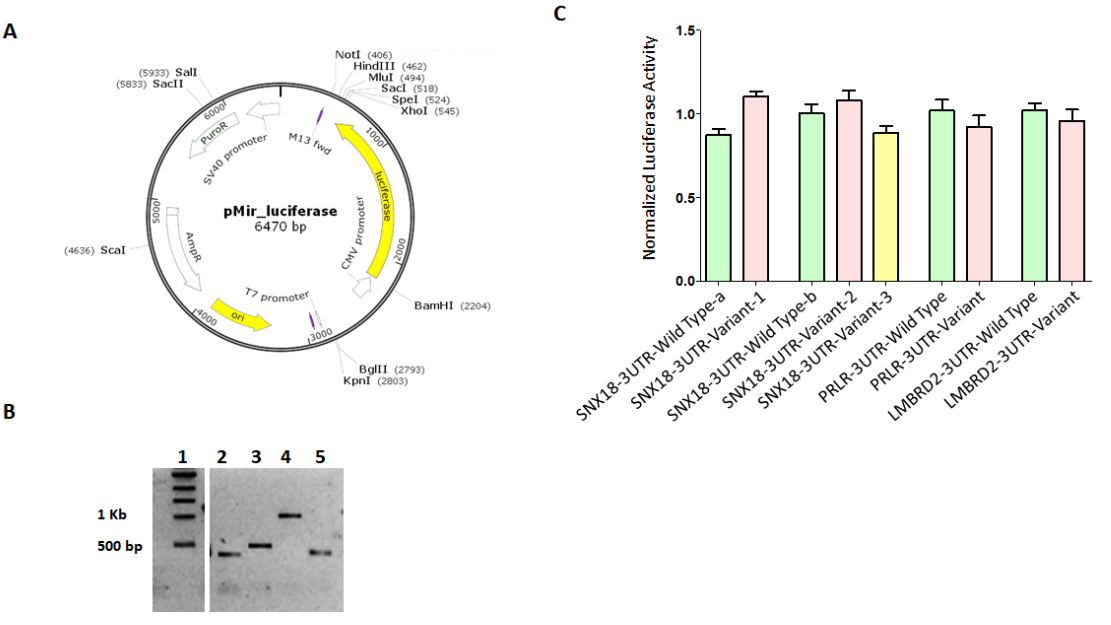


**Figure 5: Luciferase assay for variants in miRNA binding sites.** A. Plasmid map of the pMIR-luciferase vector. B. PCR amplification of the region of interest harbouring the 3'UTR variants in genes *SNX18, PRLR* and *LMBRD2* from genomic DNA. D. Normalized luciferase readings have been plotted for wildtype and variant harbouring constructs. No statistically significant difference in luciferase activities between the wildtype and variant constructs is observed. The experiment was done 4 times.

### 2.2.6 Identification of CDC20B rare and conserved missense variant

The intronic variant c.538-30C>T in *IL7R* and the 3'UTR variants, c.*1664A>G in *PRLR*, c.*1770A>G in *LMBRD2*, c.*620G>A in *SNX18*, c.*2812_*2813del in *SNX18*, and c.*3364A>G in *SNX18*, showed no effect on splicing and gene expression levels, respectively. The three variants identified in the coding exons, c.8044G>C in *CPLANE1*, c.2413A>G in *NUP155*, and c.1513G>A in *CDC20B* are missense variants: p.Gly2682Arg, p.Ile805Val and p.Ala509Thr, respectively. The variants in the genes *CPLANE1* and *NUP155* were found to be poorly conserved across species and the majority of in-silico tools predicted these variants to be benign or polymorphisms (Table 4).

The only variant which was conserved across species and showed deleterious effects on the proteins function employing in-silico predictions by different tools was c.1513G>A in *CDC20B* (Table4). This variant gives rise to a non-synonymous change, p.Ala509Thr, in the protein. Multiple sequence alignment for CDC20B across species shows that the residue alanine at the 509$^{th}$ position is highly conserved (Figure 6).

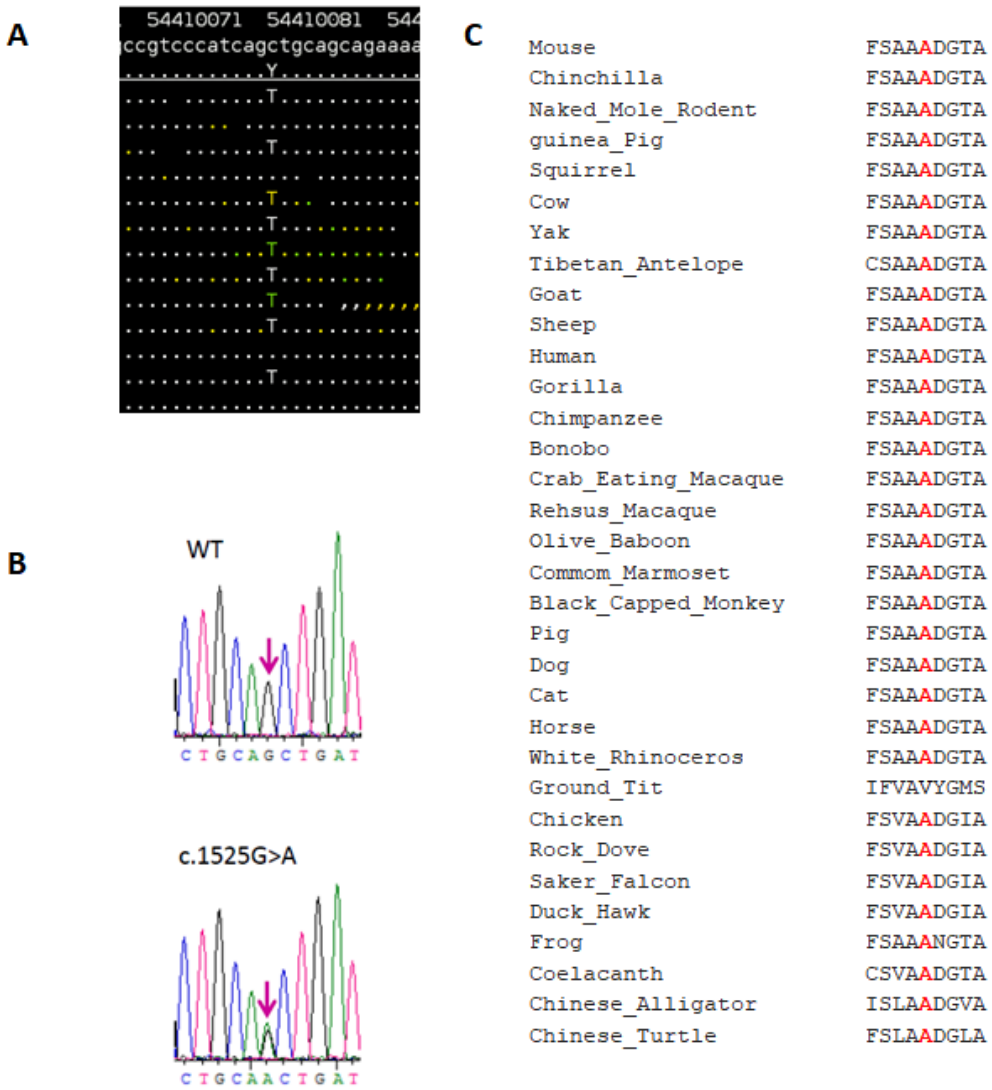



**Figure 6: The c.1525G>A (p.Ala509Thr) variant in *CDC20B*: A.** Alignment of the genome sequence reads against human genome reference sequence as viewed in tview of SAMtools v0.1.7a. The snapshot here shows that not all reads have the C>T variant, which indicates that this variant is heterozygous. The dot and comma represent the matched reference allele and the alphabet shows the variant allele. Note: *CDC20B* gene lies on the negative strand of the chromosome, hence the variant in the gene is G>A but is shown in the alignment as C>T. **B.** Electropherograms of sequence with wild-type (G/G) and variant allele (G/A) for the c.1525G>A variant. The arrow indicates the nucleotide showing heterozygous variant shown by the presence of two peaks. **C.** Multi-species protein sequence alignment of CDC20B by Clustal Omega. The alanine (A) 509 is highlighted in red.

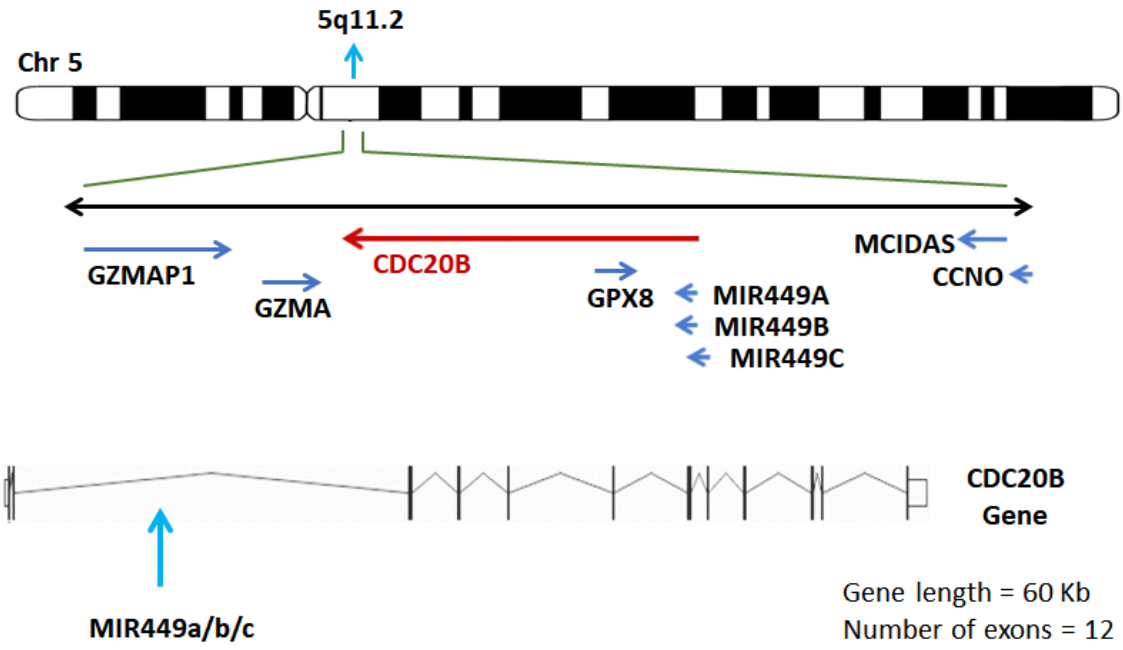



**Figure 7: Gene architecture of *CDC20B*.** The cytogenetic location of *CDC20B* and the schematics of exon-intron structure is depicted here. It is encoded on the reverse strand of chromosome 5. Other genes spanning the *CDC20B* locus have been represented. *CDC20B* gene is 60 kilobases in length and has 12 exons. The second intron of *CDC20B* harbours miRNA449 cluster (blue arrow).

### 2.2.7 *CDC20B* rare variants among JME patients examined

*CDC20B* gene encompasses 60 kilobases of the genomic region and harbours 12 exons. CDC20B encodes a 519 amino acid-long cell division cycle 20 homolog B protein. It belongs to the WD-repeat family and has 7 WD40 repeats at its C-terminus. The location of the *CDC20B* gene on chromosome 5, genetic architecture is represented in Figure 7.

The complete transcript of *CDC20B* was examined in 552 JME patients and 488 ethnically-matched control individuals. In this analysis, 25 rare variants, 13 of which are novel, were identified in 42 patients (Table 5). These variants were either absent or rare (MAF<0.005) in 1000 in-house control chromosomes and databases- dbSNP151, 1000 Genomes, ExAC and gnomAD. Out of these, c.282G>T (p.Gln94His),

c.598A>G (p.Arg200Gly), c.635T>A (p.Ile212Lys) and c.1087T>C (p.Trp363Arg) give rise to non-synonymous changes in the CDC20B protein (Figure 6, Table 6) and c.990-5C>T and c.1460-4C>A lie in splice regions.

**Table 5: Novel/rare *CDC20B* variants identified by Sanger sequencing a cohort of 552 JME patients.**

| Genomic position (Grch38) | Region | cDNA position | Amino acid position | rs ID | MAF gnomAD v2 | MAF in-house controls |
|---|---|---|---|---|---|---|
| 55172851 | Intron | c.63+87G>A | _ | Novel | _ | _ |
| **55146701** | **Exon** | **c.282G>T** | **p.Gln94His** | **Novel** | **_** | **_** |
| 55143818 | Intron | c.356-175C>A | _ | rs372181403 | 0.00006 | 0.001 |
| 55143701 | Intron | c.356-58A>G | _ | Novel | _ | _ |
| 55140487 | Intron | c.487-80A>G | | Novel | _ | _ |
| 55133584 | Intron | c.581-56A>G | _ | Novel | _ | _ |
| 55133573 | Intron | c.581-45A>G | _ | Novel | _ | _ |
| **55133511** | **Exon** | **c.598A>G** | **p.Arg200Gly** | **rs1238048521** | **_** | **_** |
| **55133474** | **Exon** | **c.635T>A** | **p.Ile212Lys** | **rs550120464** | **0.0003** | **0.001** |
| 55133413 | Exon | c.696C>T | p.Tyr232Tyr | Novel | _ | _ |
| 55133290 | Intron | c.697+122G>A | _ | rs369502428 | 0.0002 | _ |
| 55125033 | Intron | c.990-5C>T | _ | rs769802207 | 0.00001 | _ |
| **55124931** | **Exon** | **c.1087T>C** | **p.Trp363Arg** | **rs759894199** | **0.00002** | **0.002** |
| 55124923 | Exon | c.1095G>A | p.Pro365Pro | rs375917242 | 0.00005 | _ |
| 55120578 | Intron | c.1216-28G>A | _ | rs1209715691 | 0.000008 | _ |
| 55120524 | Exon | c.1242T>C | p.Ser414Ser | rs532500094 | 0.0001 | 0.002 |
| 55119765 | Intron | c.1459+36T>C | _ | rs752577948 | 0.00006 | 0.001 |
| 55114322 | Intron | c.1460-4C>A | _ | rs1370902674 | _ | _ |
| 55114272 | Exon | c.1506C>A | p.Thr502Thr | rs773815311 | 0.000008 | _ |

| **55114253** | **Exon** | **c.1513G>A** | **p.Ala509Thr** | **rs200898156** | **0.0002** | **_** |
|---|---|---|---|---|---|---|
| 55113615 | 3'UTR | c.*603T>C | _ | Novel | _ | _ |
| 55113204 | 3'UTR | c.*1014A>G | _ | Novel | _ | _ |
| 55113080 | 3'UTR | c.*1138T>C | _ | Novel | _ | _ |
| 55113074 | 3'UTR | c.*1144G>A | _ | rs531849250 | _ | _ |
| 55112919 | Downstream | c.+52G>C | _ | Novel | _ | _ |

Of the total five non-synonymous variants identified in CDC20B, p.Ala509Thr and p.Trp363Arg, lie in the c-terminus of CDC20B, which harbours 7 WD40 repeat domains. Conservation analysis across species showed that the residues Trp363 and Ala509 are highly conserved across species, with Gln94, Arg200 and Ile212 also showing moderate conservation in mammals (Figure 8). The patients harbouring the CDC20B non-synonymous variants all manifest GTCS and myoclonic seizures, which are hallmarks of JME.

**Table 6: Novel/rare missense variants identified in *CDC20B***

| Genomic position (Grch38) | cDNA position | amino acid position | rs ID | MAF gnomAD 2.1 | SIFT | Polyphen-2 | CADD phred score | UMD predictor | Mutation Taster | PROVEAN 1.1 | fathmm-MKL coding |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 55146701 | c.282G>T | p.Gln94His | - | – | Tolerated (0.60) | Benign (0) | 0.005 | Polymorphism (32) | Neutral (1) | Neutral(-0.27) | Neutral (0.00) |
| 55133511 | c.598A>G | p.Arg200Gly | rs1238048521 | – | Damaging (0.05) | Benign (0.006) | 22.6 | Pathogenic (80) | Neutral (0.99) | Damaging (-2.59) | Neutral (0.13) |
| 55133474 | c.635T>A | p.Ile212Lys | rs550120464 | 0.0003 | Damaging (0.01) | Benign (0.091) | 11.03 | Polymorphism (36) | Neutral (1) | Neutral (-0.84) | Neutral (0.06) |
| 55124931 | c.1087T>C | p.Trp363Arg | rs759894199 | 0.00002 | Damaging (0.00) | Probably damaging (1) | 27.9 | Pathogenic (84) | Damaging (0.99) | Damaging (-13.65) | Damaging (0.97) |
| 55114253 | c.1525G>A | p.Ala509Thr | rs200898156 | 0.0002 | Damaging (0.001) | Probably damaging (1) | 31 | Probably pathogenic (66) | Damaging (0.93) | Damaging (-3.61) | Damaging (0.56) |

gnomAD 2.1 was used to get MAF; SIFT, Polyphen-2 v2.2.2, CADD v1.4, UMD-Predictor, MutationTaster 2, PROVEAN 1.1 ensembl 66 and fathmm-MKL were used to make predictions on the pathogenicity of the variants.

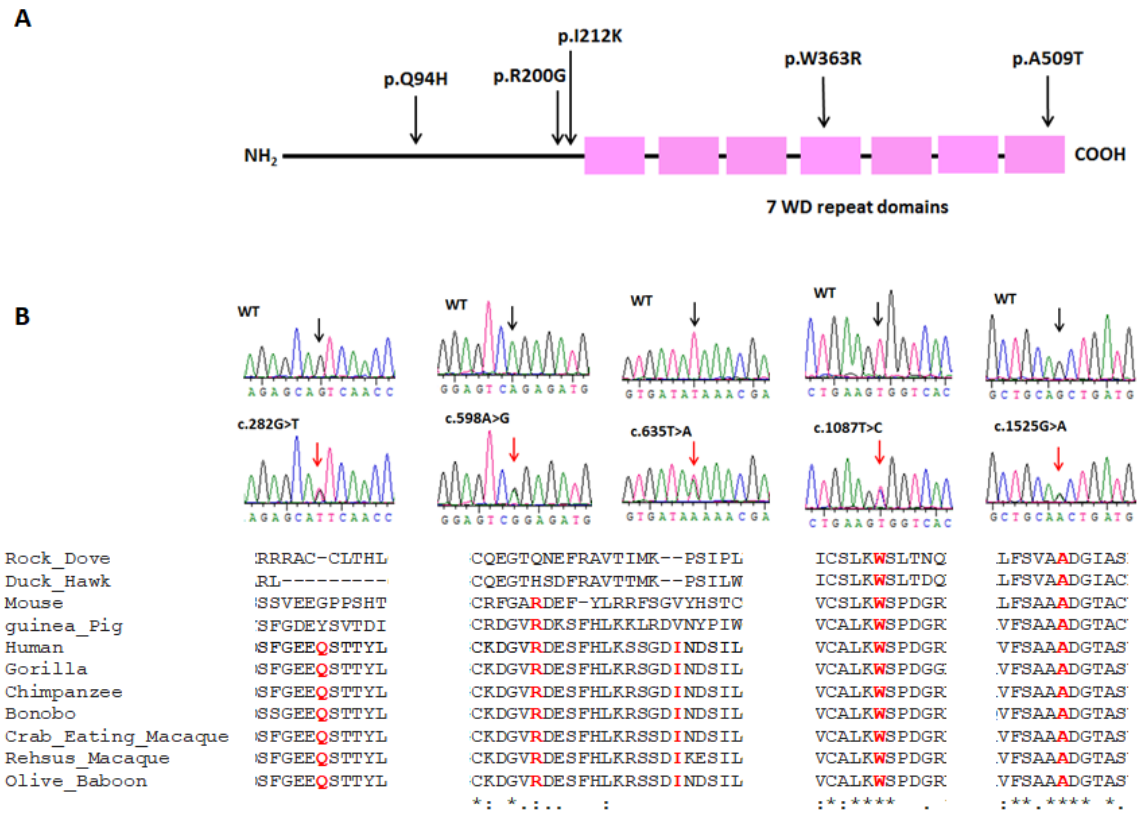



**Figure 8: Rare *CDC20B* variants in epilepsy patients: A.** CDC20B protein architecture is represented with the location of the five missense variants, p.Q94H, p.R200G, p.I212K, p.W363R and p.A509T. The carboxy-terminal of CDC20B harbours seven WD repeat domains, shown in pink boxes. Two of the patient variants, p.W363R and p.A509T, lie in WD repeat domains. **B.** Electropherograms showing DNA sequence of the respective wild-type (upper panel) and the variant (lower panel) from the genomic DNA. The presence of two peaks at one nucleotide position indicates the heterozygous variant for each variant. The substitution of the reference (black) to variant allele (red) is marked by an arrow. Alignment of CDC20B protein sequence from a few species by Clustal Omega, showing only the region near the location of mutated amino acid residue. The wild-type amino acid residue for each variant is highlighted in red.

### 2.2.8 *CDC20B* in Epi25

Epi25 is a global collaboration under which more than 9,000 epilepsy patients have been sequenced(https://epi25.broadinstitute.org/). In this study, ninety-two ultra-rare variants have been identified in *CDC20B* out of which 13 are potentially pathogenic. Seven of these variants give rise to missense changes in the protein sequence, two are loss of function variants and four lie in the splice regions (Appendix A-

2.4, A-2.5). *CDC20B* has been listed among the top 200 genes with burden of ultra-rare variants in the Epi25 collaborative, with the family variant p.Ala509Thr being identified in six epilepsy patients, three of whom have genetic generalized epilepsy (GGE). These observations provide further support for *CDC20B's* contribution to epilepsy.

### 2.2.9 CDC20B brain expression

We observed CDC20B expression in human and mouse brain samples. GTExRNAseq data and BioGPS microarray data both show that CDC20B is transcribed in the brain. Marathon-Ready full-length human brain cDNAs were used to amplify the CDC20B transcript. The amplified products were Sanger sequenced for confirmation (Figure 9-C) CDC20B expression analysis in human brain regions, cortex, hippocampus and cerebellum, and mouse whole brain was performed by Western analysis. We observed a band, immunoreactive to CDC20B antibody, at 54 kDa. Mouse testis sample, in which CDC20B is shown to be highly expressed by human protein atlas, was used as a positive control (Figure 9-D). CDC20B expression in mouse brain sections was checked by immunohistochemistry. CDC20B showed expression in mouse ependymal cells. We did not observe immunoreactivity of the CDC20B antibody anywhere else in the mouse brain (Figure 9-B).

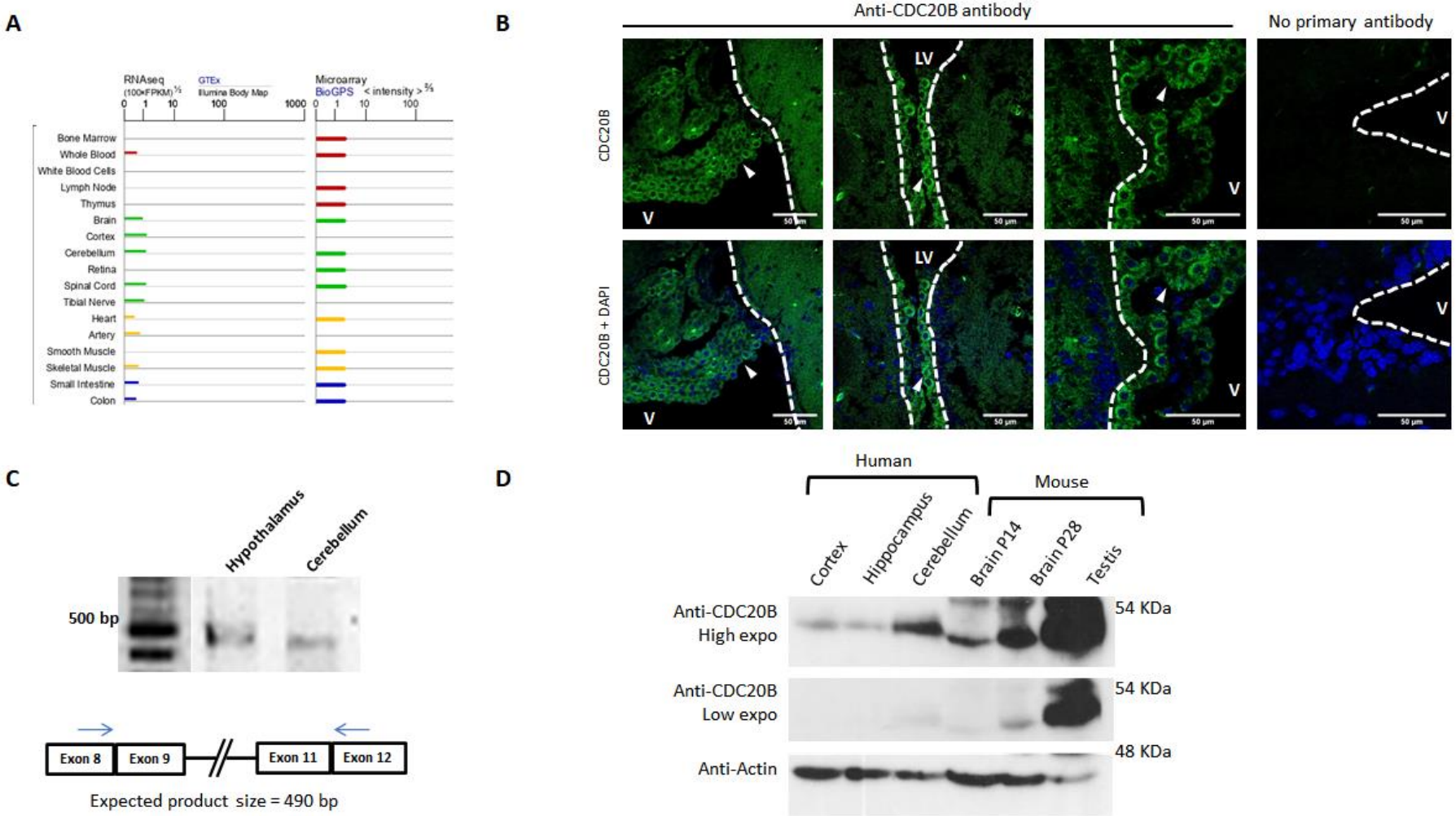

A
RNAseq
(100*FPKM) 1/2
GTEx
Illumina Body Map
Microarray
BioGPS
< intensity > 5/6
0 1 10 100 1000
0 1 10 100
Bone Marrow
Whole Blood
White Blood Cells
Lymph Node
Thymus
Brain
Cortex
Cerebellum
Retina
Spinal Cord
Tibial Nerve
Heart
Artery
Smooth Muscle
Skeletal Muscle
Small Intestine
Colon
B
Anti-CDC20B antibody
No primary antibody
CDC20B
CDC20B + DAPI
LV
V
50 µm
C
Hypothalamus
Cerebellum
500 bp
Exon 8
Exon 9
Exon 11
Exon 12
Expected product size = 490 bp
D
Human
Mouse
Cortex
Hippocampus
Cerebellum
Brain P14
Brain P28
Testis
Anti-CDC20B
High expo
54 KDa
Anti-CDC20B
Low expo
54 KDa
Anti-Actin
48 KDa

**Figure 9: CDC20B brain expression: A.** Expression of CDC20B in various human tissues shown by the RNA expression data summarized in GeneCards. GTExRNAseq data and BioGPS microarray data have been shown here. **B.** CDC20B brain expression in mouse brain sections checked by immunohistochemistry. CDC20B shows expression in mouse ependymal cells in the choroid plexus marked by arrowheads. L=Ventricle, LV=Lateral ventricle. **C.** CDC20B transcript was PCR amplified from human brain cDNAs. Forward primer binds in exon 8 and reverse primer binds in exon 12 of CDC20B. A band at 490 base pairs is expected. **D.** Western blot analysis of CDC20B protein in different human brain regions and mouse brain. Mouse testis has been used as a positive control as CDC20B is known to be highly expressed in this tissue. CDC20B antibody shows immunoreactivity at 54 KDa, corresponding to CDC20B predicted molecular weight.

### 2.2.10 CDC20B cellular localization

To get an insight into the role of CDC20B, its endogenous localization was checked across mammalian cell lines such as HeLa (epithelial cell type, derived from the human cervix), SH-SY5Y (epithelial cell type, derived from human bone marrow), U87-MG (epithelial cell type, derived from the human brain), HEK293 (epithelial cell type, derived from the human embryonic kidney) and NIH3T3 (fibroblast cell type, derived from mouse embryo), by immunocytochemistry using three different CDC20B antibodies. It was observed that CDC20B endogenous protein is present majorly in the cytoplasm with some expression in the nucleus as well, all through interphase. At different stages of mitosis, prophase, metaphase and anaphase, CDC20B is distributed uniformly across the cell body. During telophase and cytokinesis, CDC20B localizes at the midbody, an organelle responsible for the separation of the two daughter cells at the end of mitosis, marked by the midbody marker, gamma-tubulin (Figure 10).

### 2.2.11 CDC20B over-expression and cytokinesis defects

To study the effect of the patient-specific variants, the wild type and variant CDC20B protein was overexpressed in HeLa cells. In addition, the N-terminal and C-terminal truncated CDC20B protein was also overexpressed in HeLa cells. Western blot analysis of the wild type and variant CDC20B showed no significant difference in overall protein

expression. The N-terminal CDC20B protein, as opposed to the full-length protein, showed enhanced protein expression on Western analysis. Immunofluorescence analysis of the overexpressed CDC20B

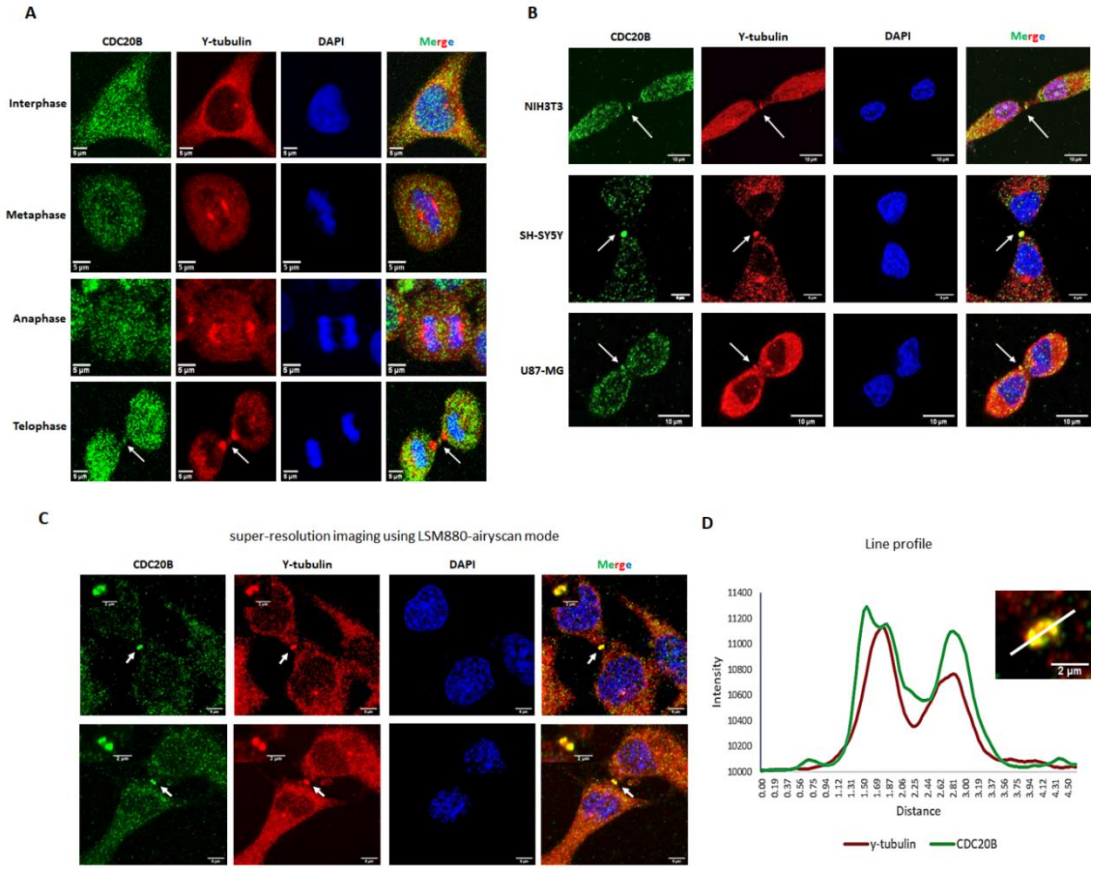



**Figure10: CDC20B cellular localization. A.** CDC20B localization in HeLa cells across different stages of the cell cycle. CDC20B endogenous protein is present majorly in the cytoplasm with some expression in the nucleus as well all through interphase. In telophase or during cytokinesis, CDC20B localizes to the midbody. **B.** CDC20B localization at the midbody in NIH3T3, SH-SY5Y and U87-MG cell lines. **C.** High/Super-resolution imaging of HeLa cells during cytokinesis showing CDC20B (green) localizing to the midbody along with gamma-tubulin (red). Arrows point to the midbody. Green: CDC20B, red: gamma-tubulin and blue: DAPI. **D.** Line profile made using Image J showing the presence of CDC20B and gamma-tubulin together at the midbody.

protein showed that it localises to the cytoplasm during interphase and to the midbody during cytokinesis both in the case of wildtype and mutant. Since the endogenous CDC20B protein localizes to the midbody during cytokinesis, we hypothesized that it may have a role in separating the two daughter cells at the end of mitosis. We decided to

overexpress the variant proteins in HeLa and check their effect on the progression of the cell cycle. It was observed that three of the variants, p.Arg200Gly, p.Trp363Arg and p.Ala509Thr, and the N-terminal CDC20B show more number of cells undergoing cytokinesis as opposed to the wildtype. This suggests that patient variant harbouring CDC20B is hampering the progression of cytokinesis (Figure 11). Also, the C-terminus, which harbours the seven WD40 domains, is essential for the CDC20B protein's function during cytokinesis.

### 2.2.12 CDC20B- interacting partners

Immuno-pulldown (IP) studies were conducted to obtain an insight into the molecular pathways CDC20B might be involved in. The colocalization of CDC20B with gamma-tubulin at the midbody lead us to check the interaction of the two proteins. FLAG-tagged CDC20B was overexpressed in HeLa cells and IP was done both with anti-FLAG antibody and anti-gamma tubulin antibody. 5% input, supernatant

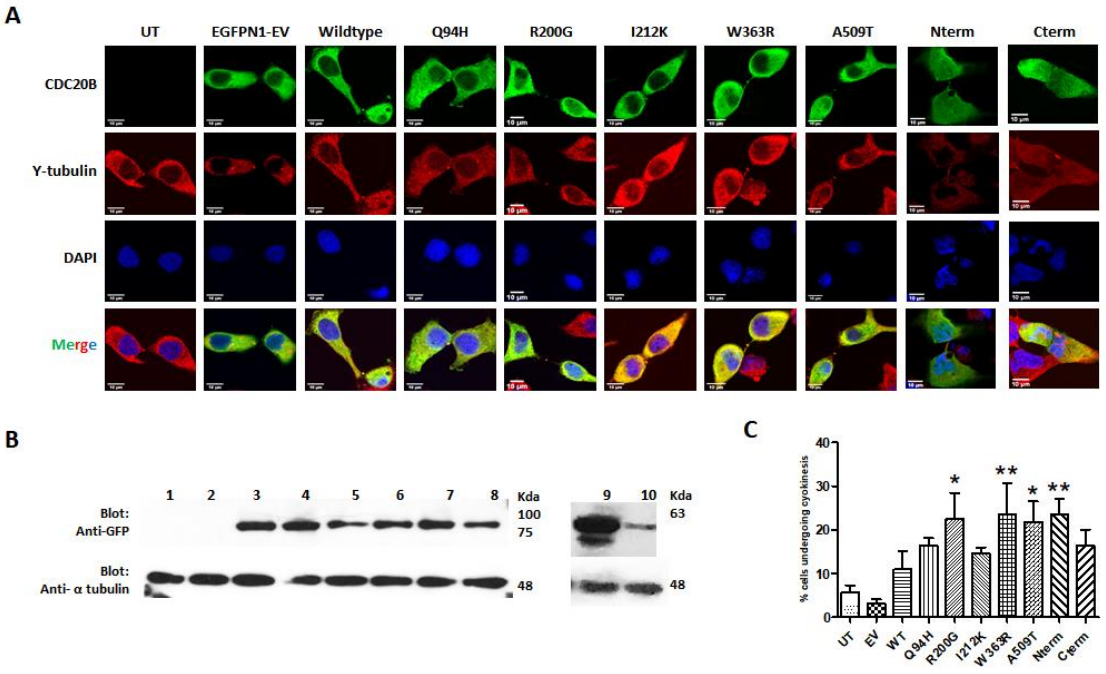



**Figure 11: CDC20B overexpression in HeLa cells.** GFP tagged CDC20B wildtype, mutants – Q94H, R200G, I212K, W363R and A509T, N-term and C-term truncations (green) were overexpressed in HeLa cells and co-stained with gamma-tubulin (red) **A.** Overexpressed CDC20B localizes to the cytoplasm and midbody.

**B.** Western blot analysis of over expressed CDC20B in HeLa cells. Lane1: un-transfected, lane2: EGFPN1-empty vector, lane3: EGFPN1-CDC20B-wildtype, lane4: EGFPN1-CDC20B-Q94H, lane5: EGFPN1-CDC20B-R200G, lane6: EGFPN1-CDC20B-I212K, lane7: EGFPN1-CDC20B-W363R, lane8: EGFPN1-CDC20B-A509T, lane9: EGFPN1-CDC20B-N-term (1st amino acid to 222nd amino acid of CDC20B), lane10: EGFPN1-CDC20B-C-term (223rd amino acid to 519th amino acid of CDC20B). The blot has been probed with an anti-GFP antibody to detect the GFP tagged CDC20B protein. Alpha-tubulin has been used as a loading control. **C.** The number of cells undergoing cytokinesis was quantified across three independent experiments with approximately 650 cells counted for each construct. Dunnett's multiple comparison test was used to calculate the P values. R200G and A509T have a P-value < 0.05 (*), W363R and N-term have a P-value <0.01 (**).

and the undenatured elute fraction were electrophoresed on an 8% SDS-PAGE gel. The blots were probed with anti-FLAG antibody to visualize CDC20B bands and anti-gamma tubulin antibody to visualize gamma-tubulin bands. We observe that CDC20B and gamma-tubulin are present in the eluted fraction, showing that they interact. FLAG-tagged CDC20B was immuno-precipitated using mouse IgG antibody. 5% input, supernatant, wash fractions, undenatured and beta-mercaptoethanol denatured elutes were electrophoresed on an 8% SDS-PAGE gel. Untransfected HeLa lysate was used to do an IP using anti-FLAG antibody. 5% input, supernatant, wash fractions, undenatured and beta-mercaptoethanol denatured elutes were electrophoresed on an 8% SDS-PAGE gel. In both control experiments, it was observed that there was no gamma-tubulin in the elutes (Figure 12).

In order to check which domain in CDC20B is responsible for its interaction with gamma-tubulin, GFP-tagged CDC20B - full length, N-terminal, C-terminal, amino acids 1-100, amino acids 101-395, amino acids 396-519, amino acids 101-519 and amino acids 1-395 were overexpressed in HeLa cells and IP was performed using anti-GFP antibody. Gamma tubulin was found in the elute fraction along with CDC20B, showing that they are present in the same complex of proteins (Figure 13). Each of the truncated CDC20B protein constructs that we expressed in HeLa cells showed interaction with gamma-tubulin. This shows that CDC20B interacts with gamma-tubulin in a

complex, where regions across CDC20B are contributing to the interaction.

Further, we wanted to check if CDC20B interacts with other proteins that localize to the midbody. The candidates we chose were, EFHC1, a known JME gene that also localizes to the midbody (Raju et al., 2017) and PLK1, a protein that recruits other proteins to the central spindle and midbody for cytokinesis (Jang et al., 2009; Tian et al., 2015). We find that CDC20B interacts with EFHC1 and PLK1 (Figure 13). We co-stained EFHC1 and PLK1 with CDC20B in HeLa cells. We find that CDC20B colocalizes with EFHC1 throughout the cytoplasm and at the midbody. PLK1 and CDC20B show colocalization at the midbody. CDC20B localizes to basal bodies in multicilliated cells, along with another protein centrin2 (Revinski et al., 2018). We checked for their interaction by co-IP and find that, under the conditions that we have used, CDC20B does not immunoprecipitate with centrin2 (Figure 14).

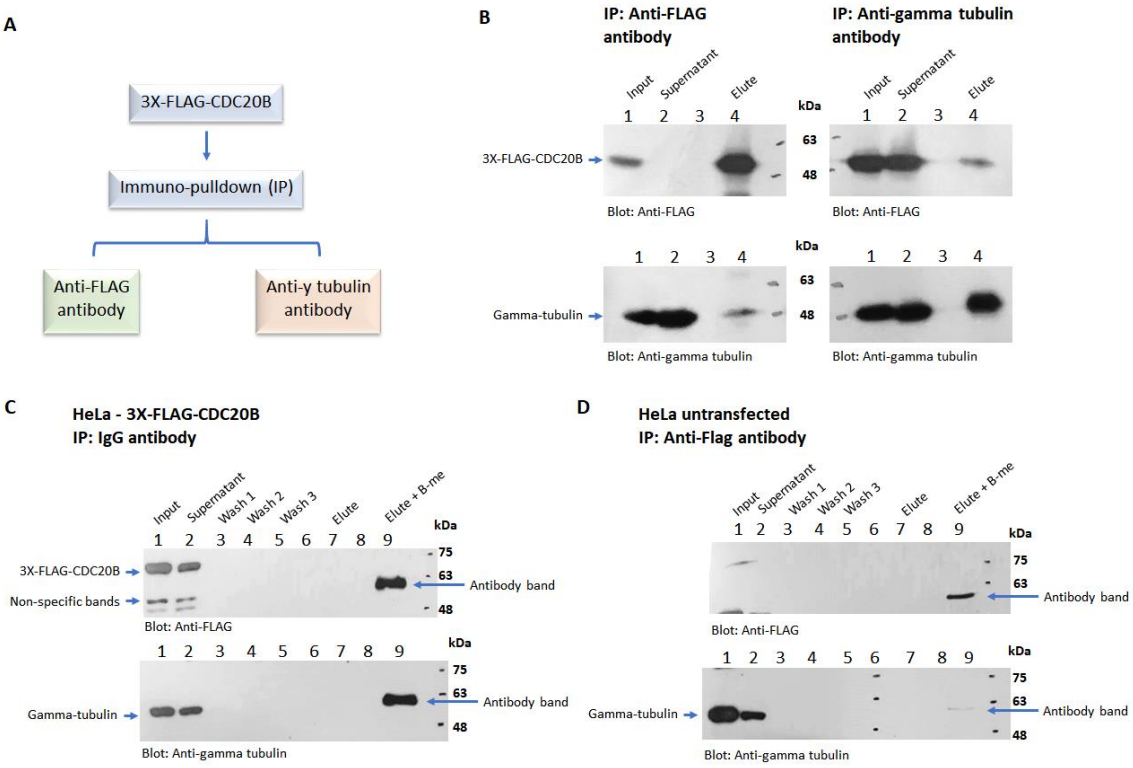



**Figure 12: CDC20B interacts with gamma-tubulin. A.** Strategy followed to check if CDC20B interacts with gamma-tubulin. FLAG-tagged CDC20B was overexpressed in HeLa cells and IP was done both with anti-FLAG antibody and

anti-gamma tubulin antibody. **B.** Western blot to analyze the IP. 5% input, supernatant and undenatured elute fraction was electrophoresed on an 8% SDS-PAGE gel. The blots were probed with anti-FLAG antibody to visualize CDC20B bands and anti-gamma tubulin antibody to visualize gamma-tubulin bands. We observe that CDC20B and gamma-tubulin are present in the eluted fraction, showing that they interact. **C.** FLAG-tagged CDC20B was immuno-precipitated using mouse IgG antibody. 5% input, supernatant, wash fractions, undenatured and beta-mercaptoethanol denatured elutes were electrophoresed on an 8% SDS-PAGE gel. **D.** Untransfected HeLa lysate was used to do an IP using anti-FLAG antibody. 5% input, supernatant, wash fractions, undenatured and beta-mercaptoethanol denatured elutes were electrophoresed on an 8% SDS-PAGE gel. In both control experiments, it was observed that there was no gamma-tubulin in the elutes.

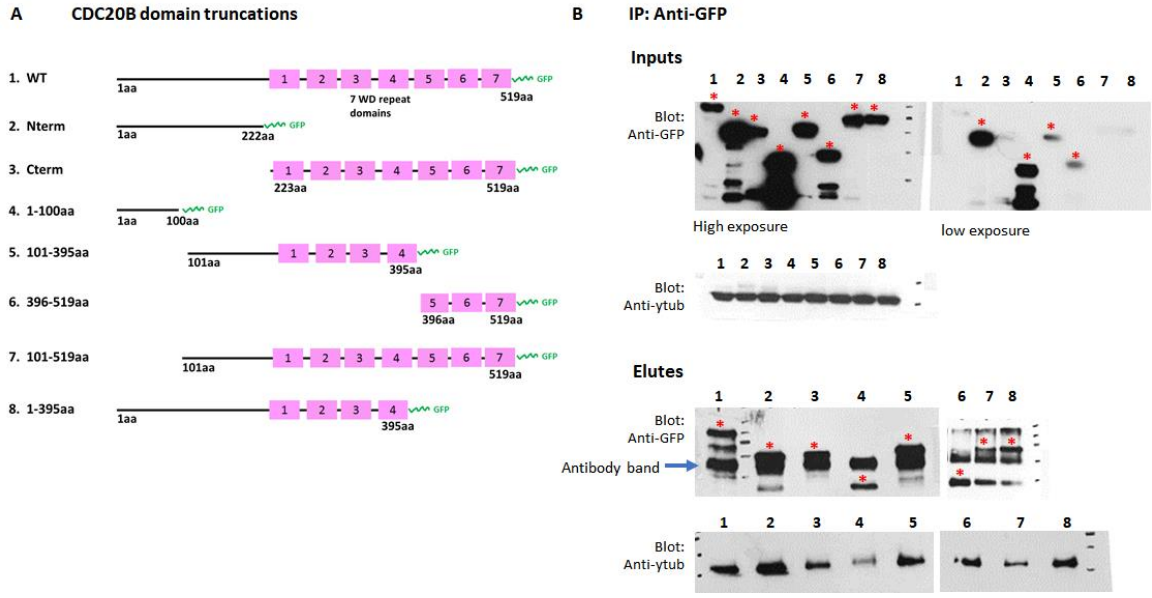



**Figure 13: Studying CDC20B domains responsible for its interaction with gamma-tubulin. A.** EGFPN1-CDC20B full length and truncated proteins used for the IP experiments. **B.** EGFPN1-CDC20B was immunoprecipitated using anti-GFP antibody. The inputs and elutes were probed with anti-GFP antibody to detect CDC20B and anti-gamma tubulin antibody. The elute fraction for the wildtype CDC20B as well as all the truncations showed the presence of gamma-tubulin. The blue arrow shows the antibody band in the elute fraction. Red stars mark the GFP-tagged CDC20B protein band in the elutes.

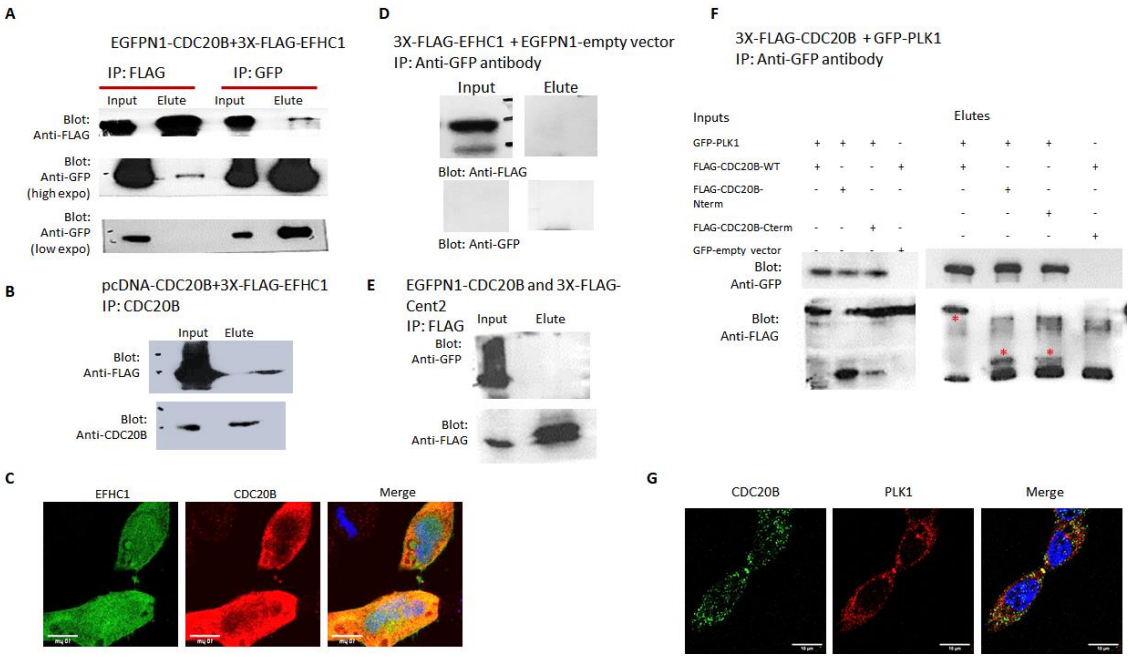



**Figure 14: CDC20B interacts with EFHC1 and PLK1. A.** GFP-tagged CDC20B and FLAG-tagged EFHC1 were co-expressed in HeLa cells. IP was performed with both anti-FLAG and anti-GFP antibodies. We find that in both Ips, CDC20B shows interaction with EFHC1. **B.** Untagged CDC20B was co-expressed with FLAG-tagged EFHC1 and IP was performed using anti-CDC20B monoclonal antibody (Sigma). A band corresponding to EFHC1 is found in the elute fraction. **C.** GFP-tagged EFHC1 and FLAG-tagged CDC20B and co-expressed in HeLa. Both these proteins co-localize in the cytoplasm and at the midbody during cytokinesis. **D.** FLAG-tagged EFHC1 and EGFPN1-empty vector were co-expressed in HeLa and IP was done using anti-FLAG antibody. On probing the elute with anti-GFP antibody, there was no band corresponding to EGFP-CDC20B. **E.** GFP-tagged CDC20B and FLAG-tagged centrin-2 were co-expressed and IP was performed using anti-FLAG antibody. We observe that CDC20B does not interact with centrin-2 under the conditions we have used. **F.** GFP-tagged PLK1 and FLAG-tagged full length, N-term and C-term CDC20B were co-expressed in HeLa cells. Anti-GFP antibody was used to perform IPs. CDC20B full-length and truncated proteins interact with PLK1. **E.** PLK1 and CDC20B colocalize at the midbody during cytokinesis.

### 2.2.13 CDC20B in-utero electroporation in mouse embryo

The patient variants, p.Trp363Arg and p.Ala509Thr were ectopically expressed in mouse embryos by performing in-utero electroporation at E14.5. Cells expressing the GFP tagged CDC20B protein were examined for their spatial localization in the section, numbers, and the length of the projections. GFP-empty vector showed high transfection

efficiency when compared to the wild type and mutant CDC20B-GFP. When we observed the spatial distribution of the transfected cells, we find that the empty vector shows distribution from the ventricular zone to the cortex. The CDC20B expressing cells are mostly concentrated in the ventricular and subventricular zones. The length of the projections of the neuronal cells was quantified using ImageJ. GFP-empty vector expressing cells show very long projections (average 480 μm). The projection lengths for CDC20B-GFP expressing cells were found to be smaller (average 190 μm). Both the mutants, p.Trp363Arg and p.Ala509Thr, showed projections significantly longer than the wildtype but shorter than the empty vector (Figure 14). Overexpression of CDC20B seems to negatively regulate the projection lengths and the variants do not show the same level of negative regulation. The change in the lengths of the neuronal progenitor projections can have an influence on neuronal migration, which in turn can be a cause for subtle brain developmental defects.

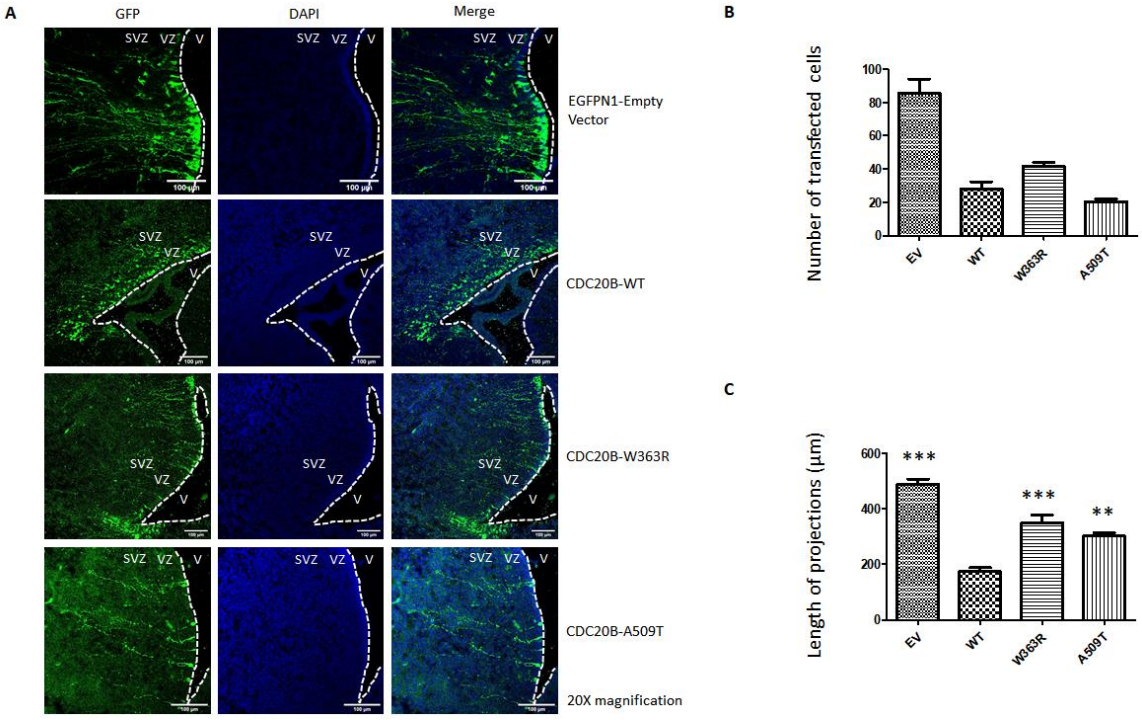



**Figure 15:** In-utero electroporation study of CDC20B wildtype and variant proteins. **A.** GFP-tagged empty vector and CDC20B have been imaged after staining with anti-GFP antibody. V: Ventricle, VZ: Ventricular zone, SVZ: Sub-ventricular zone. **B, C.** The number of transfected cells and their lengths have

been quantified. The experiment is done three times (N=3). The number of transfected cells counted per construct is 50. The number of cells used to measure the lengths of the projections per construct is 20. Dunnett's multiple comparison test was used to calculate the P values.

## 2.3 Discussion

Here, I have presented the results of a genetic analysis of a JME family, GLH35 from south India. Out of 13 members who participated in the study, four members had ABS, GTCS and MYO, two had GTCS and one, FS. Genome-wide linkage analysis of the family led to the identification of a locus at 5p15-q12. The highest lod score that we obtained was 1.7, which is due to the low penetrance value and phenocopies in the family. It has been suggested that the group of less severe epilepsies may be due to causal variants/genes of low penetrance effects (Myers et al., 2019).

The 64 MB, 5p15-q12 locus was analysed by performing whole-genome sequencing, which led to the identification of a rare missense variant, p.Ala509Thr, in *CDC20B*. On sequencing additional JME patients for *CDC20B*, 25 rare variants were identified, out of which four produced non-synonymous changes in the protein. Of the total five non-synonymous variants identified in CDC20B, two, p.Ala509Thr and p.Trp363Arg, lie in the c-terminus of CDC20B, which harbours 7 WD40 repeat domains. Conservation analysis across species showed that the residues Trp363 and Ala509 are highly conserved across species. The other three residues which lie in the N-terminus, Gln94, Arg200 and Ile212, showed moderate conservation in mammals. *CDC20B* has been listed among the top 200 genes with the burden of ultra-rare variants in the Epi25 collaborative (Epi25 Collaborative, 2019), with the family variant p.Ala509Thr identified in six epilepsy patients, three of whom had genetic generalized epilepsy (GGE), one each had developmental and epileptic encephalopathy (DEE) and non-acquired focal epilepsy (NAFE) with the type of epilepsy not given for the sixth patient. The

Epi25 study reports seven missense variants, two loss of function variants and four splice region variants in *CDC20B* (Appendix A2.4, A2.5). Eight of these variants lie in the WD-repeat domain region of CDC20B which is a highly conserved region of the protein. These data add evidence for *CDC20B*'s contribution as a potential epilepsy gene.

*CDC20B* encodes a 519 amino acid cell division cycle 20 homolog B protein. CDC20B protein harbours at its C-terminus, 7 WD40 repeat domains, which help in the formation of beta-propeller tertiary structures. These beta-propeller domains form protein-protein interaction scaffolds (Schapira et al., 2017). WD40 domain-containing proteins comprise the largest protein families (Schapira et al., 2017). These proteins are involved in a large variety of cellular processes many of which have been implicated in various disease pathogenesis (Xu and Min, 2011)

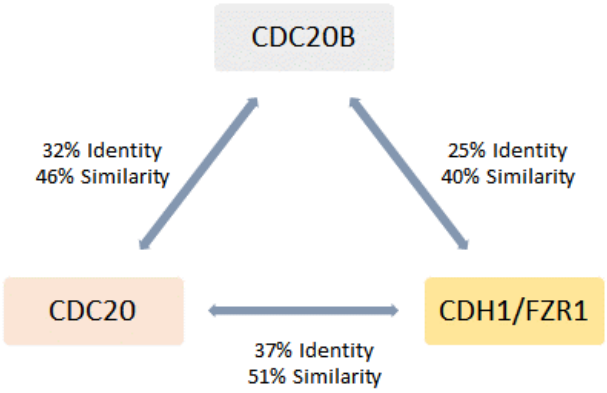



**Figure 16:** CDC20B and its paralogs CDC20 and CDH1/FZR1: CDC20B pairwise alignment was performed using Clustal Omega.

*CDC20B* has two paralogs, *CDC20* and *FZR1* (Figure 16), both of which bind to anaphase-promoting complex (APC) and help in the progression of the cell cycle during mitosis and G1 phase, respectively (Weinstein et al., 1994; Fang et al., 1998; Pfleger and Kirschner, 2000; Yamamuro et al., 2008; Sigl et al., 2009). Curiously, CDC20 protein is also reported to play a role in dendrite morphogenesis and regulation of pre-synaptic

differentiation (Kim et al., 2009; Yang et al., 2009b); *CDH1*/*FRZ1* regulates axonal growth and patterning in the mammalian brain and is essential for the survival of postmitotic neurons (Konishi et al., 2004; Almeida et al., 2005). Recent studies have also reported that *FZR1* is mutated in patients with neurodevelopmental disorders (Rodríguez et al., 2019). De-novo mutations in *FZR1* have been implicated in the causation of developmental and epileptic encephalopathies and Myoclonic Atonic Epilepsy. Drosophila fzr mutants have been reported to show abnormal neuronal patterning (Manivannan et al., 2021). These observations suggest that these WD40 domain-containing proteins have functions beyond cell cycle regulation. The molecular mechanisms by which CDC20 and FZR1 regulate these neurological functions are still not understood and need to be further deciphered.

CDC20B's cellular expression and localization experiments in various mammalian cell lines showed us that during interphase CDC20B is present in the cytoplasm. While CDC20B shows a diffused localization throughout the cell, from pro-metaphase to anaphase stages of mitosis, it exhibits co-localization with gamma-tubulin at the midbody/midbody ring during telophase and cytokinesis. When we overexpressed the patient-variant harbouring CDC20B protein in HeLa cells, we observed that the variants p.Arg200Gly, p.Trp363Arg and p.Ala509Thr and the truncated CDC20B protein lacking all the WD-40 domains showed a significantly greater number of cells undergoing cytokinesis as opposed to the wild type CDC20B. This increase in the number of cells undergoing cytokinesis suggests a delay in the time taken by the daughter cells to separate from each other in CDC20B variant expressing cells. Such defects in cytokinesis can have effects on the development of the brain as shown in the case, for example, of the CIT-K knockout mouse, a protein that localizes to the cleavage furrow/midbody and plays a role in cytokinesis (Madaule et al., 1998). CIT-K mutant mice display defective neurogenesis due to altered

cytokinesis and apoptosis. These mice have microcephaly, are ataxic and die due to severe epileptic seizures (Di Cunto et al., 2000).

The patient variants p.Trp363Arg and p.Ala509Thr were expressed in mouse embryos by performing in-utero electroporation at E14.5. Cells expressing GFP were quantified for their numbers, spatial localization in the section and the length of the projections. GFP-empty vector showed high transfection efficiency when compared to the wild type and mutant CDC20B-GFP. The length of the projections of the neuronal progenitors was also quantified using ImageJ. Both the mutants, p.Trp363Arg and p.Ala509Thr, showed cellular projections significantly longer than the wild-type controls. Overexpression of CDC20B could negatively regulate the projection lengths and the variants seem to partially abolish this regulation. The change in the neuronal progenitor projection lengths can have an effect on neuronal migration/connectivity, which in turn can be a cause for the development of seizure prone neuronal circuits (Qin et al., 2017).

In this study, we show that, in addition to gamma-tubulin (TUBG1), CDC20B interacts with EF-hand domain-containing protein 1 (EFHC1) and polo-like kinase (PLK1). *EFHC1*, another epilepsy-causing gene, localizes to the spindle poles and midbody (Suzuki et al., 2004; Bailey et al., 2017; Raju et al., 2017; Thounaojam et al., 2017). EFHC1 has been implicated in cell division, apoptosis, neuronal migration, regulation of ion channels, neurite architecture and neurotransmitter release (Suzuki et al., 2004; de Nijs et al., 2009, 2012; Rossetto et al., 2011; Nijs et al., 2013). EFHC1 null mice develop spontaneous myoclonus (Suzuki et al., 2009). PLK1, an essential mitotic kinase that regulates cytokinesis (Sunkel and Glover, 1988), has been shown to govern mitotic progression and spindle orientation in the developing mammalian cortex (Sakai et al., 2012). PLK1 regulates other cell cycle proteins such as KIF2A and ATXN-10, both of which have been shown to be important for brain development (Jang et al., 2009; Tian et al., 2015). Missense mutations in *KIF2A* cause microcephaly, early-onset

epilepsy and various malformations of cortical development (Poirier et al., 2013). A pentanucleotide repeat (ATTCT) in intron 9 of the ATXN10 gene causes spinocerebellar ataxia type 10 (SCA10), whose symptoms include cerebellar ataxia and epilepsy (Matsuura et al., 2000). Both KIF2A and ATXN10, are probable interactors of CDC20B according to a mass spectrometry study available at ProteomeXchange, ID:PXD010629 (http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD010629). It will be interesting to study CDC20B in light of the other proteins such as KIF2A, ATXN10, that it might associate with.

*CDC20B* is situated in a conserved locus that harbors several genes such as *MCIDAS*, *CCNO* and the microRNA cluster mir449, which are all known for their role in multi-ciliogenesis (Marcet et al., 2011; Lu et al., 2019; Terré et al., 2019). CDC20B expression is positively regulated by the multi-ciliogenesis proteins MCIDAS and GEMC1 (Terréet al., 2016; Lu et al., 2019). CDC20B and miR-449, a microRNA cluster present in the second intron of CDC20B, share a common promoter and are co-regulated during multi-ciliogenesis (Yang et al., 2009a). A recent report has shown that CDC20B is associated with deuterosomes during the phase of centriole amplification, and to basal bodies and cilia in fully differentiated mammalian MCCs. Mouse ependymal MCCs lacking CDC20B show no cilia formation, showing its importance in the formation of cilia in multi-ciliated cells (Revinski et al., 2018). Motile cilia of ependymal cells drive CSF flow and are required for migration of neuroblasts in the subventricular zones, and any abnormalities in these cells may cause abnormal neuroblast migration which can, in turn, lead to brain abnormalities (Sawamoto et al., 2006a). Overexpressing CDC20B patient-variants harbouring proteins in ependymal primary cultures and human airway epithelial cells can be employed as a tool to study the effect of these variants on cilia formation in MCCs.

EFHC1 localizes to the axoneme of primary cilia where it regulates neuronal activation and dopamine signalling (Loucks et al., 2019). The

CDC20B paralog CDC20 localizes to the basal body of primary cilia and regulates its length and disassembly (Wang et al., 2014b). There have been no reports yet looking at CDC20B's localization at the primary cilia, hence it will be interesting to study this aspect of CDC20B, as primary cilia are present in neurons and are centres of diverse cell signalling pathways, such as Wnt, Notch, planar cell polarity (PCP), Platelet-Derived Growth Factor Subunit A (Pdgfα), and the Hedgehog (Hh) (Goetz and Anderson, 2010).

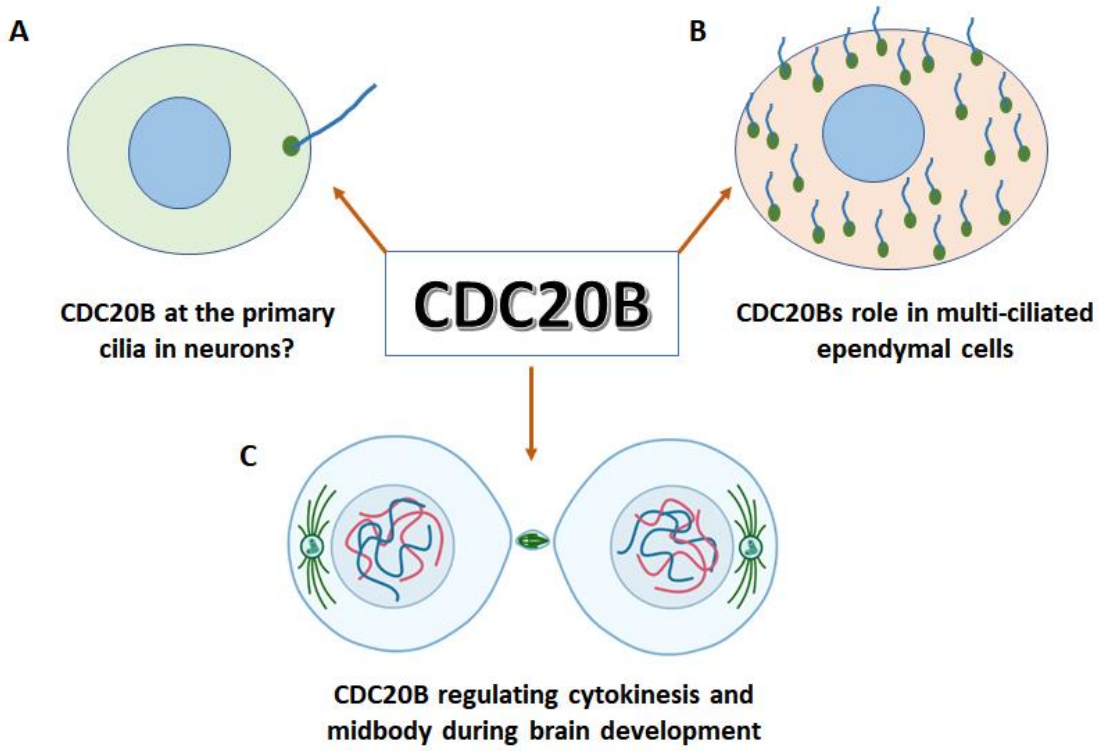



**Figure 17:** Hypothesis of CDC20B's functions based on its various known and predicted localizations in different types of cells. **A. Primary cilia in neurons:** Centres of diverse cell signalling pathways, such as Wnt, Notch, planar cell polarity (PCP), Platelet-Derived Growth Factor Subunit A (Pdgfα), and the Hedgehog (Hh) (Goetz and Anderson, 2010). **B. Multi-ciliated ependymal cells:** Motile cilia of ependymal cells drive CSF flow and are required for migration of neuroblasts in the subventricular zones (Sawamoto et al., 2006b). **C. Cytokinesis and midbody during brain development:** Cytokinesis and post-abscission midbody remnants are regulated during mammalian brain development (McNeely and Dwyer, 2020).

Based on our genetic findings discussed in this chapter, we propose *CDC20B* as a candidate for the cause of epilepsy in the family 35. We report five missense variants in *CDC20B* in this study: p.Ala509Thr in the family and p.Gln94His, p.Arg200Gly, p.Ile212Lys, p.Trp363Arg in additional unrelated JME patients. The interaction of CDC20B with gamma-tubulin, EFHC1 and PLK1 gives us a clue towards pathways that it might be regulating or be a part of. Further studies employing a CDC20B mouse knockout model may give us insights into its role in brain development and function, and its role in epilepsy pathophysiology.

# Chapter 3

## Genome sequencing reveals a mutation in *DES* at the EJM9, 2q33-q36 locus.

### Summary

A genetic locus for JME, EJM9 (2q33-q36) was identified in a three-generation south Indian family (SCT135) with several of its members affected with the disorder. I carried out whole-exome sequencing (WES) and whole-genome sequencing (WGS) to identify a potentially causative gene in the SCT135 family. Analysis of the gene variants in the 2q33-q36 region led to the identification of two novel variants, c.966A>T (p.Glu323Asp) in *DES* and c.2466+89G>C in *USP37,* co-segregating with JME. These were absent or rare (MAF>0.005) in databases such as gnomAD, dbSNP and our set of control individuals. Bioinformatic analysis suggested p.Glu323Asp in *DES* to be a damaging variant and c.2466+89G>C in *USP37* capable of disrupting normal splicing. A minigene-based splicing assay carried out for c.2466+89G>C in *USP37* in cultured HEK293T cells, showed the absence of any alternate spliced product, indicating that this intronic variant is likely to be a benign polymorphism. These results suggest that mutation in *DES* is likely to be causative of JME in the family. Further, to identify additional epilepsy-associated variants in *DES*, I examined the gene sequence in additional JME patients. I identified two new variants: p.Gly65Ser, and p.Ala337Thr among the JME patients examined. *DES* codes for a 470 amino acid, muscle-specific, type III intermediate filament. RT-PCR analysis showed the expression of desmin in different human brain regions. Desmin is known to be expressed in pericytes in the brain. Though the epilepsy-associated mutations in *DES* do not seem to affect its expression levels and filament formation, our observations, make a prima facie suggestion that they may have a damaging effect on its cellular function/s in the brain by affecting the physiology of the pericytes.

## 3.1 Materials and methods

### 3.1.1 Family ascertainment and linkage analysis

Family 135 (Figure 1) was ascertained through a 38-year proband, II:7, who developed MJ and GTCS at the age of about 20 years (Ratnapriya et al., 2010). Photoparoxysmal response was also observed for II:7. This patient's seizures were precipitated by sleep deprivation, alcohol consumption and flashing lights. He also had an antecedent history of febrile seizures. A detailed clinical history and neurological examination of all the family members were done by the study neurologist. All the affected members exhibited frequent myoclonic seizures occurring on awakening or within about two hours of waking up. The diagnosis of JME was based on the established diagnostic criteria, which were: (1) characteristic repetitive myoclonic jerks involving upper extremities, occurring after awakening, without loss of consciousness, (2) age at onset of myoclonic seizures between 8 and 25 years, (3) normal background activity and paroxysmal generalized spike and wave discharges in the EEG and (4) otherwise normal neurological status and intelligence. Proband's father (I:2) had MJ and GTCS. Proband's two sisters (II:2, II:4) and one brother (II:5) had MJ and GTCS starting at the ages between 15-17 years. II:2 and II:4, in addition, had histories of ABS. III:5 had MJ and GTCS with a history of febrile seizures. III:7 manifested MJ and GTCS from the age of 12 years. Proband's offspring, III:8 and III:9 did not have a history of seizures, but III:8 showed generalized activation of spike and wave discharges in the EEG during sleep. All family members provided written consent to participate in the study. Of the 17 members in the family, six had JME, with histories of febrile seizures in two of them, and one had generalized EEG abnormalities alone (Figure 1).

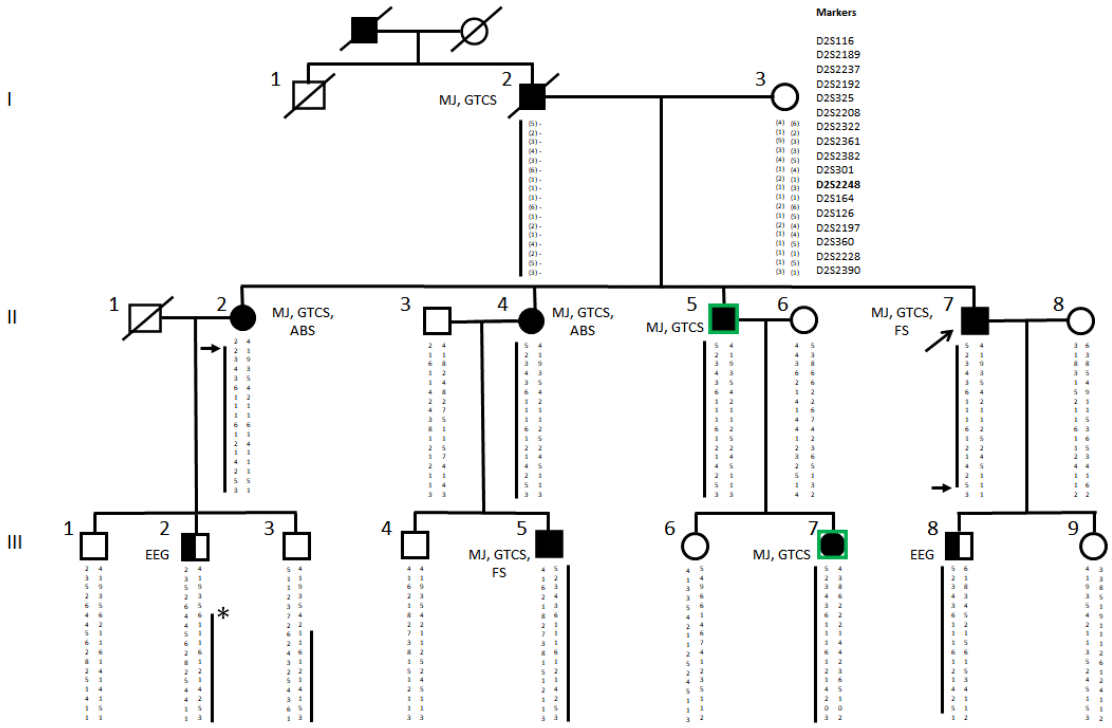



**Figure 1: SCT135 family:** Pedigree of SCT135 depicting the disease-linked haplotype (2q33-q36) shared among its eight-affected members (II:2, II:4, II:5, II:7, III:2, III:5, III:7, III:8) and one unaffected member (III:3). Males are denoted by squares and females are denoted by circles. Filled symbols represent affected individuals and empty, unaffected ones. Haplotypes are shown below the symbols. The inferred alleles are in parentheses, I:2, I:3. Arrows mark the recombinant events in individuals II:2 and II:7. The recombination event in the individual III:2 is indicated by an asterisk. Arrow marks the proband, II:7, and the green boxes indicate the individuals, II:5 and III:7, whole exome or genome sequenced. Clinical features of affected subjects are indicated along with the symbols (AB: Absences, MS: Myoclonic seizures, GTCS: Generalized tonic-clonic seizures, FS: Febrile seizures, EEG: Electroencephalography; the patient showed Generalized spike and wave discharges in their EEG recording but was clinically normal).

Ratnapriya and colleagues had performed genotyping and linkage analysis on this family. Seventeen individuals from the family, who participated in the study were genotyped (Figure 1) using 382 microsatellite markers from the ABI Prism Linkage Mapping Set $V_{2.5}$ and two- and multi-point lod score analysis was carried out.

### 3.1.2 Whole-exome sequencing experiment

#### 3.1.2.1 Library preparation

Five micrograms of genomic DNA (individuals II:5 and III:7) was fragmented (sonication at 55 pulses, 30s ON and 30s OFF) (Bioruptor-Diagenode, NJ, USA) and purified using Agencourt AMPure XP beads (Beckman Coulter, CA, USA). The target peak for base-pair size was 150-400 bp. The sheared DNA was analyzed for size distribution using Agilent DNA 1000 Bioanalyzer (Agilent Technologies, CA, USA). Successively, the sheared DNA fragments were used to construct DNA libraries using Agilent's SureSelectXT Target Enrichment System for illumina paired-end sequencing library. The constructed library was reformed by a series of steps using different enzymes to repair ends and make blunt-ended 5' phosphorylated fragments, add a single nucleotide A overhang and ligate 60 bp sequence adaptors to fragment ends. Each step was followed by a purification step using Agencourt AMPure XP beads. After ligation, the adapter-ligated fragments were enriched by PCR and concentrated using a vacuum concentrator (Eppendorf, Hamburg, Germany). The library was then hybridized to SureSelect™ biotinylated RNA baits at 65°C for 24-72 hours. Hybridized library fragments were isolated by magnetic capture using Dynal M-280 streptavidin-coated beads (Invitrogen, CA, USA) followed by purification of the capture library-bead solution using AMPure XP beads. PCR amplification was carried out to enrich the captured library and the amplified products were purified using AMPure XP beads. The amplified DNA was analyzed using the high sensitivity bioanalyzer chip (Agilent) which shows a peak in the size range of 300-400 nucleotides. The SureSelect human exome kit is designed to enrich a total of 51 Mb region in the genome. The sequencing was carried out for the captured libraries with illumina Genome Analyzer IIx platform (GAIIx) obtaining the 72 bp paired-end reads.

#### 3.1.2.2 Sequence alignment, variant calling and annotation

The raw sequence data (FASTQ) was processed through an in-house analysis pipeline. The quality check was done using SeqQC v2.1 (http:// genotypic.co.in/Products/7/Seq-QC.aspx) and the low-quality bases and 3' adapter sequences were removed. The whole-exome FASTAQ sequencing reads were aligned to human genome reference (hg19/GRch37) using BWA v-0.6.0 (Li and Durbin 2009). The reads showing at least 70% of bases with a minimum Phred score of 20, obtained by SeqQC v-2.0 (http:// genotypic.co.in/Products/7/Seq-QC.aspx), were used for alignment. Using SAMtools v-0.1.7a (Li et al 2009), duplicate reads arising possibly from PCR artefacts, were removed.

The variant calling was performed by SAMtools at a Phred-like SNP quality score of 20. The variants identified were annotated by SNPeff and filtered against the dbSNP131. Novel variants were further examined in updated databases such as dbSNP139 (http:// www.ncbi.nlm.nih.gov/SNP/), 1000 Genomes (http://browser.1000genomes.org/index.html), Ensemble (http://asia.ensembl.org/index.html) and EVS datasets (http:// evs.gs.washington.edu/EVS/). In order to obtain potential variants, which may have gotten missed at high coverage, variants up to 3x read depth were manually examined. Those transcript regions which remained uncovered by the whole-exome sequencing were manually identified and examined by Sanger sequencing.

### 3.1.3 Whole-genome sequencing (WGS)

#### 3.1.3.1 Library preparation and sequencing

The genomic DNA was checked for its integrity by electrophoresis on a 0.8% agarose-TAE gel (Figure 2). Three micrograms of genomic DNA was fragmented using Covaris to generate 300-400 base pair fragments. 100ng of fragmented DNA was used to generate a sequencing library

using NEBNext Ultra II DNA Library Prep Kit for illumina (New England Biolabs, USA, E7103). In brief, the fragmented DNA was subjected to end repair followed by A – tailing and adapter ligation. Ampure bead-based size selection was performed to obtain the library of the desired size. The size-selected DNA was enriched by PCR amplification using illumina index adapter primers (*P7 adapter 5'* – AGATCGGAAGAGCACACGTCTGA ACTCCAGTCA – 3', *P5 adapter 5'* – AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT – 3'). The amplified product was purified using Ampure beads to remove unused primers. The libraries were quantitated using Qubit DNA HS quantitation assay (Thermo Scientific) which specifically quantitates dsDNA. The library quality was checked using Agilent Bioanalyzer DNA 1000 kit (Figure 2). The QC passed libraries were sequenced using illumine HiseqX (illumina).

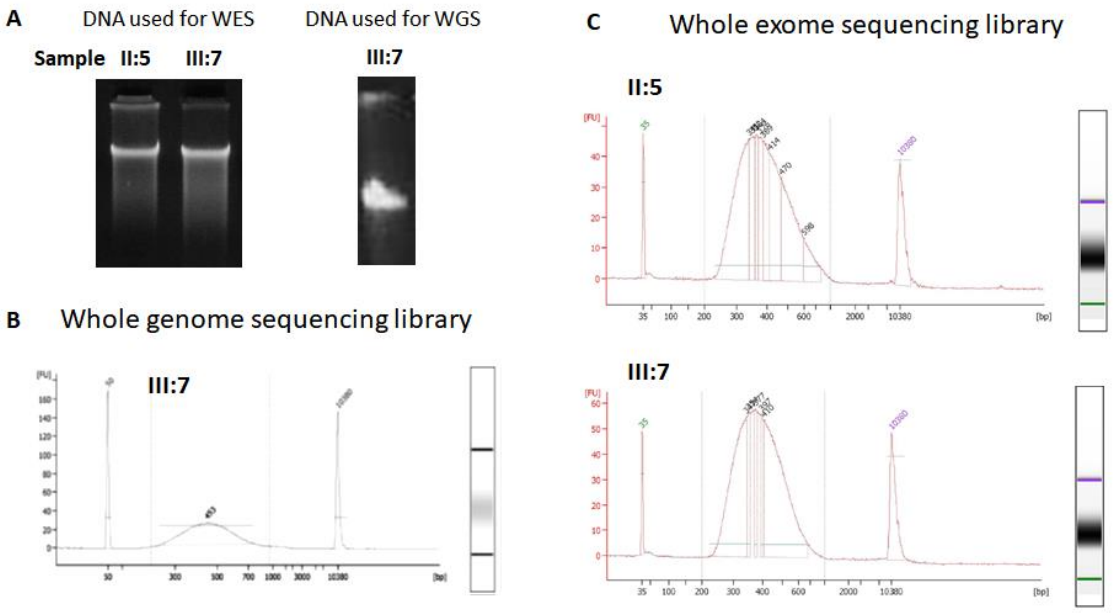



**Figure 2 – DNA and library quality check: A.** The genomic DNA (Sample II:5 and III:7) was checked for its integrity and quality by electrophoresis on 0.8% TAE-agarose gel. This genomic DNA was used for library preparation. The libraries were checked for their quality and size distribution using Agilent Bioanalyzer DNA 1000 Kit. **B.** Size distribution of the whole genome sequencing library made using the NEBNext Ultra II DNA Library Prep Kit for illumina. **C.** Size distribution of the whole-exome sequencing library made using Agilent's

SureSelectXT Target Enrichment System for illumina Paired-End Sequencing Library.

### 3.1.3.2 Sequence data quality check (QC), processing and alignment

The sequence data quality was checked using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and MultiQC (Ewels et al., 2016) software. The data was checked for base call quality distribution,% bases above Q20, Q30,%GC, and sequencing adapter contamination. The sequence data was processed using Trimgalore (https://github.com/FelixKrueger/TrimGalore) to remove adapter sequences and low-quality reads. QC passed reads were mapped to human reference genome build GRCh38 that was provided in GATK Resource Bundle using BWA mem (Li and Durbin, 2009b) algorithm with default parameters. The alignments were sorted, indexed and PCR duplicates were marked and removed using sambamba (Tarasov et al., 2015).

### 3.1.3.3 Variant calling, annotation and filtering

GATK (McKenna et al., 2010) best practice workflow for germline short variant discovery (SNPs + INDELs) using GATK v4 was followed for variant calling (https://software.broadinstitute.org/gatk/best-practices/). The process, in brief, base qualities in the bam file were recalibrated and GVCF was generated using HaplotypeCaller. GVCF was genotyped to generate vcf. SNPs and INDELs were recalibrated separately by providing variants from dbSNP, 1000 Genome phase1, hapmap, omni SNPs and Mills gold standard INDELs as known and training sets. The filter threshold was set to retain 99% of true variants, which provides the highest sensitivity while still being acceptably specific.

Functional annotation for the variants identified in the genome sequencing experiment was performed using Ensembl Variant Effect Predictor (VEP) build 94 (McLaren et al., 2016). The input was provided in the form of a tab-delimited vcf file (Variant Call Format). Transcript database from both GENCODE and NCBI was used. Allele frequencies were taken from dbSNP151, 1000 Genomes, Exome Aggregation Consortium (ExAC) and Genome Aggregation Database (gnomADv2, exomes and genomes; gnomADv3, genomes). dbSNFPv3.5 (Liu et al., 2011), a database developed for functional prediction and annotation of all potential non-synonymous single-nucleotide variants in the human genome, was also used for annotation. Samtools (Li et al., 2009) was used to view the whole genome sequence data aligned with the human reference sequence (Grch38) and each exon in the region of interest on chromosome 5 was manually checked for its coverage and read depth.

The identified variants were filtered for minor allele frequency in the databases, dbSNP151, 1000 Genomes project (Auton et al., 2015), Exome Aggregation Consortium (ExAC) (Lek et al., 2016) and the Genome Aggregation Database (gnomAD) (Karczewski et al., 2020) for their minor allele frequency (MAF) to be less than 0.005.

### 3.1.4 Sanger-based sequence validation of WGS rare variants

All novel/rare variants identified in the WGS dataset were validated by Sanger sequencing. Primers were designed spanning the variant-carrying exons/regions. The Sanger confirmed variants were further analysed in the family for their segregation with the disease phenotype. The segregating variants were examined in 488 ethnically matched control set of normal individuals for their MAF.

### 3.1.5 *DES* mutation analysis

The gene *DES*, which carried the disease co-segregating rare variant, was further analysed in a set of 192 JME affected individuals and ethnically matched control individuals. The entire gene sequence

comprising all exons, flanking intronic boundaries, and the 5'- and 3'-untranslated regions were Sanger sequenced to identify additional rare variants.

### 3.1.6 Bioinformatics analysis

The complete gene sequence of *DES* was downloaded from NCBI (NC_000002.12:219418377-219426734, Homo sapiens chromosome 2, GRCh38.p13 Primary Assembly). Primers were designed using the online tools Primer3 (Untergasser et al., 2012) and OligoCalc (Kibbe, 2007). Protein sequences for DES across different species were downloaded from NCBI to perform multiple sequence alignments using the tool Clustal Omega (Madeira et al., 2019).

### 3.1.7 Plasmids and antibodies

*DES* cDNA was amplified from human brain cDNAs (Clontech) and cloned into pcDNA3.1(+) plasmid. Site-directed mutagenesis was performed using QuikChange II XL Mutagenesis reagents (Agilent Technologies) to generate patient-specific variants p.Gly65Ser, p.Glu323Asp and p.Ala337Thr.

The antibodies used in this study were: anti-desmin (polyclonal, rabbit raised; Abcam), anti-vimentin (monoclonal, mouse raised; Sigma), anti-vinculin (Polyclonal, rabbit raised; Abcam), anti-alpha tubulin (monoclonal, mouse raised; Sigma), goat anti-rabbit Alexa Fluor 488 (Thermo Fisher Scientific), goat anti-mouse Alexa Fluor 568 (Thermo Fisher Scientific), goat anti-rabbit IgG – HRP (Genei), goat anti-mouse IgG – HRP (Genei).

### 3.1.8 Cell culture and transfection

HEK-293 cells and SH-SY5Y cells were maintained in DMEM (high glucose, Sigma) supplemented with 10% heat-inactivated fetal bovine serum (MP Biomedicals), 2 mM L-glutamine (Sigma) and antibiotics (100 U/ml Penicillin and 10 mg/ml streptomycin; Sigma) 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) coated 18 mm glass coverslips and transiently transfected with 1 µg plasmid per well using Lipofectamine 2000 transfection reagent (Thermo Fischer Scientific). Protein expression was checked 24 to 48 hours post-transfection.

### 3.1.9 Minigene assay

To study the effect of the intronic variant, c.2466+89G>C in USP37, on splicing, the exon close to the intronic variant along with 500 base pairs upstream and downstream of the exon-intron boundary was cloned in the exon trapping vector pSPL3 (Figure 4) (Tompson and Young, 2017) using the primer pairs 5' – GTGTGAATTCACACTATGACCTATGGCTG – 3' and 5' – TCAGTCAGGGATCCGAAACACTACCCACGTACC– 3'. PCR was done in a 50 µl reaction volume containing 200 µM dNTP, 0.4 µM of each primer, 100 ng of genomic DNA and 1 unit Phusion high-fidelity DNA polymerase (New England Biolabs, M0530S). Amplification was performed using a GeneAmp 9700 at following conditions: initial denaturation at 94°C (5 minutes), followed by 25 cycles of denaturation at 94°C (30 seconds), annealing at 58°C (30 seconds) and extension at 72°C (1 minute) followed by a final extension at 72°C (10 minutes). The amplified products were electrophoresed and visualized on a 0.8% agarose/TAE/EtBr gel (Figure 4). The amplified minigene-insert was purified using Multiscreen 96 well PCR purification plates (Millipore). The 1072 bp minigene-insert was cloned into a pSPL3 mammalian expression vector, using EcoRI and BamHI restriction enzyme sites. Site-directed mutagenesis was performed using QuikChange II XL site-directed mutagenesis reagents (Agilent Technologies) to generate the patient-specific variants. These wild-type and mutant constructs were used for transfection in HEK293 cells. Total RNA was isolated from the cells using the TRIzol method (Thermo Fisher Scientific). cDNA synthesis was performed using the superscript III first-strand synthesis

system (Thermo Fisher Scientific). Primers V1 (V1-F: 5'-TCTGAGTCACCTGGACAACC-3') and V2 (V2-R: 5'-ATCTCAGTGGTATTTGTGAGC-3') were used to amplify the minigene. The amplified products were electrophoresed and visualized on a 0.8% agarose/TAE/EtBr gel (Figure 4). The products were Sanger-sequenced to confirm any splicing errors.

### 3.1.10 RNA isolation

Glassware to be used during RNA isolation and Milli-Q water to be used for making solutions/reagents were Diethyl Pyrocarbonate (DEPC) treated and double autoclaved. TRIzol reagent was added to the cells and incubated 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 12,000 g at 4°C. The mixture separated into a lower red phenol-chloroform, interphase, and a colourless 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 for 10 minutes at 12,000 g at 4°C. Total RNA precipitate forms 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 7500 g 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 water bath set at 55–60°C for 10–15 minutes for dissolution.

### 3.1.11 cDNA synthesis

The SuperScript III First-Strand Synthesis System for RT-PCR was used to synthesize first-strand cDNA from total RNA. 1 µg total RNA, 5 µM oligo$(dT)_{20}$, 1 mM dNTP mix and DEPC treated water to makeup volume to 10 µl were combined in a 0.2 ml PCR tube and incubated at 65$^{o}$C for 5 minutes followed by transferring to ice. The cDNA synthesis mix comprising, 2 X RT buffer, 10 mM $MgCl_2$, 0.02 M DTT, 40 units of RNaseOUT enzyme and 200 units of SuperScript III RT enzyme, was prepared and 10 µl 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, followed by chilling on ice. 1 µl RNase-H was added to each tube and the reaction was incubated at 37$^{o}$C for 20 minutes. The synthesized cDNA was stored at -20$^{o}$C.

### 3.1.12 Western analysis

For isolating whole-cell protein, the cells/tissues were lysed using lysis buffer (150 mM NaCl, 10 mM Tris-pH 7.5, 0.1% SDS, 1% Triton X-100, 1% deoxycholate and 5 mM EDTA). The total protein content of the cells was collected by centrifuging the lysed cells at 13,000 rpm for 20 minutes at 4$^{o}$C. Total protein content was quantified using the Pierce BCA protein assay kit (Thermo Fischer Scientific). 50 µg of total protein was separated on a 10% SDS-polyacrylamide gel (Laemmli, 1970) followed by transfer to nitrocellulose membrane (Pall) at 20 volts for 1 hour. The membrane was blocked with 5% skimmed milk for 4 hours at 4$^{o}$C followed by incubation with primary antibodies, anti-desmin (1:500) or anti-alpha tubulin (1:5000) or anti-vinculin (1:5000) at 4$^{o}$C for 16 hours. HRP-conjugated secondary antibodies (1:5000) were used to probe the membrane at 4$^{o}$C for 4 hours. The membrane was washed between each incubation step with 0.05% PBST (0.05% tween-20 in 1X PBS). Clarity western ECL chemiluminescence substrate (Biorad) was used to detect the protein bands on photographic films.

### 3.1.13 Immunocytochemistry

To perform immunocytochemistry analysis, cells were grown on 18 mm poly-L-lysine coated glass coverslips. The cells were fixed with 2% paraformaldehyde (PFA) for 15 minutes at room temperature (RT) or with ice-cold methanol (for gamma-tubulin staining) at -20$^{o}$C for 5 minutes. The fixed cells were permeabilized using 0.1% Triton X-100 for 10 minutes at RT and then blocked with 3% BSA (bovine serum albumin) for 1 hour at RT. Incubations with primary antibodies, anti-desmin (1:500) and anti-vimentin (1:5000), and fluorescently labelled secondary antibodies (1:500), diluted in 1% BSA, were done for 1 hour each at room temperature. Nuclear staining was performed using 1 µg/ml 4, 6-diamidino-2-phenylindole (DAPI) for 15 minutes at RT. The coverslips were mounted on glass slides using Polyvinyl alcohol mounting medium (Sigma) and sealed with transparent nail polish. Imaging was done using Zeiss LSM 510 meta/880 confocal microscope.

## 3.2 Results

### 3.2.1 Whole-exome sequencing

To examine the 2q33-q36 region, whole-exome-based sequencing was undertaken for individuals II:5 and III:7 of the family SCT135. This region spans 24 Mb of sequence length and harbours 158 well-annotated protein-coding genes. In the whole-exome sequencing experiment, a total of 9.59 GB sequence per sample, as 72 base pair long paired-end reads, was generated from sequencing on the GAIIx where more than 98% of bases were of high quality and total coverage of 98% was obtained (Table 1). The region of our interest has 2061 coding exons and 563 non-coding exons out of which 120 coding exons and 380 non-coding exons have not been covered by the whole-exome sequencing. The 120 coding exons, 5' UTRs and the first 500 base pairs of 3' UTR were targeted manually by Sanger sequencing.

**Table1: Whole-exome sequencing statistics**

| Sample | II:5 | | III:7 | |
|---|---|---|---|---|
| **Target** | **Whole Exome Probes Targeted** | **chr2 – Whole Exome Probes Targeted** | **Whole Exome Probes Targeted** | **chr2 – Whole Exome Probes Targeted** |
| Total number of reads | 87.48 Million | | 76.00 Million | |
| Reads aligned to complete genome | 87.34 Million | | 75.87 Million | |
| %Reads aligned to complete genome | 99.84 | | 99.82 | |
| Reads aligned to the target region | 70.84 Mb | 1012 Kb | 60.80 Mb | 811 Kb |
| %Reads aligned to target region | 80.97 | 1.16 | 79.99 | 1.07 |
| Target Sequence Length | 51.54 Mb | 645 Kb | 51.54 Mb | 645 Kb |
| Target covered | 49.71 Mb | 630 Kb | 50.00 Mb | 632 Kb |
| % Target covered | **96.45** | 97.63 | 97.01 | 98.03 |
| % Target covered with at least 5X Read Depth | 89.70 | 93.81 | 91.37 | 94.80 |
| % Target covered with at least 10X Read Depth | 83.66 | 89.83 | 86.18 | 91.52 |
| % Target covered with at least 15X Read Depth | 78.33 | 85.88 | 81.33 | 87.96 |
| % Target covered with at least 20X Read Depth | 73.51 | 82.27 | 76.50 | 84.06 |
| Average Read Depth | **69.33** | 76.74 | 59.61 | 61.63 |

Percentage of genome/ region coverage calculated after sequences were aligned to the human genome reference sequence (GRCh37).

### 3.2.2 WES rare variant analysis and prioritization

Upon variant analysis 969 variants were identified out of which 940 variants were reported in dbSNP135 or 1000 Genome project datasets with MAF>0.005 and hence, were not carried forward for our study. The 29 novel variants identified were Sanger validated and then checked for their segregation in the family. Three variants, namely, c.966A>T in *DES,* c.401G>C in *USP37* and c.2908C>T in *TNS1*, co-segregated with the JME phenotype in the family. Two of the co-segregating variants, c.401G>C in *USP37* and c.2908C>T in *TNS1*, were present in normal control individuals and were therefore, not taken forward for analysis (Table 3).

### 3.2.3 Whole genome sequencing

To cover all the UTR regions which were not covered by the whole-exome sequencing experiment, the linked region 2q33-q36 was examined by whole-genome based sequencing in the member III:7. In the whole-genome sequencing experiment, a total of 90 GB sequence data, as 583 million paired end reads of 150 bp, was generated from sequencing on the illuminaHiseq X sequencer (illumina). More than 97% of bases were of high quality (Phred score > 20) and an average read depth of 26 X was obtained (Table 2). A total of 54,030 variants were identified out of which 875 variants were either absent or present at a MAF<0.005 across the databases dbSNP151, 1000 Genomes, ExAC and gnomAD (Figure 3).

**Table 2: Whole-genome sequencing statistics**

| Sample | III:7 | |
|---|---|---|
| **Fields** | **Genome** | **2q33-q36 Region** |
| Total number of reads | 583.6 Million | |
| Reads aligned to complete genome | 583.2 Million | |
| **%Reads aligned to complete genome** | 99.93 | |
| Target Sequence Length | 3.2 Gb | 24 Mb |
| Total target covered | 3.0 Gb | 22 Mb |
| **% Target Covered** | 94.93 | 99.91 |
| % Target Covered with at least 5X read depth | 93.31 | 93.91 |
| % Target Covered with at least 10X read depth | 92.09 | 93.09 |
| % Target Covered with at least 15X read depth | 90.48 | 90.9 |
| % Target Covered with at least 20X read depth | 79.1 | 80.1 |
| % Target Covered with at least 30X read depth | 39.74 | 41.29 |
| **Average read depth** | 26.97 | 27.01 |

Percentage of genome/ region coverage calculated after sequences were aligned to the human genome reference sequence (GRCh38).

### 3.2.4 WGS rare variant analysis and prioritization

Of the 875 rare variants identified at the 2q33-q36 locus, 440 lie in protein-coding genes. The variants in protein-coding transcripts (440) were further classified based on their consequences. There were 387 variants in intronic regions, 17 in UTR regions, 101 in regions upstream or downstream genes, one variant gave rise to a missense change and two variants to a synonymous change in the protein sequence. Variants in low complexity regions, which are rich in polymorphisms, regions beyond 100 base pairs from the exon-intron boundary and variants in regions upstream or downstream of genes were excluded from further analysis (Figure 3).

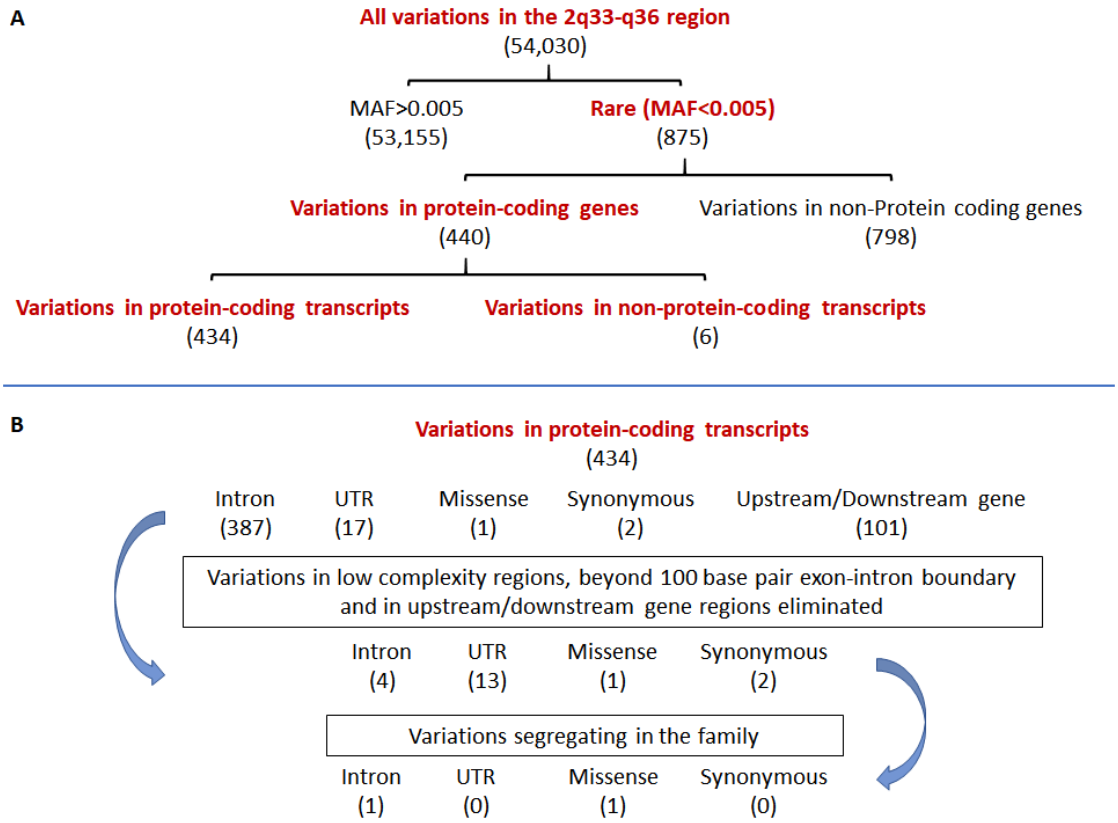



**Figure 3: Whole-exome/genome sequencing variants filtration and prioritization flow chart.** The variants identified in the 2q33-q36 locus are first filtered based on their minor allele frequency (MAF) across databases. Variants with frequency less than 0.005 are further filtered based on their annotation and their segregation with the disease phenotype in the family.

Segregation status was checked for the rare variants identified in the protein-coding genes in the WGS analysis. c.2466+89G>C in *USP37* was the additional rare segregating variant identified in the family. This variant was missed in our WES experiment (Table 3). In-silico predictions showed that the variant c.2466+89G>C in *USP37* could potentially alter splice sites and was hence taken forward for minigene assay.

**Table 3: Rare variants identified by whole-exome/genome sequencing in the 2q33-q36 locus in the SCT135 family**

| Grch38 | Gene | Sequence variant | Effect on protein | Family segregation | MAF in-house controls |
|---|---|---|---|---|---|
| 195756076 | *DNAH7* | c.11586+57G>A | _ | Non-segregating | _ |
| 195857660 | *DNAH7* | c.8131A>G | p.Ile2711Val | Non-segregating | _ |
| 196307273 | *HECW2* | c.2587-40T>C | _ | Non-segregating | _ |
| 196834398 | *PGAP1* | c.*6836C>G | _ | Non-segregating | _ |
| 197552916 | *MOB4* | c.*2271del | _ | Non-segregating | _ |
| 200823713 | *BZW1* | c.*1538dup | _ | Non-segregating | _ |
| 201389524 | *TRAK2* | c.1194-21G>A | _ | Non-segregating | _ |
| 201725395 | *ALS2* | c.3308A>G | p.His1103Arg | Non-segregating | _ |
| 201757845 | *ALS2* | c.1114-86G>A | _ | Non-segregating | _ |
| 201835573 | *CDK15* | c.731-70C>T | _ | Non-segregating | _ |
| 202561287 | *BMPR2* | c.*1346del | _ | Non-segregating | _ |
| 202788864 | *ICA1L* | c.1209T>A | p.Phe403Leu | Non-segregating | _ |
| 203297084 | *CYP20A1* | c.*176T>C | _ | Non-segregating | _ |
| 203430502 | *ABI2* | c.*3150G>A | _ | Non-segregating | _ |
| 205121812 | *PARD3B* | c.1028C>T | p.Ala343Val | Non-segregating | _ |
| 206175509 | *GPR1* | c.*671T>G | _ | Non-segregating | _ |
| 207078821 | *KLF7* | c.*2392G>T | _ | Non-segregating | _ |
| 207604477 | *CREB1* | c.*7419T>C | _ | Non-segregating | _ |
| 208240027 | *IDH1* | c.851-24C>T | _ | Non-segregating | _ |
| 209943322 | *UNC80* | c.6718-58G>A | _ | Non-segregating | _ |
| 210192812 | *ACADL* | c.1191A>G | p.Pro397Pro | Non-segregating | _ |
| 216126043 | *XRCC5* | c.798+12G>A | _ | Non-segregating | _ |
| 217831545 | *TNS1* | c.2908C>T | p.Arg970Trp | Segregating | 3/96 |
| 217835109 | *TNS1* | c.2887A>G | p.Ser963Gly | Non-segregating | _ |
| 217848829 | *TNS1* | c.1313G>A | p.Arg438Gln | Non-segregating | _ |
| 218134891 | *CXCR2* | c.90C>T | p.Pro30Pro | Non-segregating | _ |
| 218164579 | *CXCR1* | c.633C>A | p.Phe211Leu | Non-segregating | _ |
| 218357698 | *C2ORF62* | c.283T>A | p.Phe95Ile | Non-segregating | _ |
| 218465921 | *USP37* | c.2466+89G>C | _ | Segregating | 1/480 |
| 218549837 | *USP37* | c.401G>C | p.Ser134Thr | Segregating | 5/384 |
| 218699187 | *STK36* | c.3643G>A | p.Ala1215Thr | Non-segregating | _ |
| 218738556 | *TTLL4* | c.880G>A | p.Asp294Asn | Non-segregating | _ |

| | | | | | |
|---|---|---|---|---|---|
| 218823691 | *PRKAG3* | c.*71C>T | _ | Non-segregating | _ |
| 219248496 | *STK16* | c.855C>T | p.Leu330Leu | Non-segregating | _ |
| 219420899 | *DES* | c.966A>T | p.Glu323Asp | Segregating | 0/960 |
| 219567104 | *OBSL1* | c.1860A>C | p.Ala620Ala | Non-segregating | _ |
| 219633995 | *SLC4A3* | c.1642+16C>T | _ | Non-segregating | _ |
| 222569954 | *FARSB* | c.*1917G>C | _ | Non-segregating | _ |
| 224511266 | *CUL3* | c.883+88T>C | _ | Non-segregating | _ |
| 224585358 | *CUL3* | c.-349C>G | _ | Non-segregating | _ |
| 224770741 | *DOCK10* | c.6205-96G>A | _ | Non-segregating | _ |
| 224823685 | *DOCK10* | c.3037-38A>T | _ | Non-segregating | _ |
| 227008368 | *COL4A4* | c.4523-64G>T | _ | Non-segregating | _ |
| 227088729 | *COL4A4* | c.1547A>T | p.Asp516Val | Non-segregating | _ |
| 227357595 | *MFF* | c.*478C>T | _ | Non-segregating | _ |
| 227698936 | *SLC19A3* | c.779A>T | p.Asp260Val | Non-segregating | _ |
| 227817435 | *CCL20* | c.*352T>A | _ | Non-segregating | _ |

### 3.2.5 Minigene splicing analysis of the intronic variant

Minigene assay was undertaken to study the effect of the intronic variant c.2466+89G>C in *USP37* on splicing. The exon close to the intronic variant along with 500 base pairs upstream and downstream of the exon-intron boundary was cloned in the exon trapping vector pSPL3 (Figure 4). The wild type and variant minigenes were transfected in HEK293 cells. Total RNA isolation was done 48 hours post-transfection and cDNA synthesis was done using a superscript III first-strand synthesis system. The primers V1 and V2, which lie in the vector-specific exons V1 and V2, were used to amplify the minigene. If splicing takes place, the intron between the exons V1 and V2 gets spliced out. In the minigene that we have constructed, under wild-type conditions, the *USP37* exon will also be present between the vector-specific exons V1 and V2. We find that the variant minigene behaves the same as the wildtype and there are no splicing errors (Figure 4).

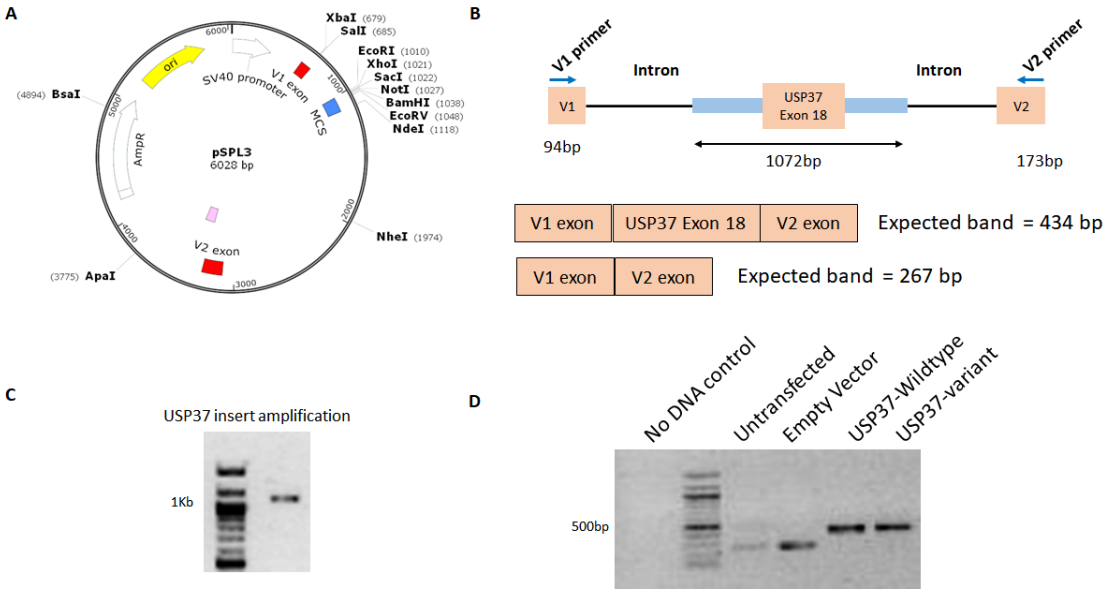



**Figure 4:** Schematic of USP37 mini-gene assay. **A.** Plasmid map of the pSPL3 vector. **B.** Schematic of the USP37 minigene. V1 and V2 are the exons present in the vector. USP37 exon 18 with the spanning intronic regions are cloned between V1 and V2. A product of 434 base pairs is expected when splicing takes place. **C.** PCR amplification of the USP37 region of interest harbouring the intronic variant from genomic DNA. **D.** RT-PCR analysis of cDNA from wildtype (WT) and variant-harbouring minigene transfected cells.

### 3.2.6 Identification of DES rare and conserved missense variant

The intronic variant c.2466+89G>C in *USP37* showed no effect on splicing. The only variant identified in the coding exons, c.966A>T in *DES* gave rise to a missense variant, p.Glu323Asp. The amino acid Glu322 is an evolutionarily conserved residue (Figure 5) and the variant p.Glu323Asp showed deleterious effects on the protein's function on performing in-silico predictions using different tools.

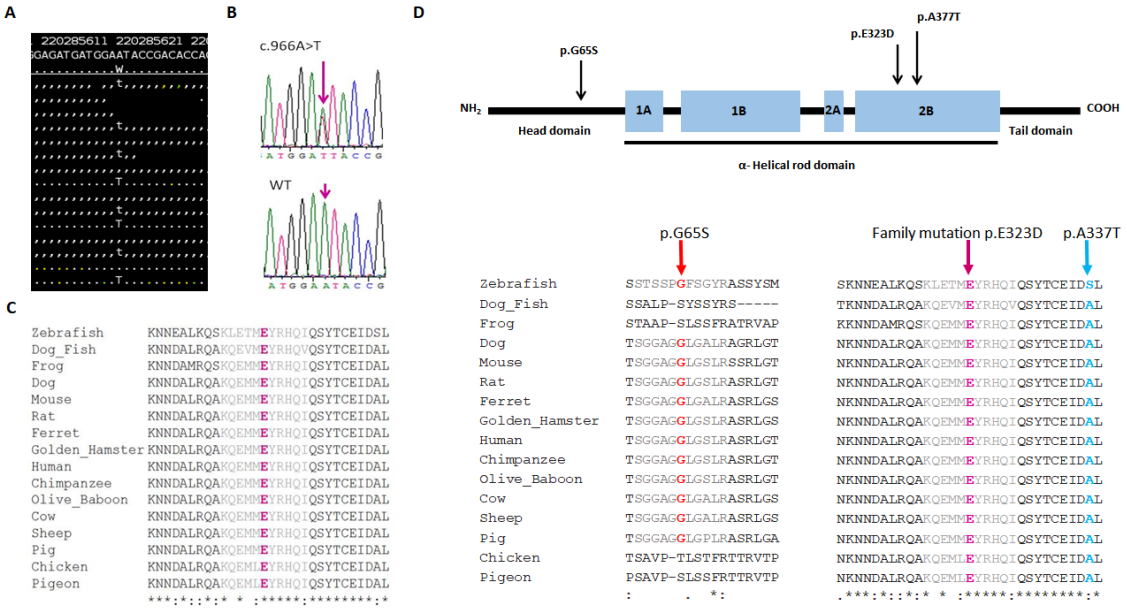



**Figure 5:** *DES* mutations associated with epilepsy: **A.** Alignment of the genome sequence reads against human genome reference sequence as viewed in tview of SAMtools v0.1.7a. The snapshot here shows that not all reads have the A>T variant, which indicates that this variant is heterozygous. The dot and comma represent the matched reference allele and the alphabet shows the variant allele. **B.** Electropherograms of sequence with wild-type (A>A) and variant allele (A>T) for the c.966A>T variant. The arrow indicates the nucleotide showing heterozygous variant shown by the presence of two peaks. **C.** DES protein architecture with the N-terminal head domain, C-terminal tail domain and middle alpha-helical rod domain is represented with the location of the three missense variants, p.G65S, p.E323D, p.A337T. **D.** Multi-species protein sequence alignment of desmin by Clustal Omega. The amino acid residues glycine (G) 65, glutamine (E) 323 and alanine (A) 337 are highlighted in red, pink, and blue, respectively.

### 3.2.7 *DES* rare variants among JME patients examined

*DES* gene encompasses an 8-kilobase genomic region and harbours 9 exons. DES encodes a 470 amino acid type III intermediate filament protein. The complete transcript of *DES* was examined in 192 JME patients and ethnically matched control individuals. In this analysis, 10 rare variants, 7 of which are novel, were identified (Table 4).

**Table 4: *DES* rare variants**

| Genomic Position (Grch38) | cDNA position | region | Amino acid position | Rs ID | Occurrence in JME cohort | MAF gnomAD v2 |
|---|---|---|---|---|---|---|
| 219418655 | c.193G>A | Exon | p.Gly65Ser | rs397516692 | 1/192 | 0.00002 |
| 219420174 | c.640+19C>T | Intron | _ | Novel | 1/192 | _ |
| 219420193 | c.640+37C/- | Intron | _ | Novel | 2/192 | _ |
| 219420659 | c.898+2GG/- | Intron | _ | rs397516699 | 2/192 | 0.002 |
| 219420855 | c.924C>T | Exon | p.Asp308= | Novel | 4/192 | _ |
| 219420939 | c.1009G>C | Exon | p.Ala337Thr | rs59962885 | 1/192 | 0.00006 |
| 219423870 | c.1288+50C>T | Intron | _ | Novel | 1/192 | _ |
| 219423976 | c.1288+156G>A | Intron | _ | Novel | 1/192 | _ |
| 219426015 | c.*25G>A | 3'UTR | _ | Novel | 1/192 | _ |
| 219426675 | c.*685G>A | 3'UTR | _ | Novel | 1/192 | _ |

**Table 5: *DES* missense variants- in-silico predictions**

| cDNA position | Amino acid position | MAF in-house controls | MAF gnomAD v2 | SIFT | PolyPhen-2 | Mutation Taster |
|---|---|---|---|---|---|---|
| c.193G>A | p.Gly65Ser | 0/576 | 0.00002 | tolerated (1) | benign (0.009) | Polymorphism (0.66) |
| c.969A>T | p.Glu323Asp | 0/576 | _ | tolerated (0.32) | possibly damaging (0.458) | disease causing (0.87) |
| c.1009G>C | p.Ala337Thr | 1/576 | 0.00006 | deleterious (0.02) | probably damaging (0.947) | disease causing (0.99) |

These variants were either absent or rare (MAF<0.005) in 1000 in-house control chromosomes and databases dbSNP151, 1000 Genomes, ExAC and gnomAD. Out of these, c.193G>A (p.Gly65Ser), and c.1009G>C (p.Ala337Thr) give rise to non-synonymous changes in the DES protein (Figure 6, Table 5). Of the total three non-synonymous variants identified in DES, p.Gly65Ser lies in the head domain of DES,

and p.Ala337Thr and p.Glu323Asp lie in the central alpha-helical rod domain. Conservation analysis across species showed that the residues Glu323 and Ala337 are highly conserved across species, with Gly65 also showing moderate conservation in mammals (Figure 6). The patients harbouring the DES non-synonymous variants all manifest GTCS and myoclonic seizures, which are hallmarks of JME (Table 6). The proband also shows tremors and absence seizures.

### 3.2.8 DES brain expression

We find DES expression in human brain samples. Marathon-Ready full-length human brain cDNAs were used to amplify DES transcripts. Primers binding in exons 1 (forward primer) and exon 5 (reverse primer) were used for amplification and an expected band of 498 base pairs was obtained, as seen in the agarose gel. The amplified products were Sanger sequenced for confirmation (Figure 6). The full-length DES transcript was amplified from human brain hippocampus and cloned into the pcDNA3.1 vector.

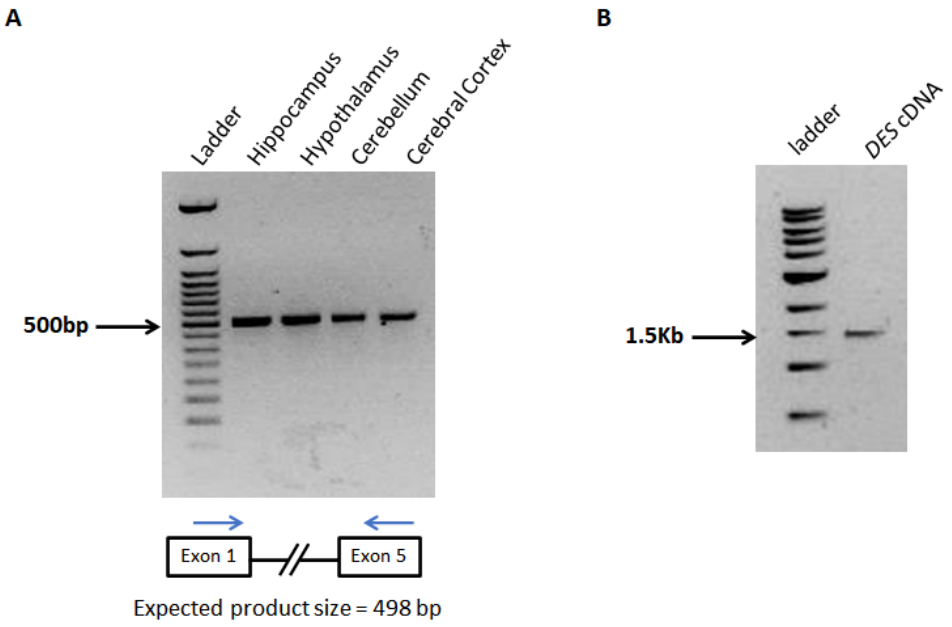



**Figure 6: A.** *DES* cDNA was amplified from total cDNA of various human brain regions. Primers binding in exons 1 (forward primer) and 5 (reverse primer) were

used for amplification and an expected band of 498 base pairs was obtained, as seen in the agarose gel image (arrow). **B. Full-**length *DES* cDNA amplification from human hippocampus total cDNA (Clontech). An expected band size of 1.4 Kb was obtained. The amplified cDNA is cloned into pcDNA3.1(+).

### 3.2.9 DES over-expression and cellular localization

On conducting an immunofluorescence study of wild-type and mutant desmin proteins in HEK293 cells, it was observed that desmin forms dense cytoplasmic filamentous structures. There is no significant difference between the wild-type and mutant desmin filament structures (Figure 7-a). On checking the levels of wild-type and mutant desmin protein, using western blot analysis, no significant difference was seen (Figure 7-b).

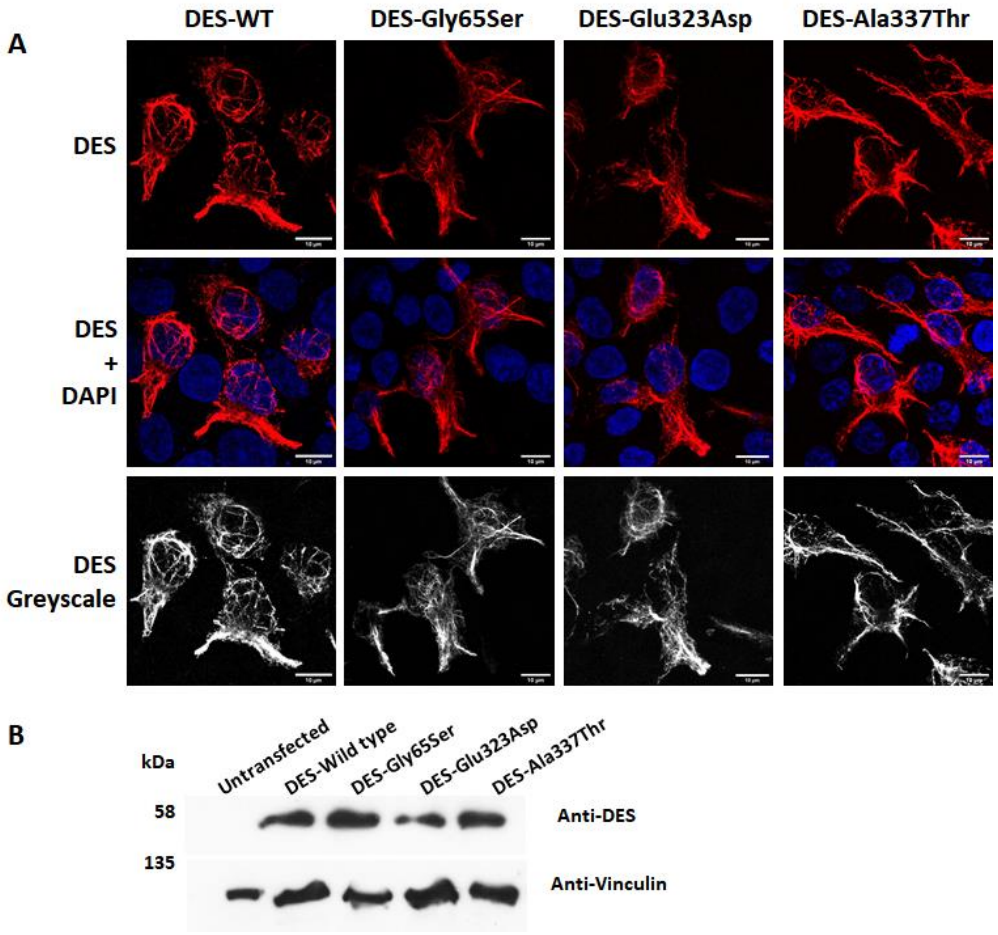


**Figure 7:** Over-expression of DES (desmin) in HEK293 cells. **A.** Immunofluorescence analysis of DES wild type and variant proteins (p.Gly65Ser, p.Glu323Asp, p.Ala337Thr) in HEK293 cells. DES (red) forms cytoplasmic filamentous structures. There is no significant difference between the wild type and variant DES filament structures. **B.** DES wild-type and variant protein expression analysis in HEK293 cells, showing no major protein expression level difference between wild type and mutant protein.

### 3.2.10 Desmin and vimentin co-localization analysis

Overexpressing DES in cells expressing endogenous vimentin gives rise to the integration of the DES filaments with the vimentin filaments. Mutant desmin that fails to make extensive filamentous structures leads to the vimentin filaments to re-organize in the nuclear perinuclear area and in some cases completely segregate from the vimentin network (Bär et al., 2006). To test the *DES* human variants identified in JME patients for their effect on vimentin filaments, the wildtype and mutant proteins were over-expressed in SH-SY5Y cells. These cells express endogenous vimentin.

We observe that the wildtype, as well as all three human variants (p.Gly65Ser, p.Glu323Asp and p.Ala337Thr), were able to form filaments that integrate with the vimentin network and there is no disruption in the vimentin network and localization. The amount of co-localization between desmin and vimentin was quantified by calculating the Pearsons's correlation coefficient, using image J plugin coloc-2, for each cell expressing both these proteins. When compared with wildtype desmin, the variant protein does not have any impact on the co-localization of vimentin with desmin (Figure 8).

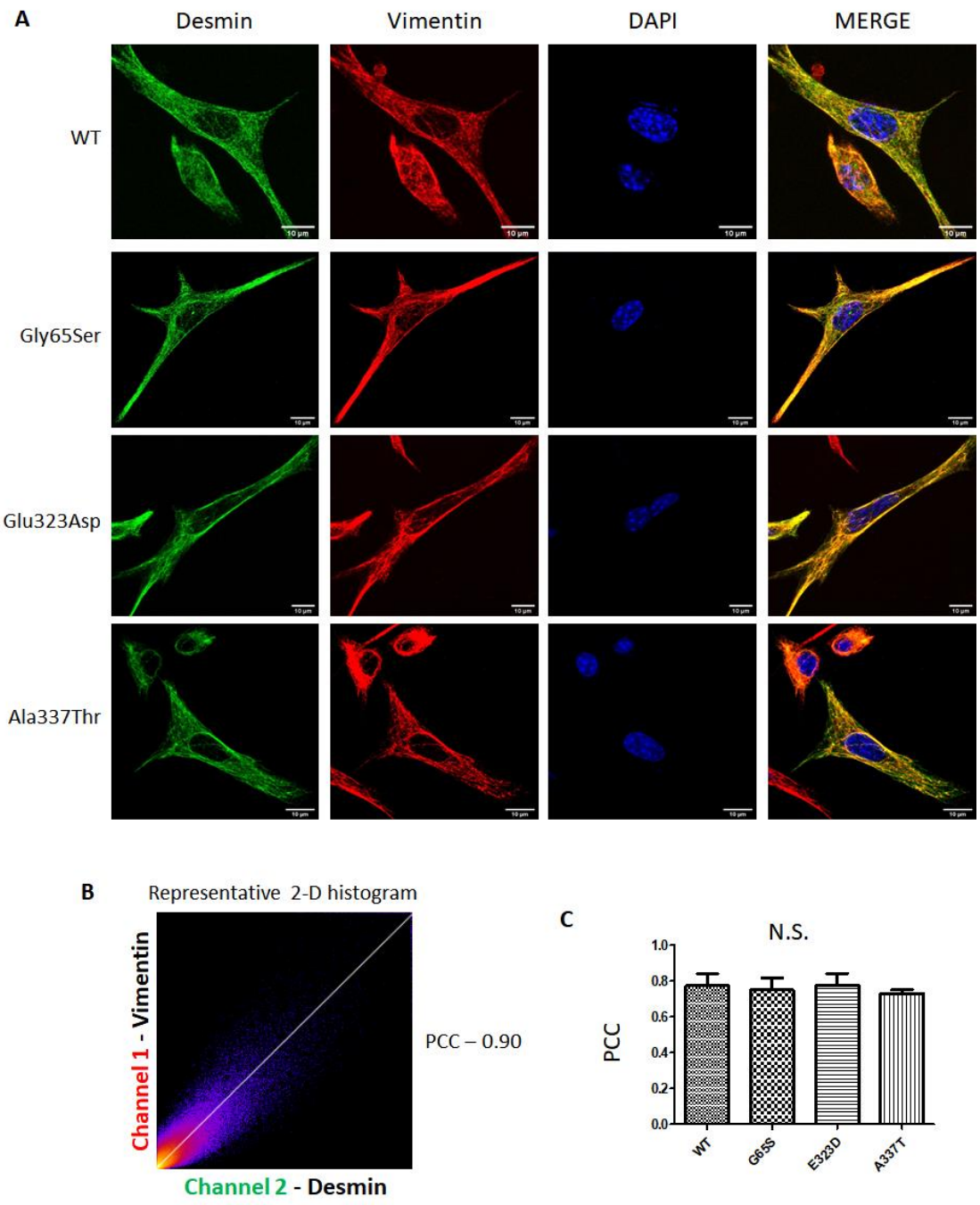



**Figure 8:** Over-expression of desmin in SH-SY5Y cells. **A.** Desmin (green) wild-type and mutant proteins are able to form filamentous structures which integrate with vimentin (red) filaments. **B.** Desmin and Vimentin colocalization represented using a 2-D histogram made using Image J Coloc-2 plugin. C. Pearson's correlation coefficient (PCC) was calculated for desmin and vimentin co-expressing cells using ImageJ Coloc-2 plugin. PCC from three independent experiments and 20 cells per experiment, per construct, were calculated. One-way ANOVA followed by Dunnet's multiple comparison test was used to calculate the statistical significance. No difference in desmin and vimentin colocalization was observed in cells expressing either wildtype or variant desmin protein.

## 3.3 Discussion

In this chapter, findings from the whole exome and genome sequencing experiment, to identify the potentially causative gene(s) at the 2q33-q36 locus and preliminary characterization of these variants have been presented. The analysis of NGS data from affected individuals of Family SCT135, revealed a rare, possibly pathogenic variant in *DES*, which appeared suitable for further studies.

The missense mutation identified in *DES* substitutes an evolutionarily conserved amino acid residue, glutamic acid at position 323 to aspartic acid; and was predicted to have a deleterious effect on the protein structure and function. This heterozygous variant was found in all the affected individuals of SCT135 and was absent in our control cohort and public databases such as dbSNP and GnomAD.

*DES* codes for a 470-amino acid, muscle-specific, type III intermediate filament. There are three major domains to the desmin protein: a conserved alpha-helix rod, a variable non-alpha helix head, and a carboxy-terminal tail. The rod domain, which consists of 308 amino acids with parallel alpha-helical coiled-coil dimers and three linkers to disrupt it, connects to the head domain. The head domain has many arginine, serine, and aromatic residues, is 84 amino acids long and is important in filament assembly and dimer-dimer interactions of desmin. The tail domain is responsible for the integration of filaments and interaction with proteins and organelles. Desmin is only expressed in vertebrates, however, homologous proteins are found in many organisms. Desmin is a subunit of intermediate filaments in cardiac muscle, skeletal muscle and smooth muscle tissue (Bär et al., 2004b). Mutations in the *DES* gene have been reported to cause human skeletal as well as cardiac myopathy, termed desmin-related myopathies (Clemen et al., 2013).

As a primary step in examining the role of *DES* in a brain disorder, its expression at transcript level was checked by qPCR in different human

brain sub-regions (Figure 6). Further, while exploring the role of *DES* as an epilepsy-linked gene, we identified two additional JME-associated novel/rare mutations, p.Gly65Ser and p.Ala337Thr in a set of JME patients. Taken together, three mutations identified in this study alter conserved amino acid residues in desmin and two of the three variants, p.Glu323Asp and p.Ala337Thr, were predicted to have a damaging effect on the protein function. These missense mutations do not seem to affect the filament formation, expression, or colocalization with vimentin of the desmin protein.

Apart from being a very well-studied skeletal and cardiac muscle protein, desmin is shown to be expressed in pericytes, which are endothelial cells that envelope the blood vessels of the brain (Hughes and Chan-Ling, 2004) (Figure 9).

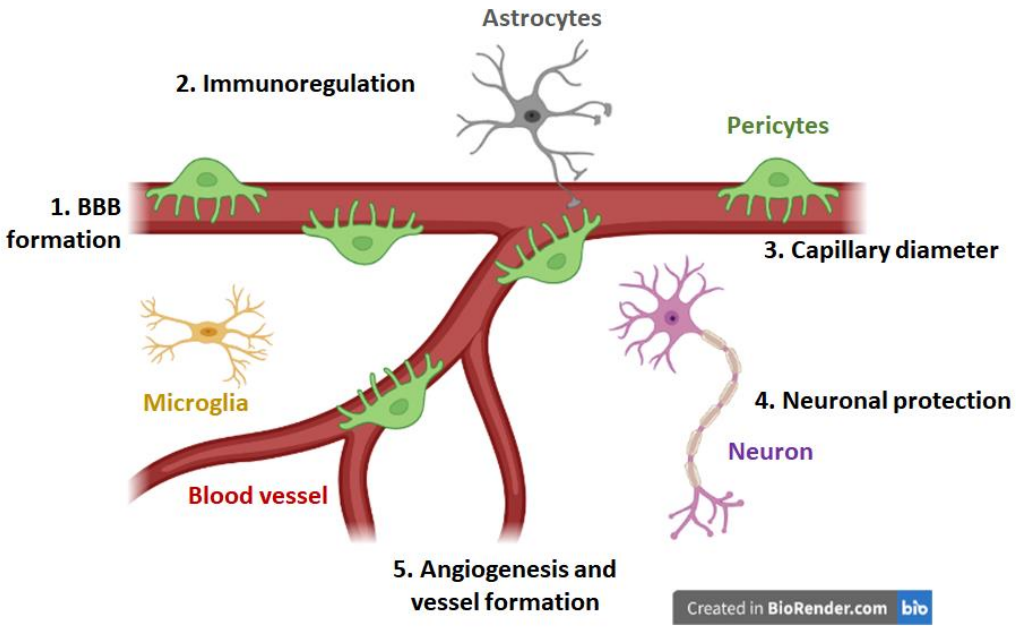



**Figure 9:** Different types of cells surrounding the blood vessels in the brain. Pericytes (shown in green colour) are the cells in which desmin is known to be expressed. Functions of brain pericytes: 1. Blood-brain barrier (BBB) formation, 2. Immune regulation, 3. Regulating capillary diameter, 4. Neuronal protection, 5. Controlling angiogenesis and vessel formation.

Pericytes control key neurovascular functions that are necessary for proper neuronal structure and function (Bell et al., 2010). Pericytes are an under-studied cell type located on capillaries. They are of crucial importance in regulating diverse microvascular functions, such as angiogenesis, the blood-brain barrier, capillary blood flow and the movement of immune cells into the brain (Figure 9). Recent studies have suggested that pericytes play a crucial role in neurological diseases (Cheng et al., 2018). It has been shown in animal models of epilepsy that focal capillary constrictions occur in close spatial association with pericytes. These constrictions are surrounded by regions of neuronal damage in the hippocampi of epileptic mice (Leal-Campanario et al., 2017). In status epilepticus mice brain, loss of pericytes is proportional to seizure severity and vascular pathology (Arango-Lievano et al., 2018). Pericytes have also been shown to regulate the health of neurons by secreting factors such as nerve growth factor (NGF), brain-derived nerve growth factor (BDNF), and pleiotrophin (Nikolakopoulou et al., 2019). These studies have shown that pericytes play a pivotal role in seizures and are potential targets for reducing pathophysiology. Cultured pericytes can be used as a model system to perform in-vitro studies to understand the effect of the *DES* patient variants on the pericyte's normal functioning (Tigges et al., 2012).

Desmin interacts with CMYA5, a gene reported to be associated with schizophrenia (Han et al., 2018). CMYA5 patient mutations have been shown to disrupt its interaction with desmin, which has been proposed by the authors as the probable mechanism by which neuritogenesis might be disrupted (Hsiung et al., 2019). Through myospryn, the protein encoded by CMYAS, desmin associates with the BLOCK-1 complex (Capetanaki et al., 2007). BLOCK-1 complex regulates endosomal trafficking that is involved in synaptic transmission and neurodevelopment as well as regulation of the cell-surface D2 dopamine receptor (DRD2), the biogenesis and fusion of synaptic vesicles, and

neurite outgrowth (Mullin et al., 2011). Even though desmin has been largely studied in the context of its role in muscle and cardiac cells, exploring desmin's functions in pericytes and other brain cells and understanding how desmin might be regulating any of the BLOCK-1 mediated processes may be considered to understand its role in epilepsy mechanism.

# Chapter 4

## Characterization of *TMEM171*, a novel JME gene identified at the EJM4, 5q12-q14 locus.

### Summary

Previous work in the laboratory has identified a genetic locus, EJM4 (5q12-q14) in a three-generation south Indian family with several of its members affected with JME. I carried out whole-genome sequencing to find a potential causative factor underlying JME in the family. Analysis of the WGS variants at the 5q12-q14 region revealed a rare variant, p.Gly159Ala in the *TMEM171*gene. The c.476G>C variant converts a highly conserved residue, glycine to alanine (p.Gly159Ala) in the predicted TMEM171 protein sequence. Further, to identify additional epilepsy-associated variants in TMEM171, I examined the gene sequence in a set of 72 JME patients. Besides p.Gly159Ala, two new variants: p.Gln90Arg and p.Arg156Trp were found. TMEM171 encodes a proline-rich transmembrane protein with an unknown biological function. TMEM171 protein has four predicted transmembrane domains in the N-terminal region and a C-terminal tail in the cytoplasm. It localizes to the plasma membrane. RT-PCR and Western analysis suggests the expression of TMEM171 in different human brain regions. Though the epilepsy-associated TMEM171 variants do not seem to affect its membrane localization or expression levels, they may have damaging effects on the functional/cellular physiology of the brain. To get an insight into TMEM171 protein's functions and pathways that it may be a part of, I carried out immuno-pulldown experiments and identified SNAP25 and SNAPIN as TMEM171-interacting partners. SNAP25 is a member of the SNARE complex. SNAPIN associates with the SNARE complex proteins and the BLOC-1 complex. Further delineation of TMEM171's role in these pathways requires further studies.

## 4.1 Materials and methods

### 4.1.1 Family ascertainment

The three-generation family, GLH5 discussed in this chapter was first reported by Kapoor and colleagues (Kapoor et al., 2007)(Figure 1). Of the 24 family members, 21 members had consented to participate in the study. They were clinically evaluated by a neurologist and diagnoses established according to the guidelines of the International League against Epilepsy (ILAE; Commission on Classification and Terminology of the International League Against Epilepsy 1989) (Scheffer et al., 2017b).

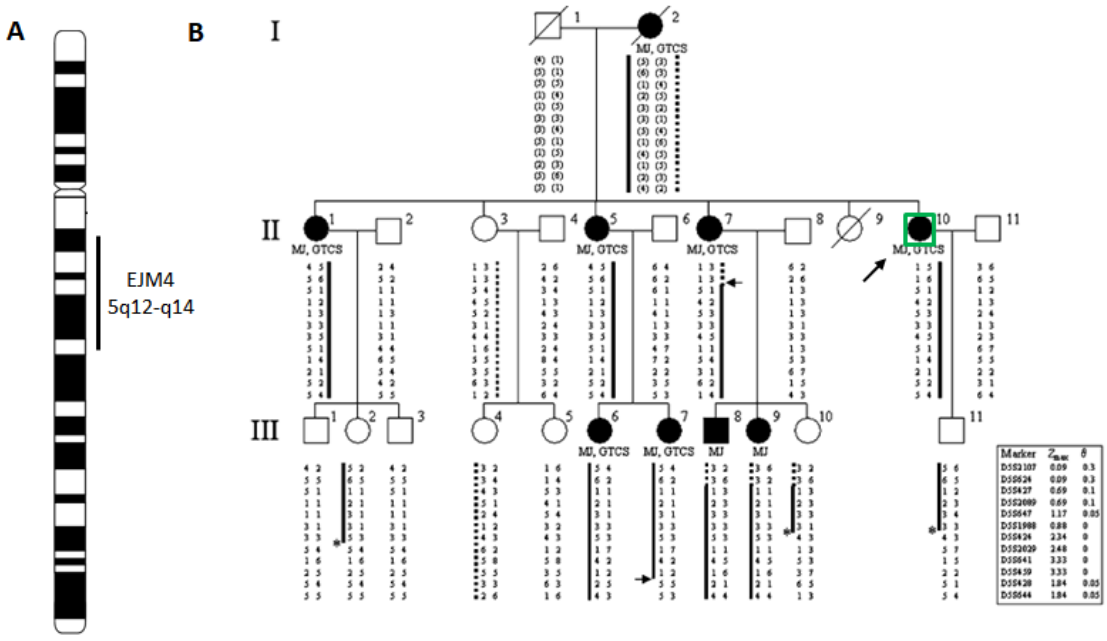



**Figure 1: The *EJM4* locus in Family GLH5: A**. Ideogram of chromosome 5 representing the location of the JME locus, *EJM4* (5q12-q14). **B.** Pedigree of GLH5 depicting the disease-linked haplotype (*EJM4*) shared among its affected members and three unaffected members (III:2, III:10 and III:11). Males and females are denoted by squares and circles, respectively. Filled symbols represent affected individuals and empty, unaffected ones. Haplotypes are shown below the symbols. Upper and lower recombination boundaries denoting the critical region are indicated by arrows in individuals II:7 and III:7. DNA from Individual II:10, marked with a green box, was used for whole-genome sequencing. Clinical features of affected subjects are indicated along with the symbols (MJ: Myoclonic jerks, GTCS: Generalized tonic-clonic seizures). Adapted from Kapoor et al 2007.

Family GLH5 was ascertained through proband (II:10), who presented distinct characteristics of JME. She had a history of early morning, myoclonic jerks (MJ) and generalized tonic-clonic seizures (GTCS) beginning at the age of 14 and 20, respectively.

Additional affected members of the family had histories of myoclonic jerks, accompanied by GTCS, with an age of onset ranging from 12-18 years. The unaffected subjects reported no history of seizures. Diagnoses for the three deceased members were inferred based on the information provided by other family members. For genetic analysis, subjects, I:2, II:1, II:5, II:7, II:10, III:6, III:7, III:8 and III:9 were considered affected, and remaining members, unaffected.

In addition, a set of 552 (480 patients analysed earlier in the lab; Manpreet Kaur, PhD thesis, 2014) JME patients were ascertained from the NIMHANS and SCTIMST and were diagnosed according to the ILAE guidelines. The control cohort comprised 480 healthy individuals from southern parts of India, above 20 years of age and with no apparent family history of neurological disorders.

### 4.1.2 Whole-genome sequencing (WGS)

#### 4.1.2.1 Library preparation and sequencing

Three micrograms of genomic DNA was fragmented using Covaris to generate 300-400 base pair fragments. 100 ng of fragmented DNA was used to generate a sequencing library using NEBNext Ultra II DNA Library Prep Kit for illumina (New England Biolabs). In brief, the fragmented DNA was subjected to end repair followed by A – tailing and adapter ligation. Ampure bead-based size selection was done to obtain the library of the desired size. The size selected DNA was enriched by PCR amplification using illumina index adapter primers (*P7 adapter 5'* – AGATCGGAAGAGCACACGTCTGAACTCCAGTCA – 3', *P5 adapter 5'* – AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT – 3'). The amplified product was purified using Ampure beads to remove unused primers.

The libraries were quantitated using Qubit DNA HS quantitation assay (Thermo Scientific) which specifically quantitates dsDNA. The library quality was checked using Agilent Bioanalyzer DNA 1000 kit (Figure 2). The QC passed libraries were sequenced using illumine HiseqX sequencer (illumina).

#### 4.1.2.2 Sequence data quality check (QC), processing and alignment

The sequence data quality was checked using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and MultiQC (Ewels et al., 2016). The data was checked for base call quality distribution, % bases above Q20, Q30, % GC, and sequencing adapter contamination. The sequence data was processed using Trimgalore (https://github.com/FelixKrueger/TrimGalore) to remove adapter sequences and low-quality reads. QC passed reads were mapped to the human reference genome build GRCh38 that was provided in GATK (Genome Analysis Toolkit) resource bundle using the BWA-MEM algorithm (Li and Durbin, 2009b) with default parameters. The alignments were sorted, indexed and PCR duplicates were marked and removed using sambamba, a high-performance robust tool and library for SAM/BAM files (Tarasov et al., 2015).

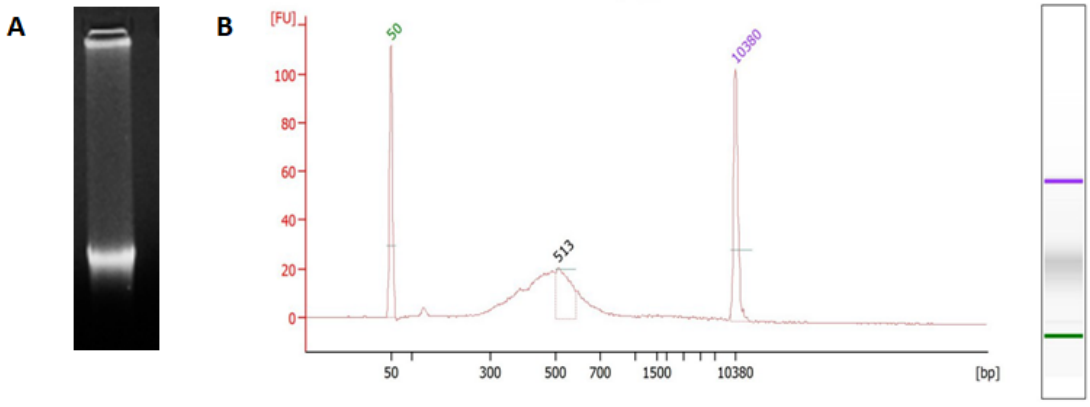



**Figure 2 – DNA and library quality check: A.** The genomic DNA was checked for its integrity and quality by electrophoresis on 0.8% TAE-agarose gel. This genomic DNA was used for library preparation. The libraries were checked for their quality and size distribution using Agilent Bioanalyzer DNA 1000 kit. **B.**

Size distribution of the whole genome sequencing library made using the NEBNext Ultra II DNA Library Prep Kit for illumina.

#### 4.1.2.3 Variant calling, annotation and filtering

GATK (McKenna et al., 2010) best practice workflow for germline short variant discovery (SNPs + INDELs) using GATK v4 was followed for variant calling (https://software.broadinstitute.org/gatk/best-practices/). The process, in brief, base qualities in the bam file were re-calibrated and genomic variant call format (GVCF) was generated using Haplotype-Caller. The GVCF was genotyped to generate variant call format (VCF). SNPs and INDELs were recalibrated separately by providing variants from dbSNP, 1000 Genome phase1, HapMap, Omni SNPs and Mills gold standard INDELs as known and training sets. The filter threshold was set to retain 99% of true variants, which provides the highest sensitivity while still being acceptably specific.

Functional annotation for the variants identified in the genome sequencing experiment was performed using Ensembl Variant Effect Predictor (VEP) build 94 (McLaren et al., 2016). The input was provided in the form of a tab-delimited vcf file. A transcript database from both GENCODE and NCBI was used. Allele frequencies were taken from dbSNP151, 1000-Genomes, Exome Aggregation Consortium (ExAC) and Genome Aggregation Database (gnomADv2, exomes and genomes; gnomADv3, genomes). dbSNFPv3.5 (Liu et al., 2011), a database developed for functional prediction and annotation of all potential non-synonymous single-nucleotide variants in the human genome, was also used for annotation. Samtools (Li et al., 2009) was used to view the whole genome sequence data aligned with the human reference sequence (Grch38) and each exon in the region of interest on chromosome 5 was manually checked for its coverage and read depth.

The identified variants were filtered for minor allele frequency in the databases, dbSNP151, 1000 Genomes project (Auton et al., 2015), Exome Aggregation Consortium (ExAC) (Lek et al., 2016) and the

Genome Aggregation Database (gnomAD) (Karczewski et al., 2020) for their minor allele frequency (MAF) to be less than 0.005.

### 4.1.3 Sanger-based sequence validation of the WGS rare variants

All novel and rare variants identified in the WGS dataset were validated by Sanger sequencing. Primers were designed spanning the variant-carrying regions. The Sanger confirmed variants were further analysed in the family for their segregation with the disease phenotype. The segregating variants were examined in 480 ethnically matched set of normal individuals to check their allele frequencies.

### 4.1.4 TMEM171 mutation analysis

*TMEM171* which carried the disease co-segregating rare variant was further analysed in a set of 552 JME affected individuals (72 patients analyzed in this study) and ethnically matched control individuals. The entire gene sequence comprising all exons, flanking intronic boundaries, and the 5'- and 3'- untranslated regions were Sanger sequenced to identify additional rare variants.

### 4.1.5 Bioinformatics analysis

The complete gene sequence of *TMEM171* was downloaded from NCBI (NC_000005.10:73120575-73131814 Homo sapiens chromosome 5, GRCh38.p13 Primary Assembly). Primers were designed using the online tools Primer3 (Untergasser et al., 2012) and OligoCalc (Kibbe, 2007). Protein sequences for TMEM171 across different species were downloaded from NCBI to perform multiple sequence alignments using the tool Clustal Omega (Madeira et al., 2019).

### 4.1.6 Plasmids and antibodies

TMEM171 cDNA (Origene) was sub-cloned into p3X-FLAG-CMV10 (3X-FLAG tag fused to the N-term of the protein) and pcDNA3.1+ (untagged) vectors. Site-directed mutagenesis was performed using QuikChange II

XL Mutagenesisreagents (Agilent Technologies) to generate patient-specific variants p.Gln90Arg, p.Arg156Trp and p.Gly159Ala. To study TMEM171transmembrane domains, premature stop sites were introduced in the cDNA (Figure 5b). Syntaxin4-myc-His (Cat. No. 12377) and pEGFP-Snapin (Cat. No. 118742) plasmids were procured from Addgene.

The antibodies used in this study were: anti-TMEM171 (monoclonal, mouse raised; Sigma, US, SAB1401999), 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), goat anti-rabbit Alexa Fluor 488 (Thermo Fisher Scientific, A11008), goat anti-mouse Alexa Fluor 568 (Thermo Fisher Scientific, A11001), goat anti-rabbit IgG – HRP (Genei, India, HPO3), goat anti-rabbit IgG – HRP (Genei, HPO5).

### 4.1.7 Cell culture and transfection

HEK-293 cells were maintained in DMEM (high glucose, Sigma) supplemented with 10% heat-inactivated fetal bovine serum (MP Biomedicals), 2 mM L-glutamine (Sigma) and antibiotics (100 U/ml Penicillin and 10 mg/ml streptomycin; Sigma) 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) coated18-mm glass coverslips and transiently transfected with 1 µg plasmid per well using Lipofectamine 2000 transfection reagent (Thermo Fischer Scientific). Protein expression was checked 24 to 48 hours post-transfection.

### 4.1.8 Western analysis

For isolating whole-cell protein, the cells/tissues were lysed using lysis buffer (150 mM NaCl, 10 mM Tris-pH 7.5, 0.1% SDS, 1% Triton X-100, 1% deoxycholate and 5 mM EDTA). The total protein content of the cells was collected by centrifuging the lysed cells at 13,000 rpm for 20 minutes at 4$^{o}$C and was quantified using a Pierce BCA protein assay kit

(Thermo Fischer Scientific). 50 ug of total protein was separated on a 10% SDS-polyacrylamide gel (Laemmli, 1970) followed by transfer to nitrocellulose membrane (Pall) at 20 volts for 1 hour. The membrane was blocked with 5% skimmed milk for 4 hours at 4°C followed by incubation with primary antibodies, anti-TMEM171 (1:500) or anti-gamma tubulin (1:5000) or anti-alpha tubulin (1:5000) or anti-actin (1:1000) or anti-FLAG (1:500) or anti-GFP (1:2500), at 4°C for 16 hours. HRP-conjugated secondary antibodies (1:5000) were used to probe the membrane at 4°C for 4 hours. The membrane was washed between each incubation step with 0.05% PBST (0.05% tween-20 in 1X PBS). The protein bands were detected using clarity western ECL Substrate (Biorad).

### 4.1.9 Immunocytochemistry

To perform immunocytochemistry analysis, cells were grown on 18 mm poly-L-lysine coated glass coverslips. The cells were fixed with ice-cold methanol at -20°C for 5 minutes. The fixed cells were blocked with 3% BSA (bovine serum albumin) for 1 hour at room temperature. Incubations with primary antibodies, anti-TMEM171 (1:500) or anti-FLAG antibody (1:500), and fluorescently labelled secondary antibodies (1:500), diluted in 1% BSA, were done for 1 hour each at room temperature. Nuclear staining was performed using 1 µg/ml 4, 6-diamidino-2-phenylindole (DAPI) for 15 minutes at room temperature. The coverslips were mounted on glass slides using Polyvinyl alcohol mounting medium (Sigma) and sealed with transparent nail polish. Imaging was done using Zeiss LSM-510 meta or -880 confocal microscope.

### 4.1.10 Immuno-pulldown assay

Total protein was isolated from cells on ice, using mild lysis buffer (150 mM KCl, 25 mM tris-pH 7.4, 5 mM EDTA, 0.5% NP40, protease inhibitor cocktail (Roche) and 1 mM PMSF). 20 µl Dynabead protein G mix (Thermo Fisher Scientific) was allowed to bind with 2 µg anti-FLAG

antibody for 4 hours with rotation (10 rpm) at 4°C. The bead-antibody complex was then incubated with 2 mg protein lysate for 12 hours at 4°C. Post-incubation, the bead-antibody-antigen complex was washed 4 times for 1 minute each, on ice, with 200 µl cold 1X PBS. Elution was done by boiling the beads in 50 µl 2X dye for 10 minutes.

## 4.2 Results

### 4.2.1 Whole-genome sequencing

To examine the critical region, 5q12-q14, whole genome-based sequencing was undertaken. This experiment generated a total of 90 GB sequence, as 0.6 billion paired-end reads of 150 bp, from sequencing on an illumine Hiseq X machine (illumina). More than 97% of bases were of high quality (Phred score>20) and an average read depth of 30X was obtained (Table 1). A total of 33,429 variants were identified out of which 559 variants were either absent or present at a MAF<0.005 across the databases dbSNP151, 1000-Genomes, ExAC and the gnomAD.

**Table 1: Sequence coverage summary for the whole-genome sequencing experiment (WGS)**

| **Sample** | **III:6** | |
|---|---|---|
| **Target region** | **Genome** | **5q12-q14 Locus** |
| Total number of reads | 678.8 million | |
| Reads aligned to complete genome | 678.3 million | |
| **% Reads aligned to complete genome** | 99.92 | 96.84 |
| Target Sequence Length | 3.2 Gb | 25 Mb |
| Total target covered | 3.0 Gb | 24 Mb |
| **% Target Covered** | **94.19** | **95.78** |
| % Target Covered with at least 5X read depth | 93.62 | 95.45 |
| % Target Covered with at least 10X read depth | 90.81 | 95.1 |
| % Target Covered with at least 20X read depth | 85.46 | 87.75 |
| % Target Covered with at least 30X read depth | 58.81 | 60.96 |
| **Average read depth** | **31.3** | **30.2** |

Percentage of genome/ region coverage calculated after sequences were aligned to the human genome reference sequence (GRCh38).

### 4.2.2 Rare variant analysis and prioritization

Of the 559 rare variants identified at the 5q12-q14 locus, 308 lie in protein-coding genes and 468 in non-protein-coding regions such as micro-RNAs, promoters, enhancers and transcription-factor binding sites. A variant at the same genomic position can have multiple annotations, which is the reason for the total number of rare variants being 559 and not 776 (308 + 468). The variants in the non-protein-coding regions were first checked for their conservation using the CADD score. Variants with CADD<20 were removed from our analysis. The variants lying in conserved regions were then analysed for their potential pathogenicity using the FATHMM-MKL prediction tool that combines sequence conservation scores with ENCODE-based functional annotations (Shihab et al., 2015). The variants were also manually checked in the UCSC genome browser to see if they lie in repeat regions that are prone to sequencing errors. None of the variants in the non-protein-coding regions were suggested to have any pathogenic effects and were hence not taken forward for further analysis.

The variants in protein-coding transcripts (308) were further classified based on their consequences. There were 270 variants in intronic regions, 12 in UTR regions, 23 in regions upstream or downstream of genes, 4 variants gave rise to missense changes and 1 variant to a synonymous change in the protein sequence. The variants in low complexity regions, which are rich in polymorphisms, regions beyond 100 base pairs from the exon-intron boundary and variants in regions upstream or downstream of genes were excluded from further analysis (Figure 3).

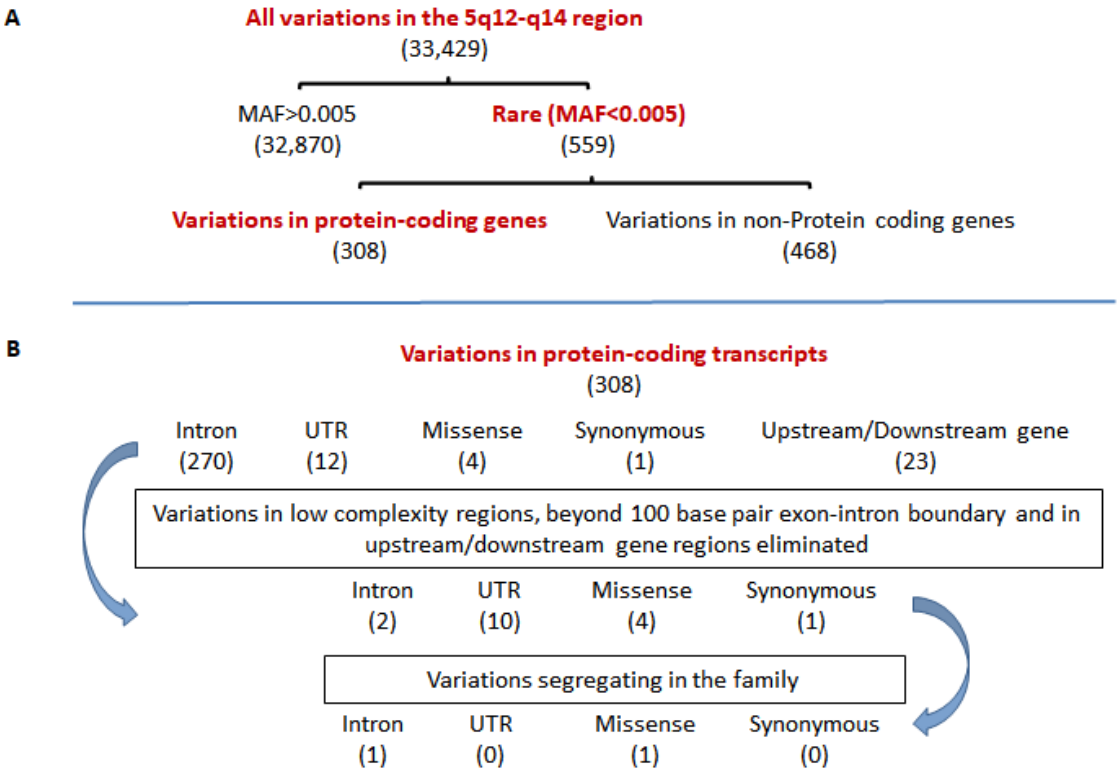



**Fig 3: Whole-genome sequencing variants filtration and prioritization flow chart.** The variants identified in the 5q12-q14 locus were first filtered based on their minor allele frequency across databases. Variants with frequency less than 0.005 are further filtered based on their annotation and their segregation with the disease phenotype in the family.

Segregation status was checked for the 17 rare variants identified in the protein-coding genes in the WGS analysis (Table 2). Two variants: c.476G>C in *TMEM171* and c.1411-22A>G in *COL4A3BP*, co-segregated with the JME phenotype in the family. Conservation and in-silico predictions were performed for these two segregating variants (Table 3). The intronic variant c.1411-22A>G in *COL4A3BP* was predicted to alter the splice site. *COL4A3BP* codes for a ceramide transporter and mutations in this gene have been associated with intellectual disability (Fitzgerald et al., 2015; Murakami et al., 2020). Hence, this variant was tested for its effect on splicing by doing a mini-gene assay. It was observed that the variant does not give rise to altered splicing events (Manpreet Kaur, PhD Thesis, 2014).

The only rare variant that altered conserved amino acid was c.476G>C in *TMEM171* (Table 3). This variant gives rise to a non-synonymous change, p.Gly159Ala, in the protein. On performing multiple sequence alignment across species for TMEM171, the residue glycine at 159$^{th}$ position was found to be highly conserved (Figure 4). In-silico predictions showed that this variant could have deleterious effects on the protein's function.

**Table 2: Rare variants identified by whole-genome sequencing in the 5q12-q14 locus in the GLH5 family**

| Grch38[a] | Gene[b] | Sequence variant[c] | Effect on protein[d] | MAF GnomAD v2.1[e] | Segregation in family[f] |
|---|---|---|---|---|---|
| 62483049 | *IPO11* | c.829-52T>A | - | 0.0006 | Non-Segregating |
| 65554836 | *CENPK* | c.72T>A | p.Leu24Leu | - | Non-Segregating |
| 65554836 | *CENPK* | c.-19T>A | - | - | Non-Segregating |
| 65666110 | *TRAPPC13* | c.*1499A>G | - | - | Non-Segregating |
| 65666110 | *SGTB* | c.*4136T>C | - | - | Non-Segregating |
| 67182182 | *CD180* | c.*675G>A | - | - | Non-Segregating |
| 72360798 | *PTCD2* | c.*2371C>T | - | - | Non-Segregating |
| 72362653 | *PTCD2* | c.*4226A>G | - | 0.0004 | Non-Segregating |
| **73123849** | ***TMEM171*** | **c.476G>C** | **p.Gly159Ala** | **-** | **Segregating** |
| 75154293 | *ANKRD31* | c.1760A>G | p.Asp587Gly | - | Non-Segregating |
| 75337269 | *HMGCR* | c.-51del | - | - | Non-Segregating |
| **75382099** | ***COL4A3BP*** | **c.1411-22A>G** | **-** | **0.0006** | **Segregating** |
| 76618250 | *F2RL2* | c.457C>G | p.His153Asp | - | Non-Segregating |
| 79375001 | *HOMER1* | c.*1008G>A | - | 0.0002 | Non-Segregating |
| 79730596 | *CMYA5* | c.1831C>A | p.Gln611Lys | 0.0006 | Non-Segregating |
| 80960476 | *RASGRF2* | c.-263G>T | - | - | Non-Segregating |
| 83056822 | *TMEM167A* | c.*262C>T | - | 0.0004 | Non-Segregating |

[a]Genomic position of the nucleotide base on chromosome 5 (Chr 5) (GRCh38, NCBI). [b]Gene harbouring the variant, [c]Nomenclature of variants 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]Minor allele frequency in GnomAD version 2.1, [f]segregation status of the variant with the disease phenotype in the family GLH5.

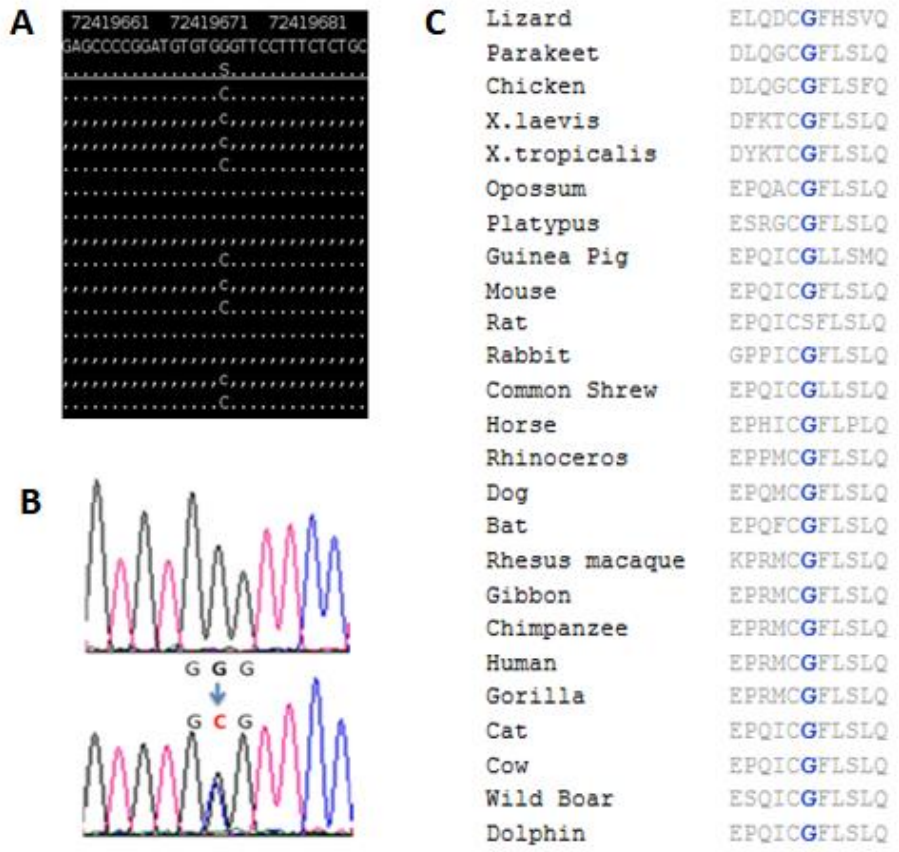



**Figure 4:The c.476G>C (p.Gly159Ala) variant in *TMEM171***: **A.** Alignment of the exome sequence reads against human genome reference sequence as viewed in tview of samtools v0.1.7a indicating the heterozygous G>C variant at g.72419676 in *TMEM171*. **B.** Electropherograms of sequence with wild-type (G) and variant allele (C) for the c.476G>C variant. The arrow indicates the nucleotide showing heterozygous variant shown by the presence of two peaks. **C.** Multi-species protein sequence alignment of TMEM171 by Clustal Omega. The glycine (G) 159 is highlighted in blue.

**Table 3: Functional predictions for the rare segregating variants identified**

| Grch38[a] | Gene[b] | Sequence variant[c] | Effect on protein[d] | Allele frequency in controls[e] | MAF gnomAD v2[f] | SIFT[g] | Polyphen2[h] | Mutation Taster[i] | Human Splicing Finder (HSF3.1)[j] |
|---|---|---|---|---|---|---|---|---|---|
| 73123849 | *TMEM171* | c.476G>C | p.Gly159Ala | 0/500 | - | Deleterious (0.01) | Probably damaging (0.91) | Deleterious (0.99) | - |
| 75382099 | *COL4A3BP* | c.1411-22A>G | - | 0/500 | 0.0006 | - | - | - | Potential alteration of splicing |

[a]Genomic position of the nucleotide base on chromosome 5 (chr 5) (GRCh38, NCBI). [b]Geneharbouring the variant, [c]Nomenclature of variants 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]Minor allele frequency of these variants in in-house, ethnically matched control individuals, [f]Minor allele frequency in GnomAD version 2.1, [g,h.i.j]predictions for pathogenicity of the variants.

### 4.2.3 *TMEM171* rare variants in JME patients

*TMEM171* gene encompasses 11.2 kilobases of the genomic region and harbours 4 exons. TMEM171 encodes a 324 amino acid transmembrane protein 171. The complete transcript of *TMEM171* was examined in additional 72 JME patients (480 patients, earlier work in the lab, Manpreet Kaur, PhD, Thesis 2014) (Table 4). In this analysis, two variants, c.269A>G (p.Gln90Arg) and c.466C>T (p.Arg156Trp) were identified which give rise to non-synonymous changes in TMEM171 (Table 5). Of the total three non-synonymous variants identified in TMEM171, p.Gln90Arg lies in the cytoplasmic arm and p.Arg156Trp and p.Gly159Ala lie in the non-cytoplasmic arm (Figure 5). The patients harbouring the TMEM171 non-synonymous variants all manifest GTCS and myoclonic seizures, which are hallmarks of JME (Table 6).

### 4.2.4 *TMEM171* in Epi25

Epi25 is a global collaboration under which more than 9,000 epilepsy patients have been sequenced (https://epi25.broadinstitute.org/). In this, ninety ultra-rare variants have been identified in *TMEM171* out of which twenty-three are pathogenic and identified only in the epileptic individuals examined. Sixteen of these variants give rise to missense changes in the protein sequence, four are loss of function variants and three lie in the splice regions (Appendix A 4.4). Figure 5 shows positions of four of the rare, pathogenic variants p.Ala62Glu, p.Leu125Pro, p.Asn139ThrfsTer7 and p.Glu299Ter.

**Table 4: Novel and rare *TMEM171* variants identified by Sanger sequencing a cohort of 552 JME patients.**

| Grch38 | Nucleotide change | Het/Hom | rs ID | Amino acid Change | Allele count in patient cohort (72 samples) | Allele count in patient cohort (552 samples) | MAF in control cohort (240 samples) | MAF gnomAD v2 |
|---|---|---|---|---|---|---|---|---|
| 73123440 | c.67T>C | het/hom | rs638333 | p.Phe23Leu | 8/144 | 122/1104 | 0.07 | 0.2 |
| 73123583 | c.210G>A | hom | rs7731777 | p.Gly70Gly | 144/144 | 1086/1104 | 1 | 1 |
| 73123629 | c.256C>G | het/hom | rs637450 | p.Arg86Gly | 51/144 | 452/1104 | 0.45 | 0.3 |
| **73123642** | **c.269A>G** | **het** | **rs543295611** | **p.Gln90Arg** | **1/144** | **3/1104** | **0** | **0.00005** |
| 73123790 | c.417C>A | het/hom | rs636926 | p.Asn139Lys | 37/144 | 394/1104 | 0.43 | 0.3 |
| **73123839** | **c.466C>T** | **het** | **TMP_ESP_5_72419666** | **p.Arg156Trp** | **1/144** | **2/1104** | **0.004** | **0.0001** |
| 73128458 | c.709G>A | het | rs186159238 | p.Ala237Thr | 1/144 | 1/1104 | 0.006 | 0.0008 |
| 73128490 | c.741G>A | het | rs150115532 | p.Pro247Pro | 1/144 | 3/1104 | 0 | 0.4 |
| 73131577 | c.819C>T | het/hom | rs61746311 | p.Asp274Asp | 5/144 | 30/1104 | 0.05 | 0.002 |
| 73131577 | c.819C>G | het/hom | rs61746311 | p.Asp274Glu | 7/144 | 52/1104 | 0.02 | 0.06 |

480 JME patients (Manpreet Kaur, PhD Thesis, 2014); 72 additional JME patients (this study)

**Table 5: Novel and rare missense patient variants identified in *TMEM171***

| Genomic position (Grch38) | cDNA position | amino acid position | rs ID | MAF GnomAD v2 | SIFT | PolyPhen | CADD phred | Mutation Taster | PROVEAN | Condel | Reference |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 73123558* | c.185C>A | p.Ala62Glu | – | – | Deleterious (0.01) | possibly damaging (0.92) | 23 | Deleterious (0.99) | Deleterious (0.73) | Deleterious (0.83) | Epi25 |
| 73123642 | c.269A>G | p.Gln90Arg | rs543295611 | 0.00005 | Tolerated (0.39) | Benign (0) | 0.89 | Neutral (1) | Neutral (0.12) | Neutral (0.01) | This study |
| 73123747* | c.374T>C | p.Leu125Pro | – | – | Deleterious (0.01) | probably damaging (0.89) | 26 | Deleterious (0.99) | Deleterious (0.89) | Deleterious (0.76) | Epi25 |
| 73123785* | c.414delC | p.Asn139ThrfsTer7 | rs756390663 | 0.00001 | Deleterious (0.01) | – | – | Deleterious (0.99) | Deleterious (0.90) | Deleterious (0.87) | Epi25 |
| 73123839 | c.466C>T | p.Arg156Trp | TMP_ESP_5_72419666 | 0.0001 | Deleterious (0.00) | Benign (0.43) | 22.4 | Neutral (1) | Deleterious (0.70) | Deleterious (0.58) | This study |
| 73123849 | c.476G>C | p.Gly159Ala | rs150702126 | 0.00009 | Deleterious (0.01) | Probably damaging (0.91) | 25.9 | Deleterious (0.99) | Deleterious (0.81) | Deleterious (0.77) | This study |
| 73131650* | c.895G>T | p.Glu299Ter | rs1261605666 | 0.00003 | Deleterious (0.00) | – | 42 | Deleterious (0.99) | Deleterious (0.87) | Deleterious (0.99) | Epi25 |

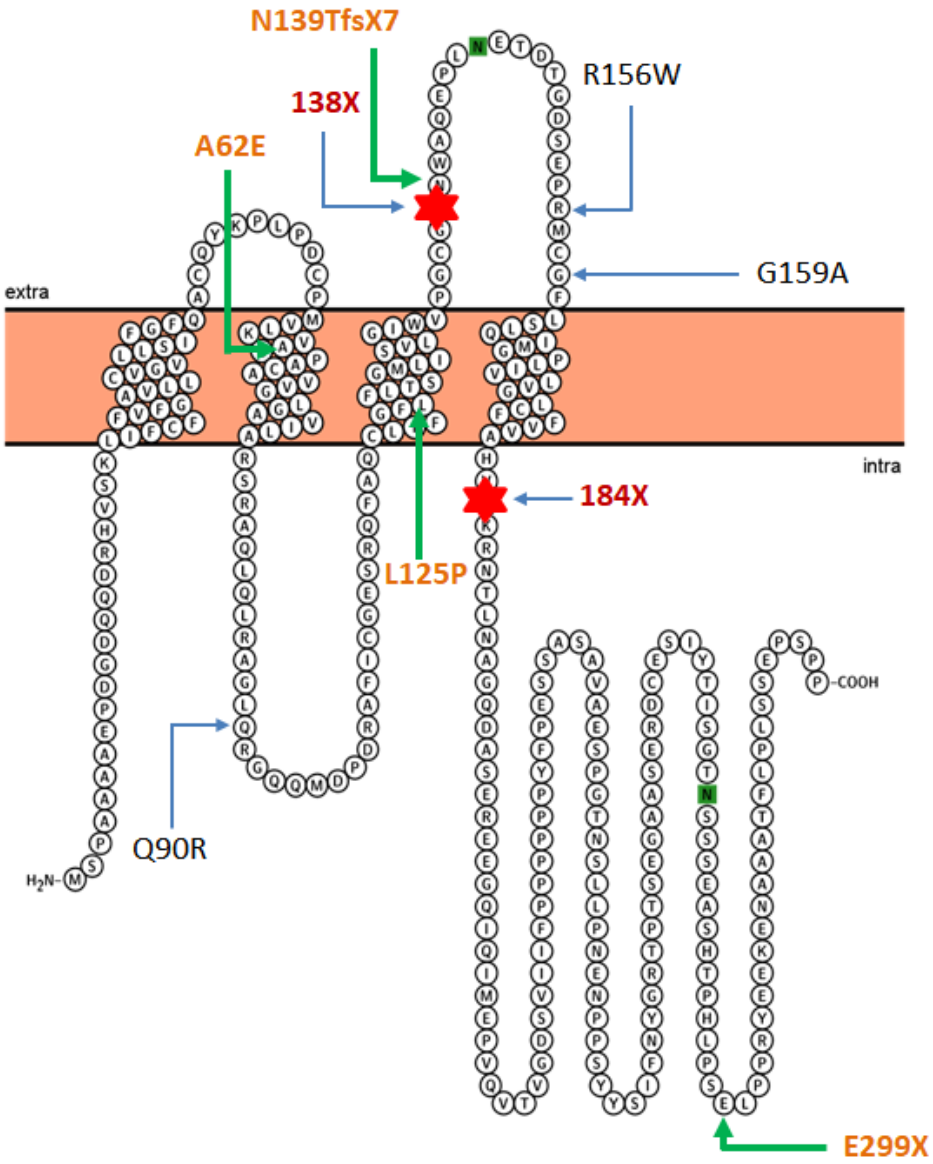



**Figure 5: TMEM171 protein architecture and mutations identified:** Schematic of TMEM171 protein prepared using the in silico human surfaceome (http://wlab.ethz.ch/surfaceome/). The N-terminal region of the protein harbours four transmembrane domains. Amino acid 1 to 20: cytoplasmic, amino acid 21 to 45: transmembrane, amino acid 46 to 56: non-cytoplasmic, amino acid 57 to 77: transmembrane, amino acid 78 to 108: cytoplasmic, amino acid 109 to 133: transmembrane, amino acid 134 to 159: non-cytoplasmic, amino acid 160 to 181: transmembrane, amino acid 182 to 324: cytoplasmic. The locations of the patient variants have been highlighted. p.Gln90Arg lies in the cytoplasmic region, p.Arg156Trp and p.Gly159Ala lie in the non-cytoplasmic region. Stop codons were inserted by site-directed mutagenesis at 138$^{th}$ and 184$^{th}$ position (shown in red) to generate truncated protein. p.Ala62Glu, p.Leu125Pro, p.Asn139ThrfsTer7 and p.Glu299Ter are variants identified in epileptic patients in the Epi25K study.

### 4.2.5 TMEM171 brain expression

RT-PCR analysis was performed to detect the expression of *TMEM171* in human brain sub-regions, using Marathon-Ready™ cDNAs (Clontech) from hippocampus, cerebral cortex, hypothalamus and cerebellum. The intron spanning primer pair (forward primer binds in the first exon and the reverse primer in the second exon) specific to *TMEM171* yielded a 400 bp PCR product from the brain sub-regions analyzed and in the positive control [*TMEM171* pcDNA3.1 (+)] (Figure 6). Sanger sequencing of the amplified products validated the sequence (NM_173490.7, CCDS4017.1). The amplified product from cDNA of hippocampus, cerebral cortex, hypothalamus and cerebellum indicates the expression of *TMEM171* at the transcript level in these human brain regions. TMEM171 protein expression analysis in human brain regions, cortex, hippocampus and cerebellum was performed by Western blot analysis. We observed a band, immunoreactive to anti-TMEM171 antibody, at 48 kDa (Figure 6).

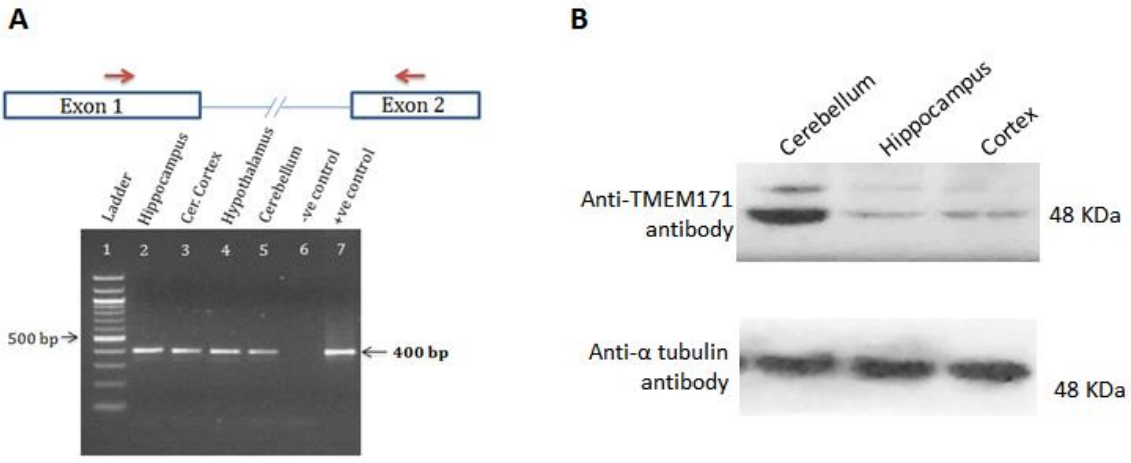



**Figure 6: Expression of *TMEM171* in the human brain regions: A**. RT-PCR analysis of cDNA from different brain regions (hippocampus, cerebral cortex, hypothalamus, and cerebellum). The arrow (lower panel) indicates the 400 bp product amplified by cDNA specific intron spanning primers located in exon 1 and exon 2 (marked by arrows in the upper panel) of *TMEM171*. Lane 6 shows the negative control (no DNA template) and lane 7 indicates the 400 bp product from the positive control [wild-type TMEM171 pcDNA3.1 (+) construct]. Lane 1 is the

100 bp ladder (Manpreet Kaur, PhD Thesis, 2014) **B.** Expression of *TMEM171* in various human brain regions (cerebellum, hippocampus, cortex) shown by Western analysis. Anti-TMEM171 antibody shows immunoreactivity at approximately 48KDa, which is the theoretical molecular weight of TMEM171 protein.

### 4.2.6 TMEM171 cellular localization

To examine the cellular localization of TMEM171 and to test whether the variants p.Gln90Arg, p.Arg156Trp and p.Gly159Ala, affect its localization, fluorescence-immunocytochemistry was carried out in cultured HEK293T cells transiently transfected with the wild-type and variant *TMEM171* constructs. Immunoreactivity was observed using mouse polyclonal anti-TMEM171 antibody (Sigma). TMEM171 protein expression was observed in the cytoplasm and on the plasma membrane of the cell (Figure 7). HEK293T transfected with *TMEM171* constructs harbouring the mutations p.Gln90Arg, p.Arg156Trp, p.Gly159Ala exhibited a similar expression profile as seen for the wild-type protein (Figure 7). This suggests that these missense mutations do not seem to alter the plasma membrane localization of TMEM171.

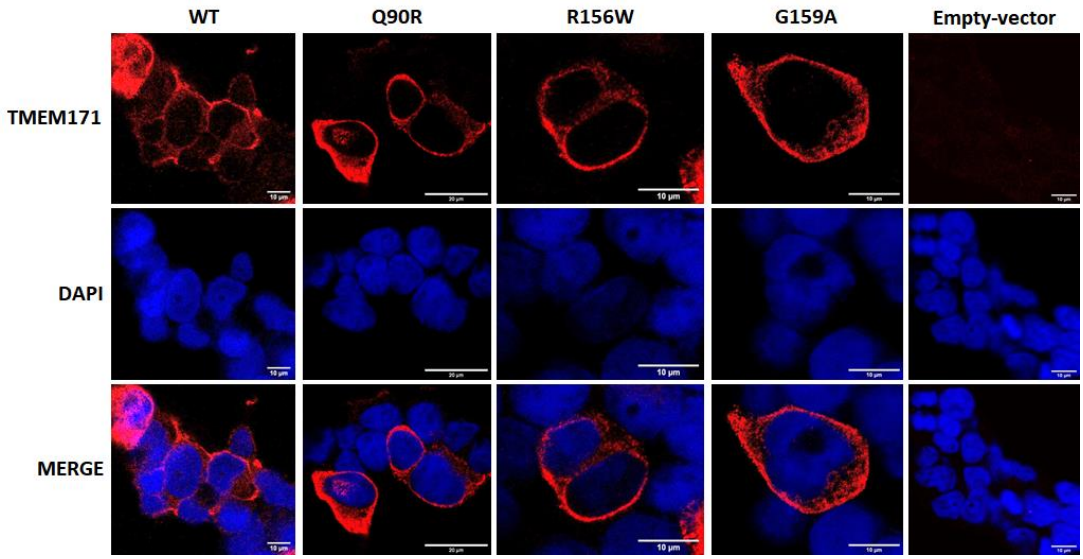



**Figure 7: Cellular localization of wildtype (WT) and mutant TMEM171 proteins.** Fluorescence immunocytochemistry performed on HEK293T transiently transfected with the wild-type and variant, p.Gln90Arg, p.Arg156Trp, p.Gly159Ala, cDNAs. Localization of TMEM171 was observed with anti-TMEM171

antibody. TMEM171 images shown here are obtained at 63x zoom with the Zeiss 880 microscope.

### 4.2.7 TMEM171-interacting partners

Immuno-pulldown (IP) experiments were conducted to get an initial insight into the molecular pathways, TMEM171 might be relevant to. The first protein checked for interaction with TMEM171 was SNAPIN. SNAPIN is predicted to be an interactor for TMEM171 (Biogrid, https://thebiogrid.org/126392/summary/homo-sapiens/tmem171.html) and its deficiency in mice is associated with developmental defects of the central nervous system (Zhou et al., 2010). Another candidate chosen was SNAP25, which is an interactor of PRRT2 (Li et al., 2015), a proline-rich transmembrane domain-containing protein known to cause benign familial infantile epilepsy (Heron et al., 2012). SNAP25 has been associated with several human neurological syndromes, including attention-deficit/hyperactivity disorder (ADHD), schizophrenia and bipolar disorder (Etain et al., 2010; Hawi et al., 2013; Antonucci et al., 2016).

FLAG-tagged full-length, 138X and 184X TMEM171 was overexpressed in HEK293 and IP was performed using anti-FLAG antibody. SNAPIN and SNAP25 were found in the elute fraction along with TMEM171, showing that they are present in the same complex of proteins (Figure 8). Truncated TMEM171 showed disruption of interaction with both SNAP25 and SNAPIN. SNAPIN showed complete disruption of interaction with the truncated TMEM171. SNAP25 showed reduced interaction with the truncated TMEM171 protein, as shown in figure 8-C. For IP with the patient variant harbouring TMEM171 proteins, p.Gln90Arg, p.Arg156Trp and p.Gly159Ala, no significant disruption, either increase or reduction, in the amount of SNAP25 (Figure 9-B) and SNAPIN (Figure 9-C) was observed.

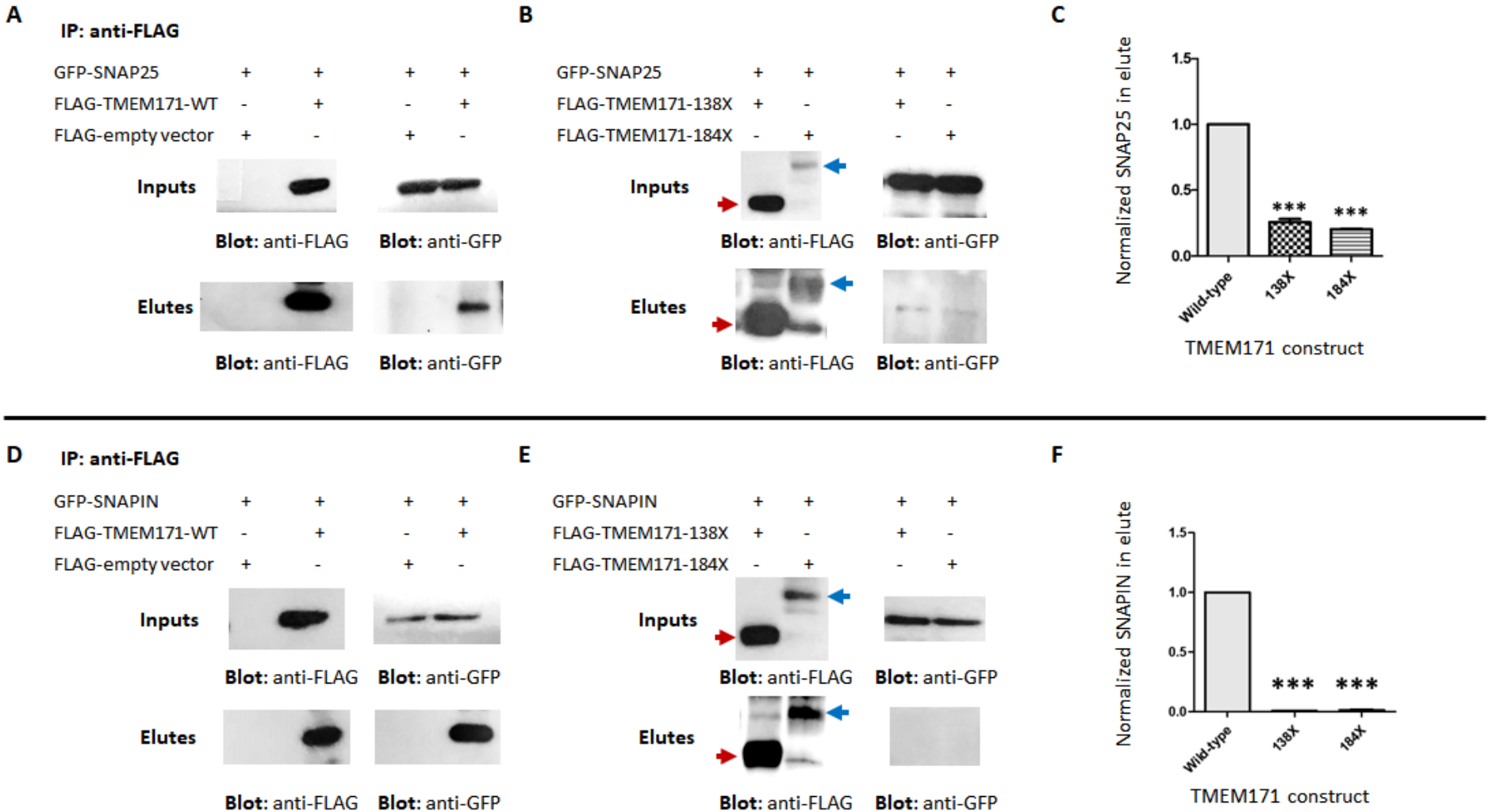

A
IP: anti-FLAG
GFP-SNAP25 + + + +
FLAG-TMEM171-WT - + - +
FLAG-empty vector + - + -
Inputs
Blot: anti-FLAG
Blot: anti-GFP
Elutes
Blot: anti-FLAG
Blot: anti-GFP
B
GFP-SNAP25 + + + +
FLAG-TMEM171-138X + - + -
FLAG-TMEM171-184X - + - +
Inputs
Blot: anti-FLAG
Blot: anti-GFP
Elutes
Blot: anti-FLAG
Blot: anti-GFP
C
Normalized SNAP25 in elute
1.5
1.0
0.5
0.0
***
***
Wild-type
138X
184X
TMEM171 construct
D
IP: anti-FLAG
GFP-SNAPIN + + + +
FLAG-TMEM171-WT - + - +
FLAG-empty vector + - + -
Inputs
Blot: anti-FLAG
Blot: anti-GFP
Elutes
Blot: anti-FLAG
Blot: anti-GFP
E
GFP-SNAPIN + + + +
FLAG-TMEM171-138X + - + -
FLAG-TMEM171-184X - + - +
Inputs
Blot: anti-FLAG
Blot: anti-GFP
Elutes
Blot: anti-FLAG
Blot: anti-GFP
F
Normalized SNAPIN in elute
1.5
1.0
0.5
0.0
***
***
Wild-type
138X
184X
TMEM171 construct

**Figure 8: Immuno-pulldown of TMEM171 wildtype and truncated proteins: A, B.** IP of FLAG-tagged TMEM171 (wildtype, 138X or 184X) or 3X-FLAG empty vector co-expressed with SNAP25. IP is performed with anti-FLAG antibody. **C.** Normalized values of GFP-SNAP25 present in the elute fractions have been plotted for wild-type, 138X and 184X TMEM171 protein IPs. The experiment was done three times and band intensities were measured using ImageJ. **D, E.** IP of FLAG-tagged TMEM171 (wildtype, 138X or 184X) or 3X-FLAG empty vector co-expressed with SNAPIN. IP is performed with anti-FLAGantibody. **F.** Normalized values of GFP-SNAPIN present in the elute fractions have been plotted for wild-type, 138X and 184X TMEM171 protein IPs. The experiment was done three times and band intensities were measured using ImageJ. Red arrow: band corresponding to TMEM171-138X truncated protein; Blue arrow: band corresponding to TMEM171-184X truncated protein.

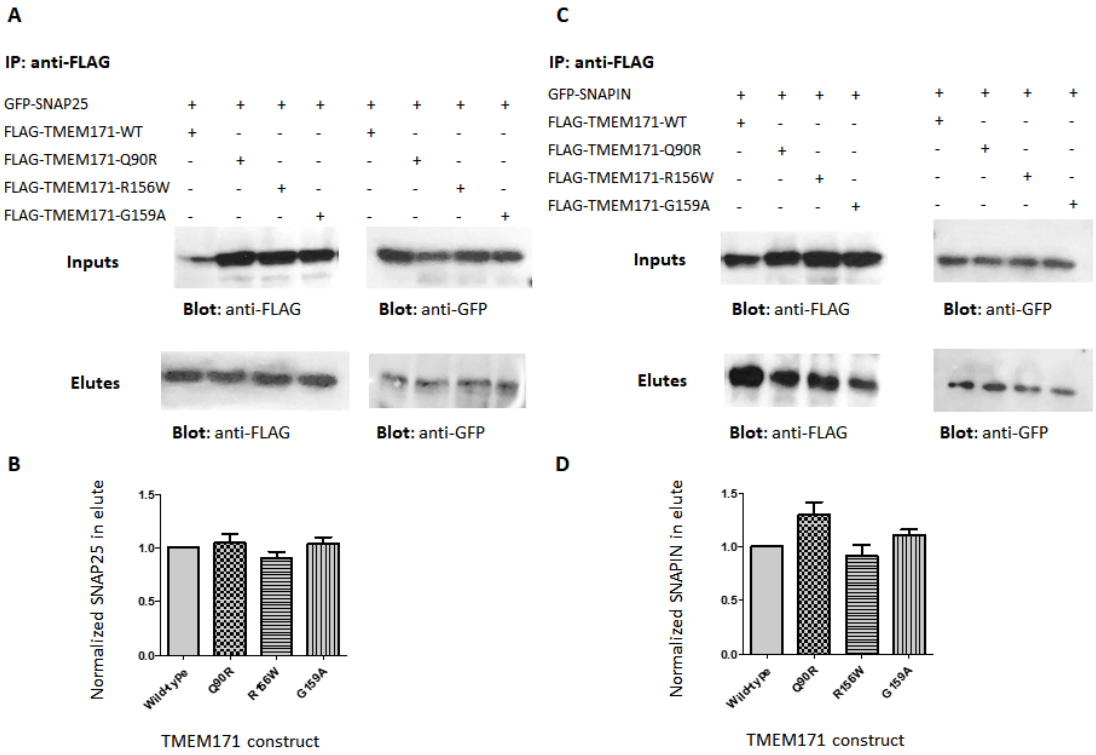



**Figure 9: Immunopulldown of TMEM171 wildtype and human variant harbouring proteins: A.** IP of FLAG-tagged TMEM171 (wildtype, p.Gln90Arg, p.Arg156Trp or p.Gly159Ala) co-expressed with SNAP25. IP is performed with anti-FLAG antibody. **B.** Normalized values of GFP-SNAP25 present in the elute fractions have been plotted for wild-type, p.Gln90Arg, p.Arg156Trp and p.Gly159Ala TMEM171 protein IPs. The experiment was done three times and band intensities were measured using ImageJ. **C.** IP of FLAG-tagged TMEM171 (wildtype, p.Gln90Arg, p.Arg156Trp or p.Gly159Ala) or 3X-FLAG empty vector co-expressed with SNAPIN. IP is performed with anti-FLAG antibody. **D.** Normalized values of GFP-SNAPIN present in the elute fractions have been plotted for wild-type, p.Gln90Arg, p.Arg156Trp and p.Gly159Ala TMEM171 protein IPs. The experiment was done three times and the band intensities were measured using ImageJ.

### 4.2.8 Co-expression of TMEM171 wildtype and variant proteins with its interacting partners, SNAP25 and SNAPIN

FLAG-tagged wildtype, patient-variant harbouring and truncated TMEM171 proteins were co-expressed with GFP-tagged SNAP25 and SNAPIN. GFP-SNAP25 showed localization at the plasma membrane. FLAG-tagged TMEM171 wildtype protein, visualized by staining with anti-FLAG antibody, was observed to localize to the cytoplasm as well the plasma membrane. TMEM171 and SNAP25 showed a high degree of colocalization at the plasma membrane (Figure 10-A Pearson's correlation coefficient (PCC) of 0.80). Patient variants harbouring TMEM171 protein also showed co-localization with SNAP25 at the plasma membrane (Figure 11). TMEM171 truncated protein 138X, which has the first 3 transmembrane domains, showed a diffused localization throughout the nucleus and cytoplasm and reduction in its localization to the plasma membrane. The co-localization of SNAP25 is also reduced with the truncated TMEM171 protein 138X at the plasma membrane, as seen in figure 11 and quantified by measuring PCC. However, 184X, which has all four transmembrane domains of TEM171 but lacks the cytoplasmic C-terminal tail, shows localization almost similar to the wildtype protein and co-localizes with SNAP25 at the plasma membrane (Figure 11). GFP-SNAPIN localizes to both the nucleus and cytoplasm and does not show a distinct localization at the plasma membrane (Figure 10-B). Co-expression of TEME171 with SNAPIN shows that the two proteins are expressed in the cytoplasm and colocalize to an extent, as shown by the 2-D scatter plot and line profile generated using ImageJ (Figure 10-B; PCC of 0.59 suggests partial colocalization). TMEM171 truncated protein 138X, which localizes more to the cytoplasm showed slightly enhanced localization with SNAPIN, as seen in the PCC graph in figure 12.

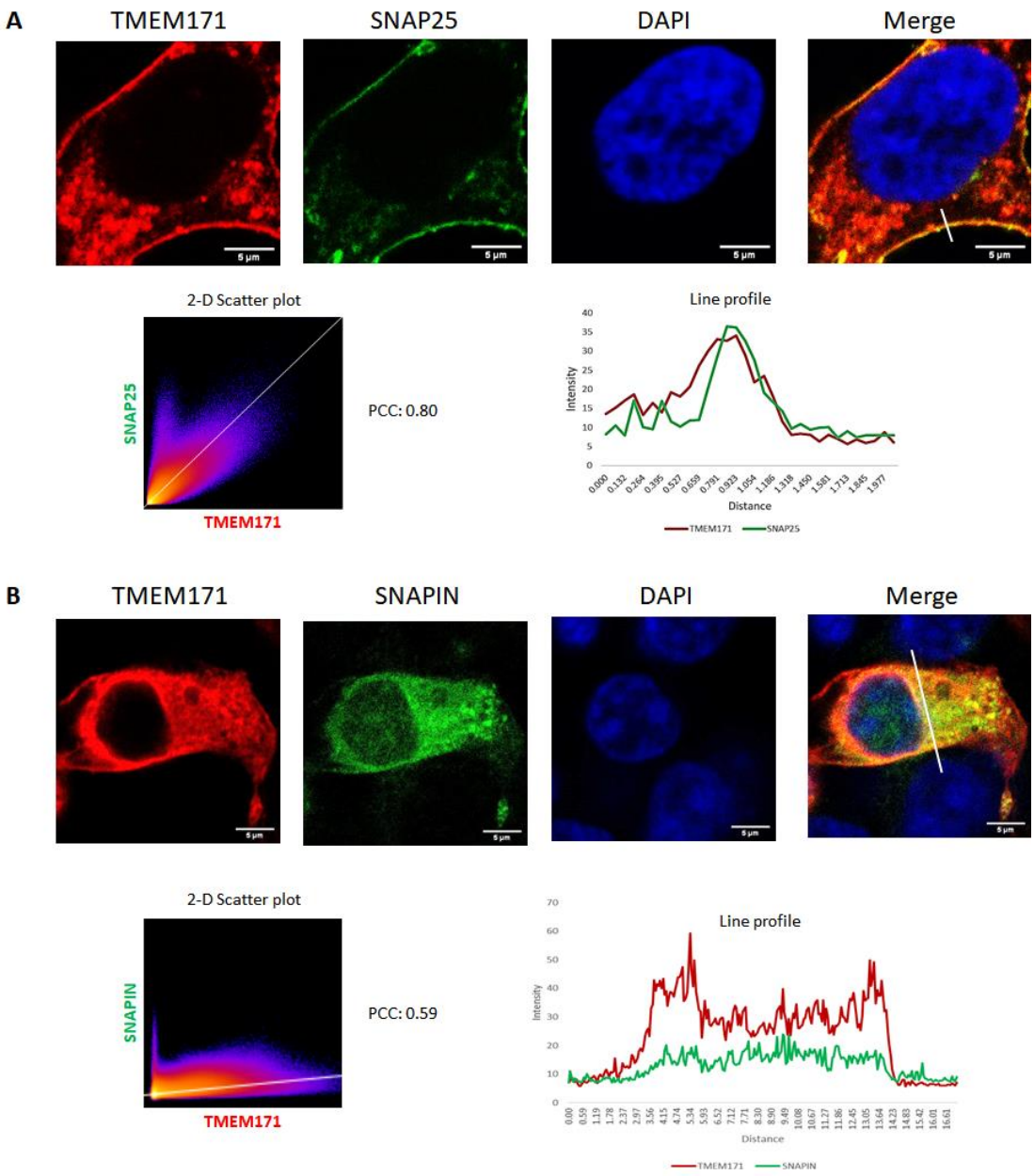



**Figure 10: Co-expression of TMEM171 with SNAP25 and SNAPIN in HEK293: A.** FLAG-tagged TMEM171 was co-expressed with GFP tagged SNAP25. Red – TMEM171, Green – SNAP25, Blue – DAPI. 2-D scatter plot and line profile have been shown for TMEM171 and SNAP25 **B.** FLAG-tagged TMEM171 co-expressed with GFP tagged SNAPIN. 2-D scatter plot and line profile have been shown for TMEM171 and SNAP25. Red – TMEM171, Green – SNAP25, Blue – DAPI. Imaging was done at 63X magnification. PCC: Pearson's correlation coefficient.

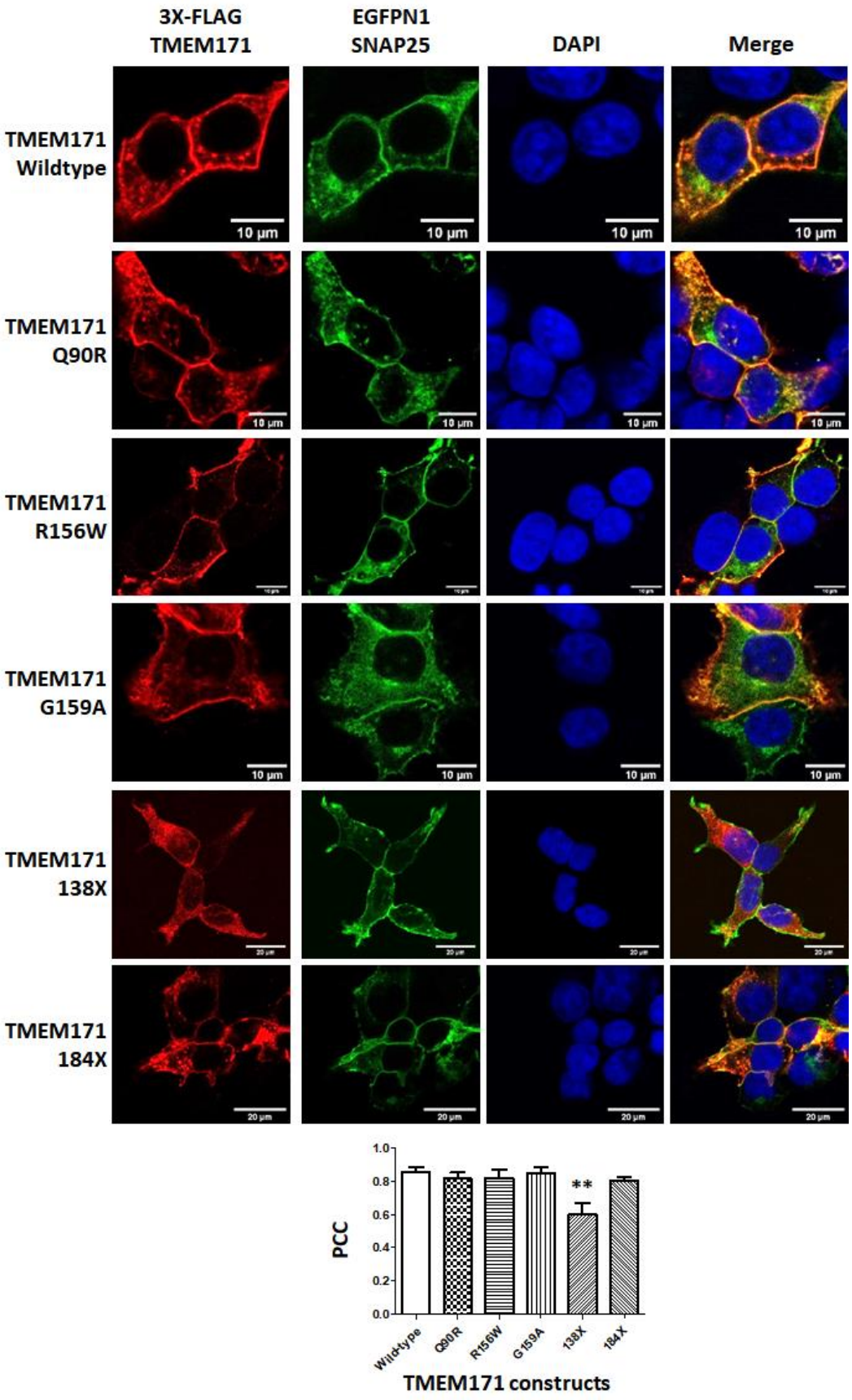



**Figure 11: Co-expression of TMEM171 and SNAP25 proteins in HEK293: A.** FLAG-tagged TMEM171 (wildtype, p.Gln90Arg, p.Arg156Trp, p.Gly159Ala, 138X

or 184X) co-expressed with GFP-tagged SNAP25. Red – TMEM171, Green – SNAP25, Blue – DAPI. Imaging was done at 63X magnification. PCC was calculated from three independent experiments and 30 cells. One-way ANOVA followed by Dunnet's multiple comparison test was used to calculate the statistical significance.

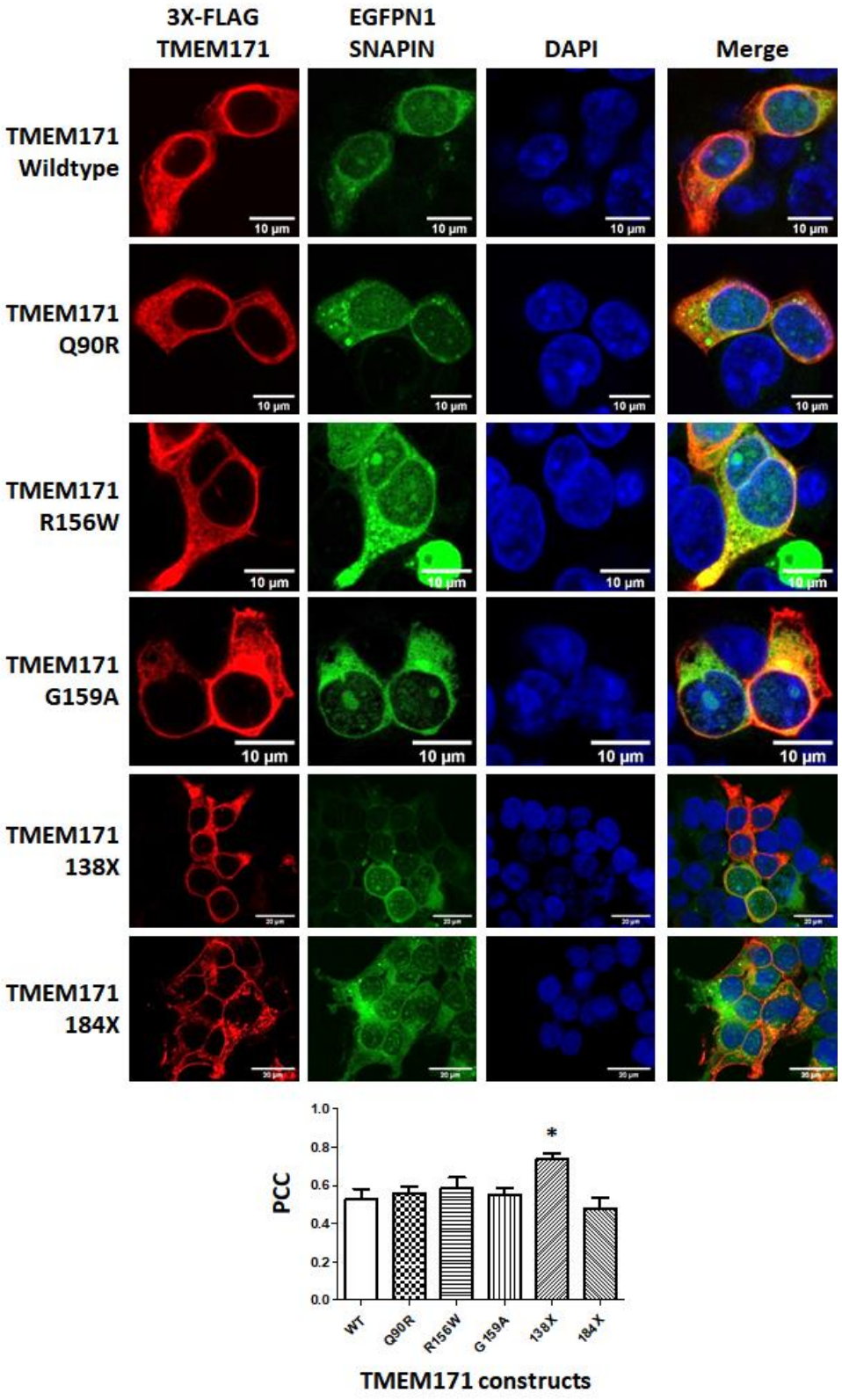


**Figure 12: Co-expression of TMEM171 and SNAPIN proteins in HEK293: A.** FLAG-tagged TMEM171 (wildtype, p.Gln90Arg, p.Arg156Trp, p.Gly159Ala, 138X or 184X) co-expressed with GFP-tagged SNAPIN. Red – TMEM171, Green – SNAP25, Blue – DAPI. Imaging was done at 63X magnification. PCC from three independent experiments and 30 cells were calculated. One-way ANOVA followed by Dunnet's multiple comparison test was used to calculate the statistical significance.

## 4.3 Discussion

In this chapter, findings of a whole-genome sequencing study, to identify potentially causative gene at EJM4 (Kapoor et al., 2007) and the first line of functional characterization of the gene, have been presented. The WGS analysis of data from an affected individual in the family, II:7, led to the identification of a rare, possibly pathogenic variant in *TMEM171*, which was considered suitable for further studies.

The missense variant in *TMEM171* substitutes an evolutionarily conserved amino acid residue, glycine at position 159 to alanine; and is predicted to have a deleterious effect on the protein structure/ function. This heterozygous variant was found in all the affected individuals of the family and was absent in our set of control individuals and public databases such as dbSNP and gnomADv2.

TMEM171 encodes a 324 amino acid proline-rich transmembrane protein, also known as PRP2 (Proline-rich protein 2). The structure and cellular functions of TMEM171 protein are not known so far. It is predicted to have four transmembrane domains in the N-terminal region and a long C-terminal tail in the cytoplasm (Figure 5). Immuno-fluorescence studies done in HEK293 show that TMEM171 localizes to the plasma membrane (Figure 7, 10, 11). As a primary step in examining the role of *TMEM171* in a brain disorder, we identified its expression at the transcript level and protein level from different human brain regions (Figure 6). So far, no study has either characterized the biological function of TMEM171 or reported its role in a brain-related disorder.

Further, while exploring the role of *TMEM171* as an epilepsy-linked gene, Manpreet, a former graduate student in the lab had identified two additional JME-associated novel/rare variants, p.Gln90Arg and p.Arg156Trp in a set of 480 JME patients. On sequencing the gene in additional 72 JME patients, I found the same missense variants, p.Gln90Arg and p.Arg156Trp in two patients. Taken together, the three mutations identified in this study alter conserved amino acid residues in TMEM171 and two of the three variants, p.Arg156Trp and p.Gly159Ala, are predicted to have a damaging effect on the protein's function. While these missense variants do not affect the membrane targeting or expression of the protein, these may affect its biological functions, which are yet to be characterized. In addition to the variants identified in our study, four rare, pathogenic variants p.Ala62Glu, p.Leu125Pro, p.Asn139ThrfsTer7 and p.Glu299Ter have been identified in the Epi25 consortium study. This finding supports our observations regarding the role of *TMEM171* as an epilepsy-causing gene.

Interestingly, mutations in the proteins belonging to the group of uncharacterized transmembrane proteins, such as *PRRT2* which encodes a proline-rich transmembrane domain protein 2 of an unknown function, have been implicated in neurological disorders. Mutations in *PRRT2* have been identified for disease phenotypes such as paroxysmal kinesigenic dyskinesia (PKD) (Chen et al., 2011), paroxysmal kinesigenic dyskinesia with infantile convulsions (PKD/IC) (Lee et al., 2012), benign familial infantile epilepsy (BFIE) and infantile convulsions and choreoathetosis (ICCA) (Heron et al., 2012). Mutations in other TMEM proteins have also been found such as in *TMEM216* (Valente et al., 2010) and *TMEM237* (Huang et al., 2011) for the Joubert syndrome-related disorders; and in *TMEM240* for spinocerebellar ataxia with mental retardation and severe cognitive impairment (Delplanque et al., 2014; Homa et al., 2020). Ultra-rare deleterious variants have been identified in epilepsy patients in TMEM171 as well as additional TMEM proteins such as *TMEM62, TMEM59L, TMEM187, TMEM220,* and

*TMEM255A* (Epi25 collaborative). Certain variants identified in TMEM171 lie in highly conserved transmembrane regions of the protein and are predicted to disrupt the protein's function. This goes to show that the transmembrane group of proteins may play an important role in normal brain functioning.

To get an insight into the molecular pathways TMEM171 might be involved in, immuno-pulldown (IP) studies were conducted which led to the identification of two potential interactors, SNARE-associated protein (SNAPIN) and synaptosomal-associated protein-25 (SNAP25). SNAPIN is predicted to be an interactor for TMEM171 (Biogrid) and its deficiency in mice is associated with developmental defects of the central nervous system (Zhou et al., 2010). Another candidate tested was SNAP25, which is an interactor of PRRT2 (Li et al., 2015), a proline-rich transmembrane domain-containing protein known to cause benign familial infantile epilepsy (Heron et al., 2012). SNAP25 has been associated with several human neurological syndromes, including attention-deficit/hyperactivity disorder (ADHD) (Hawi et al., 2013), schizophrenia and bipolar disorder (Etain et al., 2010). *SNAP25* mutations in human patients with epilepsy and intellectual disability have also been identified (Rohena et al., 2013). Truncated TMEM171 showed disruption in its interaction with both SNAP25 and SNAPIN.

SNAP25, a component of the SNARE protein complex, regulates neurotransmitter release during synaptic transmission. Apart from exocytosis, it also regulates various voltage-gated calcium channels (VGCCs) (Catterall and Few, 2008) and negatively controls neuronal calcium responsiveness through voltage-gated calcium channel inhibition(Pozzi et al., 2008). SNAP25 not only functions at the pre-synapse to control exocytic and endocytic processes but also regulates postsynaptic receptor trafficking (Selak et al., 2009), spine morphogenesis (Fossati et al., 2015) and plasticity (Lau et al., 2010). SNAPIN is a coiled-coil-forming protein that associates with the SNARE (soluble N-ethylmaleimide-sensitive fusion protein attachment protein

receptor) complex of proteins and the BLOC-1 (biogenesis of lysosome-related organelles) complex. As part of the SNARE complex, it is required for vesicle docking and fusion and regulates neurotransmitter release (Pan et al., 2009). The BLOC-1 complex is required for the biogenesis of specialized organelles such as melanosomes and platelet dense granules in non-neuronal cells. In the brain, BLOC-1 is capable of interacting with a subset of SNARE (soluble *N*-ethylmaleimide-sensitive factor attachment protein receptor) proteins previously implicated in neurite outgrowth, and that primary hippocampal neurons deficient in BLOC-1 display neurite extension defects (Ghiani et al., 2010) (Figure 13).

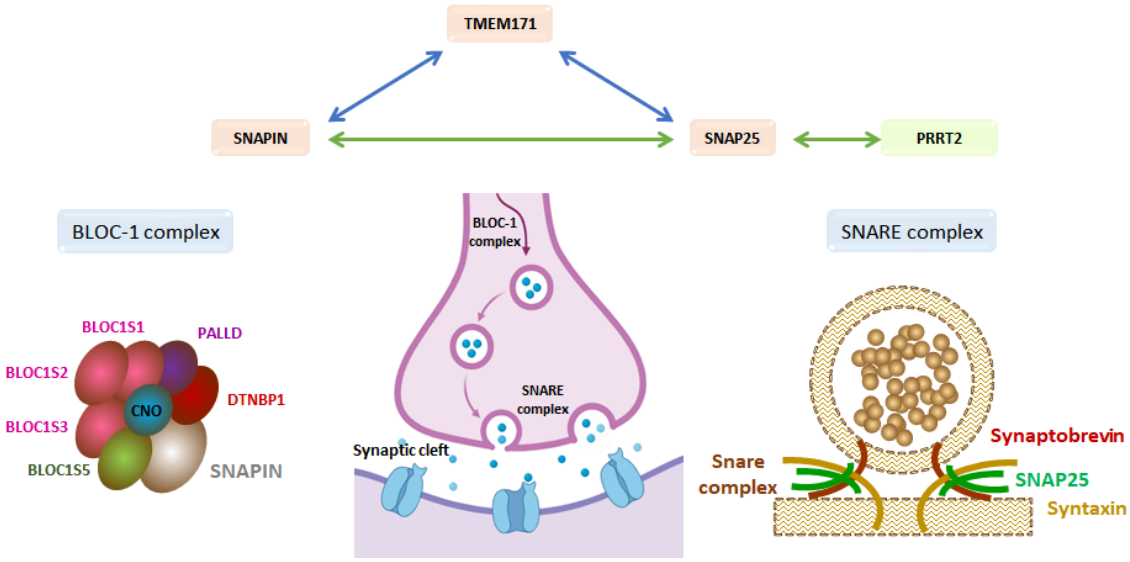



**Figure 13: TMEM171, its interactors and probable pathways:** TMEM171 interacts with SNAPIN and SNAP25, and SNAPIN and SNAP25 have been shown to interact with each other (Ilardi et al., 1999). SNAP25 is an interactor of PRRT2, another epilepsy causing gene. SNAPIN is a member of the BLOC-1 complex and is associated with some SNARE complex proteins. SNAP25 is a member of the SNARE complex. TMEM171, through its association with these two complexes, may regulate normal neuronal functioning.

Based on the findings discussed in this chapter, a strong suggestion is made for *TMEM171* as the underlying genetic cause for epilepsy at the EJM4 (5q12-q14) locus. We report three missense variants in *TMEM171* in this study, p.Gly159Ala in the family; and p.Gln90Arg and

p.Arg156Trp in three additional unrelated JME patients. These novel/rare epilepsy-linked variants in *TMEM171* provide genetic evidence for its involvement in the causation of IGE/JME. The interaction of TMEM171 with SNAPIN and SNAP25 provides additional suggestions for its functions in neurons. Further studies using knockout cell lines or mouse models or both, shall help understand molecular and cellular roles of TMEM171 in greater depth.

## Conclusions

Epilepsy is a common neurological disorder characterized by recurrent seizures usually unprovoked by any immediately identifiable cause. Genetic generalized epilepsies (GGE) constitute a heterogeneous group of epilepsies with generalized seizures. Juvenile myoclonic epilepsy (JME) is one of the most common GGE with a prevalence of 18% of all GGEs. To date, only a few genes for JME have been mapped and identified in large families. In the present study, I aimed to identify potentially disease-causing genes at the three genomic loci, 5p15-q12, 2q33-q36 (EJM9) and 5q12–q14 (EJM4), linked to juvenile myoclonic epilepsy in families from the southern parts of India. In this thesis, I have presented work done for (i) identification of potentially causative variants/genes in the disease-linked intervals by employing whole exome/genome sequencing experiments and (ii) to evaluate cell biological correlates of the variants in the genes proposed to be responsible for epilepsy.

In the initial search for genes causing epilepsies, ion channel genes emerged to be the leading cause. This led to an era of channelopathy for genetic epilepsies. In recent times, studies have led to the identification of other genes associated with genetic epilepsies. These genes are part of varied mechanisms which are involved in the pathophysiology of epilepsy. In our study too, we have identified non-channelopathy candidate causative genes, *CDC20B* at 5p15-q12, *DES* at 2q33-q36 and*TMEM171* at 5q12-q14, respectively.

The first gene that I have presented in this thesis is *CDC20B,* identified in a 3-generation south Indian family, GLH 35. Four additional rare, missense variants, p.Gln94His, p.Arg200Gly, p.Ile212Lys and p.Trp363Arg were identified among 552 JME patients examined for *CDC20B*. Interestingly, *CDC20B* is among the top 200 genes identified with the burden of ultra-rare variants in the Epi25 collaborative. CDC20B encodes a 591-amino acid long, WD repeat domain-containing

protein. During late telophase and cytokinesis, the protein localizes to the midbody, during other cell cycle stages, it is present in the cytoplasm. We found that CDC20B co-immunoprecipitates with gamma-tubulin (TUBG1), EF-hand domain-containing protein 1 (EFHC1), and polo-like kinase (PLK1). Cytokinesis defects were observed on over-expression of the CDC20B variants, p.Arg200Gly, p.Trp363Arg and p.Ala509Thr, suggesting a cell cycle associated role for the protein. Interestingly, two variants p.Trp363Arg and p.Ala509Thr expressed in the mouse embryos at E14.5 by in-utero electroporation, showed longer projections in the neuronal progenitors than the ones observed for the wildtype CDC20B allele. The projection lengths of the neuronal cells may influence micro-developmental aspects of neural circuits which in turn can predispose to epilepsy.

The next candidate gene, *DES*, was identified at the EJM9, 2q33-q36, JME locus. This locus was identified in a three-generation south Indian family (SCT135) with several of its members affected by the disorder. *DES* codes for a 470 amino acid, muscle-specific, type III intermediate filament. RT-PCR analysis showed the expression of desmin in different human brain regions. Desmin is known to be expressed in pericytes in the brain. Though the epilepsy-associated mutations in *DES* do not seem to affect its expression levels and filament formation, our observations, make a prima facie case that they may have a damaging effect on its cellular function/s in the brain by affecting the physiology of the pericytes.

Lastly, *TMEM171* was identified at the locus EJM4 (5q12-q14). TMEM171 encodes a proline-rich transmembrane protein with an unknown biological function. TMEM171 protein has four predicted transmembrane domains in the N-terminal region and a C-terminal tail in the cytoplasm. It localizes to the plasma membrane. The patient variant, p.Gln90Arg lies in the second cytoplasmic loop, whereas the other two variants, p.Arg156Trp and p.Gly159Ala, lie in the second external loop of the protein. RT-PCR and Western analysis suggests the

expression of TMEM171 in different human brain regions. Though the epilepsy-associated TMEM171 variants do not seem to affect its membrane localization or expression levels, they may have damaging effects on the functional/cellular physiology of the brain. To get an insight into TMEM171 protein's functions and pathways that it may be a part of, I carried out immuno-pulldown experiments and identified SNAP25 and SNAPIN as two TMEM171-interacting partners. SNAP25 is a member of the SNARE complex. SNAPIN associates with the SNARE complex proteins and the BLOC-1 complex. Further delineation of TMEM171's role in these pathways would require additional studies.

In conclusion, the candidate genes identified in this thesis provide insights into the various pathways that may be important in the pathophysiology of JME. The three genes identified, *CDC20B*, *DES* and *TMEM171*, all seem to regulate very different pathways in the cell ranging from cell cycle, ciliogenesis, blood-brain barrier maintenance through pericytes to synaptic regulation. These discoveries will not only help in developing therapeutic interventions for epilepsy patients but will also be useful in developing diagnostic gene panels for genetic epilepsies.

## Appendix I

### A 2.1 Rare gene variants identified at the 5p15-q12 locus (GLH 35 family)

| Gene name | Genomic Position (Grch 38) | Allele | Consequence | cDNA change | Amino acid change | rs ID |
|---|---|---|---|---|---|---|
| *C1QTNF3* | 34018448 | C/T | 3' UTR | c.*2135G>A | - | rs528976830 |
| *CDC20B* | 55114253 | C/T | missense | c.1513G>A | p.Ala505Thr | rs200898156 |
| *C1QTNF3* | 34018448 | C/T | 3' UTR | c.*2135G>A | - | rs528976830 |
| *NUP155* | 37314221 | T/C | missense | c.2236A>G | p.Ile746Val | rs376806446 |
| *DNAH5* | 13931214 | G/T | synonymous | c.88C>A | p.Arg30= | rs114220185 |
| *PDE4D* | 59586417 | C/T | 5' UTR | c.-328G>A | - | rs369106859 |
| *PDE4D* | 59586417 | C/T | 5' UTR | c.-334G>A | - | rs369106859 |
| *PDE4D* | 60004290 | G/A | 5' UTR | c.-5C>T | - | rs556348371 |
| *CCL28* | 43375949 | C/T | downstream | - | - | rs777025587 |
| *SLC9A3* | 473251 | C/T | 3' UTR | c.*128G>A | - | - |
| *TARS* | 33452960 | A/- | intron | c.330-317del | - | rs79630452 |
| *TARS* | 33453598 | G/A | intron | c.453+186G>A | - | rs553354497 |
| *CDC20B* | 55114253 | C/T | missense | c.1525G>A | p.Ala509Thr | rs200898156 |
| *CDC20B* | 55114253 | C/T | missense | c.1399G>A | p.Ala467Thr | rs200898156 |
| *SNX18* | 54543800 | G/A | 3' UTR | c.*620G>A | - | - |
| *SNX18* | 54544033 | -/T | 3' UTR | c.*854dup | - | - |
| *SNX18* | 54544468 | TTTAC/- | 3' UTR | c.*1286_*1290del | - | - |
| *SNX18* | 54545992 | TG/- | 3' UTR | c.*2812_*2813del | - | - |
| *SNX18* | 54546544 | A/G | 3' UTR | c.*3364A>G | - | rs186817959 |
| *11-Mar* | 16117375 | T/C | intron | c.694-26294A>G | - | rs1417277884 |
| *11-Mar* | 16149490 | CTA/- | intron | c.693+28236_693+28238del | - | - |
| *11-Mar* | 16156976 | A/G | intron | c.693+20750T>C | - | rs116824407 |
| *ANKH* | 14769792 | CC/- | upstream | - | - | rs1307759653 |
| *ANKH* | 14769796 | AA/- | upstream | - | - | rs1217430727 |
| *ANKH* | 14769799 | -/CGGCGAC | upstream | - | - | rs1255618085 |
| *ANKH* | 14769803 | TT/- | upstream | - | - | rs1337495788 |
| *ANKH* | 14769806 | T/- | upstream | - | - | rs1214856162 |
| *SLC9A3* | 473251 | C/T | missense | c.2383G>A | p.Ala795Thr | - |
| *PLEKHG4B* | 109094 | C/T | intron | c.46-4157C>T | - | - |
| *TTC33* | 40568560 | C/T | 3' UTR | c.*844G>A | - | rs1429871231 |
| *TTC33* | 40568560 | C/T | intron | c.326+35372G>A | - | rs1429871231 |
| *TTC33* | 40600444 | A/- | intron | c.326+3488del | - | rs137967063 |
| *TTC33* | 40641632 | C/A | intron | c.222-28251G>T | - | rs543273410 |
| *PRDM9* | 23455682 | T/A | intron | c.-265-6990T>A | - | rs535859370 |
| *PRDM9* | 23463952 | G/A | intron | c.-130+1145G>A | - | rs1450029786 |
| *PRDM9* | 23490052 | T/A | intron | c.-129-3419T>A | - | rs562687617 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *PRLR* | 35063425 | T/C | 3' UTR | c.*1664A>G | - | - |
| *PRLR* | 35063425 | T/C | 3' UTR | c.*1664A>G | - | - |
| *IL31RA* | 55858516 | C/T | upstream | - | - | rs375213312 |
| *BASP1* | 17092052 | G/A | intron | c.-73+2891G>A | - | rs959954745 |
| *BASP1* | 17105905 | G/C | intron | c.-73+16744G>C | - | - |
| *CPLANE1* | 37153907 | C/G | missense | c.5242G>C | p.Gly1748Arg | rs768671286 |
| *SLC9A3* | 473251 | C/T | downstream | - | - | - |
| *TARS1* | 33453598 | G/A | downstream | - | - | rs553354497 |
| *SELENOP* | 42825470 | A/- | intron | c.-13-17104del | - | rs575053433 |
| *SELENOP* | 42828511 | A/G | intron | c.-13-20145T>C | - | rs1357928065 |
| *SELENOP* | 42847360 | AAAT/- | intron | c.-13-38994_-13-38991del | - | - |
| *MYO10* | 16936160 | G/C | 5' UTR | c.-352C>G | - | - |
| *NUP155* | 37314221 | T/C | missense | c.2413A>G | p.Ile805Val | rs376806446 |
| *TRIO* | 14290132 | G/A | upstream | - | - | rs1296872989 |
| *TRIO* | 14522399 | -/T | intron | c.7611-9332dup | - | rs397788787 |
| *CPLANE1* | 37153907 | C/G | missense | c.1320G>C | p.Gly441Arg | rs768671286 |
| *DAB2* | 39445086 | ATATACAC/- | intron | c.-102+16992_-102+16999del | - | rs1383148855 |
| *FYB1* | 39245451 | G/T | intron | c.-28+28952C>A | - | rs531894684 |
| *ZDHHC11B* | 713271 | A/G | intron | c.*8-989T>C | - | - |
| *RPL37* | 40833567 | A/G | downstream | - | - | rs551681386 |
| *CPLANE1* | 37153907 | C/G | missense | c.8044G>C | p.Gly2682Arg | rs768671286 |
| *CDH18* | 20024603 | A/G | intron | c.-517-32589T>C | - | - |
| *CDH18* | 20172213 | T/C | intron | c.-518+83231A>G | - | rs866935060 |
| *CDH18* | 20172228 | A/G | intron | c.-518+83216T>C | - | rs1451244688 |
| *CDH18* | 20496061 | A/G | intron | c.-580+79401T>C | - | rs545190198 |
| *CDH18* | 20563902 | T/C | intron | c.-580+11560A>G | - | rs1561135176 |
| *ZNF131* | 43081493 | CA/- | intron | c.-185+13277_-185+13278del | - | rs1195955139 |
| *ZNF131* | 43122948 | G/A | downstream | - | - | rs575153747 |
| *MYO10* | 16936160 | G/C | 5' UTR | c.-352C>G | - | - |
| *TAS2R1* | 9710519 | C/T | intron | c.-242+1653G>A | - | rs568740699 |
| *SELENOP* | 42825470 | A/- | 5' UTR | c.-43del | - | rs575053433 |
| *SUB1* | 32568731 | G/A | intron | c.-1-19781G>A | - | rs555293599 |
| *SUB1* | 32578000 | T/C | intron | c.-1-10512T>C | - | - |
| *IRX4* | 1887353 | C/T | upstream | - | - | rs1253069350 |
| *TTC23L* | 34894052 | -/T | intron | c.1078-2717dup | - | - |
| *PDE4D* | 59586417 | C/T | upstream | - | - | rs369106859 |
| *SLC1A3* | 36630351 | C/A | downstream | - | - | rs753010927 |
| *IL7R* | 35873450 | C/T | upstream | - | - | rs1165054639 |
| *FBXL7* | 15503437 | AAG/- | intron | c.37+2724_37+2726del | - | - |
| *FBXL7* | 15582013 | TGCAGTGGCATGATCTTAGCTCAT/- | intron | c.38-33989_38-33966del | - | rs566181914 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *FBXL7* | 15595778 | C/A | intron | c.38-20205C>A | - | rs561989652 |
| *FBXL7* | 15679558 | -/T | intron | c.127+63505dup | - | rs34375125 |
| *FBXL7* | 15713599 | C/A | intron | c.127+97527C>A | - | - |
| *FBXL7* | 15760428 | AA/- | intron | c.127+144367_127+144368del | - | rs74868826 |
| *FBXL7* | 15793861 | A/- | intron | c.128-134025del | - | rs532552908 |
| *FBXL7* | 15820770 | G/A | intron | c.128-107120G>A | - | rs1447340017 |
| *FBXL7* | 15875336 | -/CAGGACATA | intron | c.128-52554_128-52553insCAGGACATA | - | - |
| *FBXL7* | 15884234 | A/G | intron | c.128-43656A>G | - | rs547948490 |
| *FBXL7* | 15895635 | TT/- | intron | c.128-32232_128-32231del | - | rs903652361 |
| *PRLR* | 35192501 | G/C | intron | c.-293+37767C>G | - | rs906344590 |
| *PRLR* | 35098543 | T/C | intron | c.-43-8880A>G | - | rs967185855 |
| *PRLR* | 35101986 | C/T | intron | c.-43-12323G>A | - | - |
| *PDE4D* | 59908631 | G/A | intron | c.272+79857C>T | - | - |
| *PDE4D* | 59928054 | G/A | intron | c.272+60434C>T | - | rs575721152 |
| *PDE4D* | 59970096 | C/T | intron | c.272+18392G>A | - | rs532020112 |
| *PDE4D* | 60004290 | G/A | intron | c.43-15573C>T | - | rs556348371 |
| *PDE4D* | 60019798 | C/T | intron | c.43-31081G>A | - | rs565170779 |
| *PDE4D* | 60034442 | G/A | intron | c.43-45725C>T | - | rs534545515 |
| *PDE4D* | 60040335 | G/C | intron | c.43-51618C>G | - | rs532796587 |
| *PDE4D* | 60056643 | G/A | intron | c.43-67926C>T | - | rs569001317 |
| *PDE4D* | 60129629 | C/A | intron | c.42+55928G>T | - | rs540763015 |
| *PDE4D* | 60132322 | G/C | intron | c.42+53235C>G | - | rs571589405 |
| *PDE4D* | 60207817 | G/T | intron | c.-89-22130C>A | - | - |
| *PDE4D* | 60246375 | C/T | intron | c.-89-60688G>A | - | rs574656004 |
| *PDE4D* | 60299119 | -/T | intron | c.-89-113433dup | - | rs973044082 |
| *PDE4D* | 60340277 | -/T | intron | c.-90+147664dup | - | - |
| *PDE4D* | 60377171 | T/C | intron | c.-90+110771A>G | - | rs553402147 |
| *MYO10* | 16936160 | G/C | upstream | - | - | - |
| *ARL15* | 53913286 | G/A | intron | c.-75-26573C>T | - | rs147664991 |
| *ARL15* | 53964826 | T/G | intron | c.-75-78113A>C | - | rs538305319 |
| *ARL15* | 54128095 | C/T | intron | c.-284-14685G>A | - | - |
| *ARL15* | 54180021 | AAA/- | intron | c.-344-25382_-344-25380del | - | rs11336058 |
| *BRD9* | 891308 | A/- | 5' UTR | c.-221del | - | - |
| *BRD9* | 891308 | A/- | intron | c.268-21del | - | - |
| *MIER3* | 56970934 | C/A | intron | c.-105+699G>T | - | rs546939194 |
| *C5orf67* | 56512242 | AAAAAAAATATATAT/- | intron | c.370-506_370-492del | - | - |
| *C5orf67* | 56528976 | ATATATATATACAC/- | intron | c.205+4106_205+4119del | - | rs761310547 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *PDZD2* | 31677334 | -/ATGTGATGGAGTACTATTTAGTGATAAAAAGAATGAGCTATTGACACATGCAACAAT | intron | c.-361+37897_-361+37898insATGTGATGGAGTACTATTTAGTGATAAAAAGAAATGAGCTATTGACACATGCAACAAT | - | - |
| *PDZD2* | 31792342 | T/C | intron | c.-360-6547T>C | - | - |
| *PDZD2* | 31913592 | A/T | intron | c.477-69563A>T | - | - |
| *PDZD2* | 31913915 | T/G | intron | c.477-69240T>G | - | - |
| *PDZD2* | 31916023 | G/T | intron | c.477-67132G>T | - | - |
| *PDZD2* | 31990518 | G/A | intron | c.979-5058G>A | - | rs367734526 |
| *PDZD2* | 32024001 | G/A | intron | c.1408-13230G>A | - | rs533028685 |
| *PDZD2* | 32094643 | T/- | intron | c.7845+1619del | - | rs539102907 |
| *ELOVL7* | 60797646 | AGG/- | intron | c.-34-10215_-34-10213del | - | - |
| *CPLANE1* | 37132327 | G/C | intron | c.8630+6393C>G | - | rs535402320 |
| *CPLANE1* | 37144634 | G/A | intron | c.8300-2154C>T | - | rs1298991140 |
| *CPLANE1* | 37147973 | AA/- | intron | c.8299+208_8299+209del | - | rs11284644 |
| *CPLANE1* | 37153907 | C/G | missense | c.8044G>C | p.Gly2682Arg | rs768671286 |
| *CPLANE1* | 37159097 | TTT/- | intron | c.7691-752_7691-750del | - | rs34035923 |
| *CPLANE1* | 37178230 | T/C | intron | c.5821-530A>G | - | rs560006401 |
| *CPLANE1* | 37190180 | C/G | intron | c.3812-2338G>C | - | - |
| *GPBP1* | 57188674 | AAACACGTCAGGC/- | intron | c.-58+12263_-58+12275del | - | rs575100643 |
| *UBE2QL1* | 6466115 | G/A | intron | c.354+16868G>A | - | rs528545349 |
| *UBE2QL1* | 6478416 | G/A | intron | c.355-12802G>A | - | rs566577541 |
| *MROH2B* | 41033375 | C/T | intron | c.2242-215G>A | - | rs528977438 |
| *MAP3K1* | 56875527 | ACTGTTTT/- | intron | c.1965+213_1965+220del | - | rs542544816 |
| *SEMA5A* | 9071874 | G/T | intron | c.2074-5228C>A | - | rs536703501 |
| *SEMA5A* | 9131659 | AAAAAAAAAAAAAAAAAAAAC/- | intron | c.1599+4845_1599+4870del | - | rs1304510069 |
| *SEMA5A* | 9146360 | C/T | intron | c.1481+8128G>A | - | rs980924251 |
| *SEMA5A* | 9323659 | TT/- | intron | c.225-5242_225-5241del | - | rs35243676 |
| *SEMA5A* | 9351399 | C/A | intron | c.125-13587G>T | - | rs576042681 |
| *SEMA5A* | 9392870 | A/G | intron | c.-77-12847T>C | - | rs1044264159 |
| *SEMA5A* | 9432179 | G/A | intron | c.-78+5577C>T | - | rs183822633 |
| *SEMA5A* | 9492674 | T/C | intron | c.-175+52910A>G | - | rs540206987 |
| *SEMA5A* | 9525726 | A/T | intron | c.-175+19858T>A | - | rs534444657 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *CDH18* | 19843530 | A/G | intron | c.-256-4288T>C | - | rs563538642 |
| *CDH18* | 19848326 | C/T | intron | c.-256-9084G>A | - | - |
| *CDH18* | 19891206 | G/A | intron | c.-256-51964C>T | - | rs532113616 |
| *CDH18* | 19905408 | TTTAT/- | intron | c.-256-66166_-256-66162del | - | - |
| *CDH18* | 19917893 | T/- | intron | c.-257+63167del | - | rs534152701 |
| *CDH18* | 19984427 | T/C | intron | c.-375-3249A>G | - | rs753841968 |
| *CDH12* | 22035384 | TG/- | intron | c.231+43062_231+43063del | - | rs146264911 |
| *CDH12* | 22093651 | G/C | intron | c.-186-14789C>G | - | rs570478581 |
| *CDH12* | 22153889 | -/TATATGTAT | intron | c.-187+58608_-187+58609insATACATATA | - | - |
| *CDH12* | 22226528 | T/C | intron | c.-332-13885A>G | - | - |
| *CDH12* | 22230133 | T/C | intron | c.-332-17490A>G | - | - |
| *CDH12* | 22260909 | A/G | intron | c.-332-48266T>C | - | rs1038898494 |
| *CDH12* | 22277484 | T/C | intron | c.-332-64841A>G | - | rs752655476 |
| *CDH12* | 22308933 | AGAAAGAGAGAGAGAGGAGAGAGGAGAGAGAGAGAGAGAGAGAA/- | intron | c.-332-96290_-332-96238del | - | - |
| *CDH12* | 22314954 | TTTTTTTTTTTTT/- | intron | c.-333+90303_-333+90315del | - | rs759734847 |
| *CDH12* | 22327432 | TGTG/"* | intron | - | - | - |
| *CDH12* | 22332080 | G/T | intron | c.-333+73177C>A | - | - |
| *CDH12* | 22335580 | A/G | intron | c.-333+69677T>C | - | - |
| *CDH12* | 22350951 | G/T | intron | c.-333+54306C>A | - | rs192043592 |
| *CDH12* | 22366769 | G/C | intron | c.-333+38488C>G | - | - |
| *CDH12* | 22425079 | TAG/"* | intron | - | - | - |
| *CDH12* | 22456903 | T/A | intron | c.-428+48367A>T | - | rs1004330329 |
| *CDH12* | 22461473 | A/T | intron | c.-428+43797T>A | - | rs547972371 |
| *CDH12* | 22653992 | CCTT/- | intron | c.-522-148628_-522-148625del | - | rs1023647653 |
| *CDH12* | 22661765 | A/G | intron | c.-522-156401T>C | - | rs568742391,COSV66434584 |
| *CDH12* | 22696952 | A/- | intron | c.-523+156106del | - | rs546877552 |
| *CDH12* | 22704576 | C/T | intron | c.-523+148482G>A | - | rs933948012 |
| *CDH12* | 22708319 | A/G | intron | c.-523+144739T>C | - | - |
| *CDH12* | 22724541 | A/C | intron | c.-523+128517T>G | - | rs934347497 |
| *CDH12* | 22752574 | TTTTTTTTTTTTTC/- | intron | c.-523+100484_-523+100501del | - | rs1369642200 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *CDH12* | 22778114 | -/T | intron | c.-523+74943dup | - | rs367585244 |
| *CDH12* | 22781706 | C/G | intron | c.-523+71352G>C | - | rs1278758828 |
| *CDH12* | 22828612 | T/A | intron | c.-523+24446A>T | - | rs566834004 |
| *C1QTNF3* | 34018448 | C/T | downstream | - | - | rs528976830 |
| *NUP155* | 37314221 | T/C | missense | c.2236A>G | p.Ile746Val | rs376806446 |
| *SNX18* | 54543800 | G/A | 3' UTR | c.*368G>A | - | - |
| *SNX18* | 54544033 | -/T | 3' UTR | c.*602dup | - | - |
| *SNX18* | 54544468 | TTTAC/- | 3' UTR | c.*1034_*1038del | - | - |
| *SNX18* | 54545992 | TG/- | 3' UTR | c.*2560_*2561del | - | - |
| *SNX18* | 54546544 | A/G | 3' UTR | c.*3112A>G | - | rs186817959 |
| *SNX18* | 54547387 | G/C | downstream | - | - | rs192548012 |
| *CDC20B* | 55114253 | C/T | missense | c.1525G>A | p.Ala509Thr | rs200898156 |
| *IL6ST* | 55994097 | T/- | intron | c.-104+687del | - | rs770968541 |
| *PLCXD3* | 41312738 | TCCTTCCTTCTG/- | 3' UTR | c.*879_*890del | - | - |
| *CCL28* | 43392828 | A/G | intron | c.65-4352T>C | - | - |
| *PDE4D* | 59586417 | C/T | 5' UTR | c.-42G>A | - | rs369106859 |
| *SPEF2* | 35724558 | G/C | intron | c.2915-3117G>C | - | rs1162065975 |
| *SPEF2* | 35806386 | C/A | intron | c.5011-321C>A | - | rs1193292484 |
| *DDX4* | 55749754 | TTA/- | intron | c.127+3531_127+3533del | - | - |
| *DDX4* | 55749767 | GGATTTTTTGCC/- | intron | c.127+3535_127+3546del | - | - |
| *DDX4* | 55749771 | CC/- | intron | c.127+3551_127+3552del | - | - |
| *DDX4* | 55783665 | G/A | intron | c.524-1632G>A | - | rs1345403140 |
| *DDX4* | 55786494 | C/T | intron | c.763-24C>T | - | rs199606403 |
| *ADAMTS12* | 33583973 | A/G | intron | c.2610+4626T>C | - | rs540891822 |
| *ADAMTS12* | 33629755 | T/G | intron | c.1888+7822A>C | - | rs545737603 |
| *ADAMTS12* | 33643814 | T/C | intron | c.1480-344A>G | - | rs1034794433 |
| *ADAMTS12* | 33645268 | G/C | intron | c.1480-1798C>G | - | rs1275581645 |
| *ADAMTS12* | 33714597 | A/C | intron | c.635-30542T>G | - | rs548891171 |
| *ADAMTS12* | 33717968 | G/A | intron | c.634+33436C>T | - | rs538396868 |
| *ADAMTS12* | 33751838 | A/G | intron | c.490-290T>C | - | - |
| *ADAMTS12* | 33789340 | C/T | intron | c.490-37792G>A | - | rs567228341 |
| *FYB1* | 39122637 | C/T | intron | c.1934-235G>A | - | rs1459733480,COSV60944363 |
| *FYB1* | 39148380 | -/TTTTTTTT | intron | c.1292+5067_1292+5068insAAAAAAAA | - | - |
| *DROSHA* | 31441968 | A/C | intron | c.2883-4670T>G | - | - |
| *TRIO* | 14272956 | T/C | intron | c.232+2057T>C | - | rs535946536 |
| *TRIO* | 14290132 | G/A | intron | c.541-584G>A | - | rs1296872989 |
| *TRIO* | 14293478 | C/T | intron | c.1176+344C>T | - | - |
| *TRIO* | 14338153 | G/A | intron | c.2046+1426G>A | - | - |
| *TRIO* | 14448346 | C/T | intron | c.5204-12673C>T | - | rs1231237757 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *ANKRD55* | 56112511 | -/AAAAAAAAAAAAAAAC | intron | c.966-730_966-729insGTTTTTTTTTTTTTTT | - | - |
| *ANKRD55* | 56152633 | A/G | intron | c.483+7200T>C | - | rs192608624 |
| *ANKRD55* | 56172754 | T/C | intron | c.313-1951A>G | - | rs375272895 |
| *ANKRD55* | 56210564 | AAAA/- | intron | c.58+22292_58+22295del | - | rs59566826 |
| *ANKRD55* | 56227676 | A/G | intron | c.58+5180T>C | - | rs373142766 |
| *ANKRD55* | 56233926 | T/C | upstream | - | - | rs371573990 |
| *PDE4D* | 59019870 | -/CTATCTAT | intron | c.921+18981_921+18988dup | - | rs70973175 |
| *PDE4D* | 59053664 | T/G | intron | c.809-14693A>C | - | rs866347206 |
| *PDE4D* | 59077542 | C/T | intron | c.809-38571G>A | - | rs974596527 |
| *PDE4D* | 59105886 | A/G | intron | c.809-66915T>C | - | - |
| *PDE4D* | 59140163 | T/A | intron | c.808+40432A>T | - | rs1035386339 |
| *PDE4D* | 59167215 | C/T | intron | c.808+13380G>A | - | rs558439068 |
| *PDE4D* | 59170881 | -/T | intron | c.808+9713dup | - | rs1260829199 |
| *PDE4D* | 59186404 | T/C | intron | c.685-1142A>G | - | - |
| *PDE4D* | 59226856 | T/- | intron | c.456-10888del | - | rs143154289 |
| *PDE4D* | 59233883 | C/T | intron | c.456-17915G>A | - | rs1008502587 |
| *PDE4D* | 59243704 | C/T | intron | c.456-27736G>A | - | rs914755044 |
| *PDE4D* | 59311504 | AAAAA/- | intron | c.456-95536_456-95532del | - | rs35020719 |
| *PDE4D* | 59349286 | T/C | intron | c.456-133318A>G | - | rs888828502 |
| *PDE4D* | 59352845 | G/C | intron | c.456-136877C>G | - | - |
| *PDE4D* | 59359032 | T/- | intron | c.456-143064del | - | rs543194281 |
| *PDE4D* | 59372748 | T/G | intron | c.456-156780A>C | - | - |
| *PDE4D* | 59553491 | G/T | intron | c.456-337523C>A | - | - |
| *PDE4D* | 59565672 | C/T | intron | c.455+327496G>A | - | rs578151880 |
| *PDE4D* | 59574435 | T/C | intron | c.455+318733A>G | - | - |
| *PDE4D* | 59581899 | A/G | intron | c.455+311269T>C | - | rs549436448 |
| *PDE4D* | 59586417 | C/T | intron | c.455+306751G>A | - | rs369106859 |
| *PDE4D* | 59610748 | C/T | intron | c.455+282420G>A | - | rs548231317 |
| *PDE4D* | 59616945 | A/C | intron | c.455+276223T>G | - | rs547471208 |
| *PDE4D* | 59621960 | G/A | intron | c.455+271208C>T | - | rs376764926 |
| *PDE4D* | 59632723 | T/A | intron | c.455+260445A>T | - | rs376420199 |
| *PDE4D* | 59728660 | GAAG/- | intron | c.455+164508_455+164511del | - | - |
| *PDE4D* | 59735830 | A/C | intron | c.455+157338T>G | - | rs568577960 |
| *PDE4D* | 59746749 | T/C | intron | c.455+146419A>G | - | - |
| *PDE4D* | 59762376 | TATATGTGTATA | intron | c.455+130792_455+130817del | - | - |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | TGGGTACACATGTG/- | | | | |
| *PDE4D* | 59773905 | C/T | intron | c.455+119263G>A | - | - |
| *PDE4D* | 59781984 | C/T | intron | c.455+111184G>A | - | - |
| *PDE4D* | 59867152 | A/T | intron | c.455+26016T>A | - | rs567765673 |
| *PDE4D* | 59868224 | T/C | intron | c.455+24944A>G | - | rs541446917 |
| *PDE4D* | 59885851 | -/TA | intron | c.455+7315_455+7316dup | - | - |
| *ADCY2* | 7474962 | -/TCCC | intron | c.409-45774_409-45771dup | - | - |
| *ADCY2* | 7522739 | -/AAAAAAAA | intron | c.570+1871_570+1877dup | - | - |
| *ADCY2* | 7643032 | T/A | intron | c.720+16716T>A | - | - |
| *ADCY2* | 7718443 | C/T | intron | c.1703+1206C>T | - | rs377371975 |
| *ADCY2* | 7775679 | C/- | intron | c.2384+2579del | - | rs527504724 |
| *ADCY2* | 7779239 | T/C | intron | c.2385-5126T>C | - | rs569591465 |
| *TTC33* | 40718602 | T/C | intron | c.436-2104A>G | - | rs539001663 |
| *MIER3* | 56939875 | T/G | intron | c.100-858A>C | - | rs567619032 |
| *MIER3* | 56945273 | G/T | intron | c.99+1653C>A | - | rs575669813 |
| *IL6ST* | 55934732 | T/C | downstream | - | - | - |
| *MARCHF11* | 16117375 | T/C | intron | c.694-26294A>G | - | rs1417277884 |
| *MARCHF11* | 16149490 | CTA/- | intron | c.693+28236_693+28238del | - | - |
| *MARCHF11* | 16156976 | A/G | intron | c.693+20750T>C | - | rs116824407 |
| *PLCXD3* | 41312738 | TCCTTCCTTCTG/- | splice donor, 3' UTR | c.*879_*888+2del | - | - |
| *PLCXD3* | 41339258 | G/C | intron | c.813-25488C>G | - | rs13175338 |
| *PLCXD3* | 41367092 | A/G | intron | c.812+14734T>C | - | - |
| *PLCXD3* | 41428374 | -/TTTT | intron | c.104-45844_104-45841dup | - | rs373244601 |
| *PLCXD3* | 41440215 | -/TTTTTTTTTTTTT | intron | c.104-57696_104-57682dup | - | - |
| *SNX18* | 54530820 | C/T | intron | c.1774+11094C>T | - | rs977446647 |
| *SNX18* | 54543800 | G/A | downstream | - | - | - |
| *SNX18* | 54544033 | -/T | downstream | - | - | - |
| *NIM1K* | 43210431 | T/A | intron | c.-695+18020T>A | - | - |
| *NIM1K* | 43269181 | -/A | intron | c.293-7876_293-7875insA | - | rs921212513 |
| *HMGCS1* | 43286838 | C/T | downstream | - | - | - |
| *BASP1* | 17244696 | TT/- | intron | c.-10+26899_-10+26900del | - | rs79126102 |
| *CDC20B* | 55114253 | C/T | missense | c.1399G>A | p.Ala467Thr | rs200898156 |
| *EGFLAM* | 38260072 | G/A | intron | c.97+1221G>A | - | - |
| *EGFLAM* | 38270314 | T/G | intron | c.97+11463T>G | - | rs116479992 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *EGFLAM* | 38289633 | A/T | intron | c.97+30782A>T | - | rs540261377 |
| *EGFLAM* | 38433286 | A/G | intron | c.2167-1851A>G | - | rs545019420 |
| *DAB2* | 39384951 | A/G | intron | c.688-1680T>C | - | - |
| *DAB2* | 39401528 | T/G | intron | c.-101-7107A>C | - | - |
| *SLC38A9* | 55687079 | T/C | intron | c.113+10767A>G | - | - |
| *SLC38A9* | 55696666 | CCTCCCGGACGGGGCGGCTGGCCGGGCAGAGGGGCTCCTCACTTCCCAGTAGGGGCGGCCGGGCAGAGGCGCCCCTCA/- | intron | c.113+1180_113+1257del | - | rs1561434832 |
| *AHRR* | 380446 | C/G | intron | c.363+3730C>G | - | rs1310675162 |
| *ANXA2R* | 43043110 | G/A | 5' UTR | c.-1360C>T | - | rs1310545162 |
| *C7* | 40952548 | G/- | intron | c.1093+2535del | - | rs1355449459 |
| *C7* | 40982325 | AG/- | 3' UTR | c.*751_*752del | - | rs1348389884 |
| *ZNF131* | 43122948 | G/A | intron | c.125-261G>A | - | rs575153747 |
| *ZNF131* | 43143920 | TTTTTTTTTTTTTTTTTTTTTTTT/- | intron | c.371+4606_371+4627del | - | rs79246574 |
| *ZNF131* | 43147394 | -/T | intron | c.371+8093dup | - | rs769222605 |
| *RETREG1* | 16485298 | A/G | intron | c.459-1826T>C | - | - |
| *RETREG1* | 16506426 | G/A | intron | c.459-22954C>T | - | rs1012423896 |
| *RETREG1* | 16530862 | A/- | intron | c.458+34901del | - | rs1218730697 |
| *CTNND2* | 11045019 | T/G | intron | c.2789-22040A>C | - | rs550742264 |
| *CTNND2* | 11062974 | G/A | intron | c.2788+19722C>T | - | - |
| *CTNND2* | 11097710 | T/C | intron | c.2637+865A>G | - | - |
| *CTNND2* | 11116373 | A/G | intron | c.2277+1077T>C | - | rs774015984 |
| *CTNND2* | 11250491 | -/TCTCTCTCTCTCTATATATA | intron | c.1629-13669_1629-13668insTATATATAGAGAGAGAGAGA | - | rs869141186 |
| *CTNND2* | 11297839 | C/T | intron | c.1628+48533G>A | - | rs1351006057 |
| *CTNND2* | 11324142 | G/C | intron | c.1628+22230C>G | - | rs866657412 |
| *CTNND2* | 11333738 | C/T | intron | c.1628+12634G>A | - | rs559186586 |
| *CTNND2* | 11394301 | C/G | intron | c.612+2730G>C | - | - |
| *CTNND2* | 11398101 | G/A | intron | c.440-898C>T | - | - |
| *CTNND2* | 11477645 | G/A | intron | c.288-65576C>T | - | - |
| *CTNND2* | 11482102 | A/G | intron | c.288-70033T>C | - | - |

| | | | | | | |
|---|---|---|---|---|---|---|
| *CTNND2* | 11546869 | C/T | intron | c.287+18075G>A | - | rs533782301 |
| *CTNND2* | 11556643 | T/C | intron | c.287+8301A>G | - | rs542804647 |
| *CTNND2* | 11559035 | A/G | intron | c.287+5909T>C | - | - |
| *CTNND2* | 11571585 | G/A | intron | c.175-6529C>T | - | - |
| *CTNND2* | 11578681 | AATAAATAAATA/- | intron | c.175-13625_175-13614del | - | rs57654816 |
| *CTNND2* | 11596595 | G/A | intron | c.175-31539C>T | - | rs186082986 |
| *CTNND2* | 11638652 | C/A | intron | c.175-73596G>T | - | rs550098357 |
| *CTNND2* | 11651864 | A/C | intron | c.174+80272T>G | - | - |
| *CTNND2* | 11728609 | A/G | intron | c.174+3527T>C | - | - |
| *CTNND2* | 11743344 | G/A | intron | c.38-11072C>T | - | - |
| *SLC6A19* | 1216131 | G/A | intron | c.888-427G>A | - | - |
| *HCN1* | 45278940 | G/A | intron | c.1619-11687C>T | - | - |
| *HCN1* | 45281314 | TC/- | intron | c.1619-14061_1619-14060del | - | rs368864355 |
| *HCN1* | 45326876 | T/C | intron | c.1378-23037A>G | - | rs561738528 |
| *HCN1* | 45374948 | C/G | intron | c.1230+21544G>C | - | rs1275825931 |
| *HCN1* | 45401657 | -/TT | intron | c.1012-4949_1012-4948dup | - | rs77219002 |
| *HCN1* | 45416642 | A/G | intron | c.1012-19932T>C | - | rs534688990 |
| *HCN1* | 45436929 | A/G | intron | c.1011+24917T>C | - | rs531929024 |
| *HCN1* | 45454685 | G/T | intron | c.1011+7161C>A | - | - |
| *IL7R* | 35857185 | T/C | intron | c.82+126T>C | - | - |
| *IL7R* | 35873450 | C/T | intron | c.538-30C>T | - | rs1165054639 |
| *IL31RA* | 55858516 | C/T | intron | c.7-993C>T | - | rs375213312 |
| *IL31RA* | 55881724 | TTTTTTTT/- | intron | c.398-1306_398-1299del | - | rs34217814 |
| *IL31RA* | 55883700 | A/G | intron | c.549+505A>G | - | rs368572634 |
| *IL31RA* | 55896985 | -/AAAAA | intron | c.795+568_795+572dup | - | rs1317185291 |
| *RICTOR* | 38960737 | T/A | intron | c.1716-204A>T | - | - |
| *RICTOR* | 38966108 | -/A | intron | c.1299+532dup | - | rs1009891875 |
| *RICTOR* | 38985128 | C/T | intron | c.584-3092G>A | - | rs1192515020 |
| *RICTOR* | 38994277 | T/G | intron | c.456+2542A>C | - | rs951656923 |
| *CDC20B* | 55114253 | C/T | missense | c.1513G>A | p.Ala505Thr | rs200898156 |
| *CDC20B* | 55135614 | G/A | intron | c.581-2086C>T | - | rs894501872 |
| *NDUFS4* | 53596962 | C/A | intron | c.99-6490C>A | - | - |
| *NDUFS4* | 53622362 | C/T | intron | c.177+18832C>T | - | rs1445764128 |
| *NDUFS4* | 53643336 | G/T | intron | c.178-2897G>T | - | - |
| *NDUFS4* | 53668854 | C/G | intron | c.424+10230C>G | - | rs553885656 |
| *ANKRD33B* | 10638989 | GCGGA/- | intron | c.637+818_637+822del | - | rs1560984717 |
| *ANKRD33B* | 10641898 | A/G | intron | c.637+3730A>G | - | - |
| *ANKRD33B* | 10646596 | TCTT/- | intron | c.638-2671_638-2668del | - | rs1047577203 |
| *RANBP3L* | 36266001 | G/A | intron | c.269-481C>T | - | rs1275313660 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *RANBP3L* | 36294669 | C/- | intron | c.91+6657del | - | - |
| *LMBRD2* | 36102276 | T/C | 3' UTR | c.*1770A>G | - | - |
| *NDUFAF2* | 60948948 | C/G | intron | c.127+3566C>G | - | rs538879169 |
| *NDUFAF2* | 60968001 | C/T | intron | c.127+22619C>T | - | - |
| *NDUFAF2* | 60994051 | C/T | intron | c.127+48669C>T | - | - |
| *NDUFAF2* | 61005184 | A/G | intron | c.127+59802A>G | - | rs1383724607 |
| *NDUFAF2* | 61005188 | A/T | intron | c.127+59806A>T | - | - |
| *NDUFAF2* | 61013195 | A/T | intron | c.128-59930A>T | - | rs548290091 |
| *NDUFAF2* | 61040290 | ACGCG/"* | intron | - | - | - |
| *NDUFAF2* | 61051919 | C/T | intron | c.128-21206C>T | - | - |
| *NDUFAF2* | 61065931 | A/G | intron | c.128-7194A>G | - | rs561945293 |
| *NDUFAF2* | 61087822 | A/G | intron | c.218-11170A>G | - | rs540821390 |
| *NDUFAF2* | 61125832 | A/G | intron | c.258+26800A>G | - | - |
| *NDUFAF2* | 61135141 | A/- | intron | c.259-17552del | - | rs1427360387 |
| *NDUFAF2* | 61150678 | A/G | intron | c.259-2026A>G | - | rs1017600370 |
| *SLC45A2* | 33978634 | C/- | intron | c.562+3602del | - | rs1231793266 |
| *SLC45A2* | 33979453 | T/C | intron | c.562+2783A>G | - | - |
| *ITGA2* | 52997555 | C/T | intron | c.64+8023C>T | - | rs573953177 |
| *ICE1* | 5435665 | T/G | intron | c.85-753T>G | - | rs1171881305 |
| *ICE1* | 5458274 | G/T | intron | c.1101+533G>T | - | rs539934672 |
| *ICE1* | 5478278 | C/T | intron | c.6520+2199C>T | - | rs535623205 |
| *ANKH* | 14769792 | CC/- | intron | c.97-601_97-600del | - | rs1307759653 |
| *ANKH* | 14769796 | AA/- | intron | c.97-605_97-604del | - | rs1217430727 |
| *ANKH* | 14769799 | -/CGGCGAC | intron | c.97-609_97-608insGTCGCCG | - | rs1255618085 |
| *ANKH* | 14769803 | TT/- | intron | c.97-612_97-611del | - | rs1337495788 |
| *ANKH* | 14769806 | T/- | intron | c.97-615del | - | rs1214856162 |
| *ANKH* | 14796453 | -/A | intron | c.97-27263dup | - | rs56898794 |
| *ANKH* | 14797075 | T/C | intron | c.97-27884A>G | - | - |
| *ANKH* | 14808252 | -/T | intron | c.97-39062dup | - | rs35952510 |
| *ANKH* | 14848221 | G/A | intron | c.96+23131C>T | - | - |
| *ANKH* | 14849534 | A/T | intron | c.96+21818T>A | - | - |
| *PLEKHG4B* | 142677 | C/A | intron | c.410-370C>A | - | - |
| *PLEKHG4B* | 148359 | A/T | intron | c.838-3154A>T | - | rs538418751 |
| *RAB3C* | 58629688 | A/G | intron | c.252+11818A>G | - | - |
| *RAB3C* | 58643625 | G/A | intron | c.252+25755G>A | - | rs549106770 |
| *RAB3C* | 58649469 | C/T | intron | c.252+31599C>T | - | rs191854778 |
| *RAB3C* | 58725295 | A/G | intron | c.253-707A>G | - | rs553148416 |
| *RAB3C* | 58755660 | C/T | intron | c.371+29540C>T | - | - |
| *RAB3C* | 58838667 | A/C | intron | c.497-12497A>C | - | rs1291380619 |
| *ITGA1* | 52832783 | -/TGTC | intron | c.62-16579_62-16578insCTGT | - | - |
| *ITGA1* | 52910977 | T/C | intron | c.1857+558T>C | - | rs1218815089 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *CCNO* | 55233064 | -/AGA | intron | c.381+78_381+79 insTCT | - | rs538223997 |
| *CCNO* | 55233067 | CC/- | intron | c.381+76_381+77 del | - | rs1395605645 |
| *NIPBL* | 37004400 | T/- | intron | c.3855+1064del | - | rs1209687707 |
| *NIPBL* | 37007662 | C/T | intron | c.4239+188C>T | - | - |
| *NIPBL* | 37046208 | A/T | intron | c.6589+9A>T | - | rs370709104 |
| *NIPBL* | 37055390 | AGAAA /- | intron | c.7264-1790_7264-1786del | - | rs1365490802 |
| *NIPBL* | 37055836 | T/C | intron | c.7264-1350T>C | - | - |
| *PARP8* | 50769421 | A/G | intron | c.518+6179A>G | - | rs556271521 |
| *PARP8* | 50830137 | C/T | intron | c.2233+176C>T | - | - |
| *UGT3A1* | 35965154 | G/A | intron | c.843+232C>T | - | rs571016143 |
| *UGT3A1* | 35986272 | A/G | intron | c.196+2178T>C | - | - |
| *OSMR* | 38852718 | -/TTT | intron | c.-14+6359_-14+6361dup | - | - |
| *OSMR* | 38882303 | G/A | intron | c.418+539G>A | - | rs573897797 |
| *RPL37* | 40833567 | A/G | intron | c.224+614T>C | - | rs551681386 |
| *MYO10* | 16677417 | T/- | intron | c.4576-1263del | - | rs796474728 |
| *MYO10* | 16739628 | G/A | intron | c.1929+15200C>T | - | rs1386999752 |
| *MYO10* | 16821924 | A/"* | intron | - | - | - |
| *MYO10* | 16828108 | G/A | intron | c.121-9941C>T | - | rs923531194 |
| *MYO10* | 16848633 | G/A | intron | c.120+28976C>T | - | rs1325685583 |
| *MYO10* | 16850136 | G/A | intron | c.120+27473C>T | - | rs1197589668 |
| *MYO10* | 16857836 | G/A | intron | c.120+19773C>T | - | rs534921840 |
| *MYO10* | 16859056 | T/A | intron | c.120+18553A>T | - | rs574463541 |
| *MYO10* | 16893616 | G/A | intron | c.22-15909C>T | - | rs1242492830 |
| *MYO10* | 16893654 | -/AAATTATT | intron | c.22-15955_22-15948dup | - | rs1491144924 |
| *MYO10* | 16936160 | G/C | 5' UTR | c.-352C>G | - | - |
| *CDH18* | 19494081 | C/T | intron | c.1630+8911G>A | - | - |
| *CDH18* | 19573881 | T/A | intron | c.1000-2049A>T | - | - |
| *CDH18* | 19576090 | G/T | intron | c.1000-4258C>A | - | - |
| *CDH18* | 19605489 | -/T | intron | c.811+6944_811+6945insA | - | - |
| *CDH18* | 19605490 | A/T | intron | c.811+6944T>A | - | - |
| *CDH18* | 19643242 | A/G | intron | c.644-30641T>C | - | rs568565468 |
| *CDH18* | 19667509 | AC/"* | intron | - | - | - |
| *CDH18* | 19684818 | A/G | intron | c.643+36529T>C | - | rs534445947 |
| *ROPN1L* | 10454986 | G/A | intron | c.417+4873G>A | - | rs572607220 |
| *SLC1A3* | 36630351 | C/A | intron | c.319+764C>A | - | rs753010927 |
| *TARS1* | 33452960 | A/- | intron | c.330-317del | - | rs79630452 |
| *TARS1* | 33453598 | G/A | intron | c.453+186G>A | - | rs553354497 |
| *TARS1* | 33470528 | G/C | downstream | - | - | rs530680034 |
| *RAI14* | 34664416 | AA/- | intron | c.-49+7954_-49+7955del | - | rs11311801 |
| *RAI14* | 34757154 | G/C | intron | c.37-314G>C | - | - |

| | | | | | | |
|---|---|---|---|---|---|---|
| *RAI14* | 34762933 | TGTG/"* | intron | - | - | - |
| *RAI14* | 34781992 | G/A | intron | c.168-13947G>A | - | rs566862592 |
| *WDR70* | 37387998 | C/T | intron | c.176-4002C>T | - | rs531453489 |
| *WDR70* | 37471351 | A/C | intron | c.687-8483A>C | - | - |
| *WDR70* | 37475060 | A/G | intron | c.687-4774A>G | - | - |
| *WDR70* | 37511834 | A/G | intron | c.841-4680A>G | - | rs531683665 |
| *WDR70* | 37523755 | C/T | intron | c.917+7165C>T | - | rs565133674 |
| *WDR70* | 37563093 | CCTCCCGGACGGGGCGGCTGGCCGGGCGGGGGGCTGACCCCCCCGCCTC/- | intron | c.918-41998_918-41949del | - | - |
| *WDR70* | 37597187 | A/G | intron | c.918-7877A>G | - | - |
| *WDR70* | 37605881 | A/G | intron | c.1092+643A>G | - | rs916993226 |
| *WDR70* | 37666973 | A/G | intron | c.1093-30682A>G | - | - |
| *WDR70* | 37667889 | A/C | intron | c.1093-29766A>C | - | - |
| *WDR70* | 37738949 | A/G | intron | c.1877+11904A>G | - | rs924426680 |
| *WDR70* | 37744707 | A/T | intron | c.1878-7779A>T | - | - |
| *DNAH5* | 13736456 | G/A | intron | c.11456-524C>T | - | rs551371231 |
| *DNAH5* | 13777404 | G/C | intron | c.8952-49C>G | - | rs533347451 |
| *DNAH5* | 13852165 | T/G | intron | c.4951-1350A>C | - | - |
| *DNAH5* | 13931214 | G/T | synonymous | c.88C>A | p.Arg30= | rs114220185 |
| *NPR3* | 32751993 | G/A | intron | c.1059+12963G>A | - | rs532152369 |
| *NPR3* | 32760089 | G/A | intron | c.1060-14619G>A | - | rs10461914 |
| *NPR3* | 32771050 | -/T | intron | c.1060-3646dup | - | rs540902606 |
| *NPR3* | 32772651 | G/T | intron | c.1060-2057G>T | - | - |
| *CDH6* | 31300889 | G/A | intron | c.812-1222G>A | - | - |
| *GOLPH3* | 32149890 | G/A | intron | c.226-6010C>T | - | rs1041348241 |
| *ZFR* | 32367018 | C/G | intron | c.2836-2743G>C | - | - |
| *ZFR* | 32408326 | -/C | intron | c.785-1306_785-1305insG | - | rs1554073133 |
| *ZFR* | 32415828 | G/A | intron | c.566-641C>T | - | - |
| *ERCC8* | 60873441 | G/C | downstream | - | - | rs547479737 |
| *ERCC8* | 60919668 | T/G | intron | c.276-1280A>C | - | rs4647083 |
| *ERCC8* | 60941218 | C/T | intron | c.77+3714G>A | - | rs566583898 |
| *DEPDC1B* | 60618416 | GAG/- | intron | c.899-12560_899-12558del | - | - |
| *DEPDC1B* | 60625945 | C/T | intron | c.898+12805G>A | - | - |
| *DEPDC1B* | 60632237 | C/T | intron | c.898+6513G>A | - | rs531078156 |
| *DEPDC1B* | 60633945 | C/T | intron | c.898+4805G>A | - | - |
| *DEPDC1B* | 60696410 | A/G | intron | c.48+3636T>C | - | - |

| | | | | | | |
|---|---|---|---|---|---|---|
| *SLC9A3* | 473251 | C/T | 3' UTR | c.*128G>A | - | - |
| *MTMR12* | 32307746 | A/- | intron | c.81+5012del | - | rs1281665765 |
| *GPBP1* | 57229230 | -/AAAAAAA | intron | c.85-1585_85-1579dup | - | rs543005934 |
| *PLPP1* | 55443192 | -/AAAAAAAAAAAAATATAT | intron | c.495-1285_495-1284insATATATATTTTTTTTTTTTTT | - | rs1204621666 |
| *PLPP1* | 55532345 | T/A | intron | c.58+2227A>T | - | rs547367114 |
| *MTRR* | 7880693 | T/C | intron | c.861+2371T>C | - | - |
| *MTRR* | 7882263 | T/G | intron | c.862-892T>G | - | rs574484046 |
| *FGF10* | 44308614 | T/C | intron | c.429+1813A>G | - | rs546765823 |
| *FGF10* | 44329080 | C/T | intron | c.326-18550G>A | - | rs1443965218 |
| *FGF10* | 44337520 | A/C | intron | c.326-26990T>G | - | - |
| *NNT* | 43676998 | A/C | intron | c.2795-727A>C | - | rs967327212 |
| *CDH10* | 24529417 | AA/- | intron | c.814+5695_814+5696del | - | rs979727439 |
| *CDH10* | 24553344 | G/A | intron | c.232-15670C>T | - | rs529464745 |
| *CDH10* | 24580216 | G/A | intron | c.231+13044C>T | - | rs1217382005 |
| *CDH10* | 24586844 | T/- | intron | c.231+6416del | - | rs1001227038 |
| *CDH10* | 24595626 | C/T | intron | c.-123-2013G>A | - | rs554020550 |
| *C6* | 41194012 | C/T | intron | c.587+1780G>A | - | rs577937091 |
| *LIFR* | 38528982 | T/- | intron | c.143-142del | - | rs1246932957 |
| *C9* | 39289177 | AAT/- | intron | c.1417-226_1417-224del | - | - |
| *C9* | 39333140 | C/- | intron | c.477-1326del | - | rs753503600 |
| *NUP155* | 37314221 | T/C | missense | c.2413A>G | p.Ile805Val | rs376806446 |
| *NUP155* | 37317008 | A/- | intron | c.2305+980del | - | rs1263259747 |
| *NUP155* | 37352277 | G/A | intron | c.556+460C>T | - | - |
| *PRLR* | 35063425 | T/C | intron | c.1009+2524A>G | - | - |
| *AGXT2* | 34997687 | G/A | downstream | - | - | rs532629004 |
| *AGXT2* | 35013621 | A/C | intron | c.1096+366T>G | - | rs550872430 |
| *AGXT2* | 35023177 | AAG/- | intron | c.963+2586_963+2588del | - | rs1474371115 |
| *AGXT2* | 35036596 | C/T | intron | c.486+346G>A | - | - |
| *C1QTNF3* | 34028970 | C/A | intron | c.352-87G>T | - | - |
| *CDH9* | 26910223 | -/TAT | intron | c.524-3386_524-3385insATA | - | rs1166873308 |
| *CDH9* | 26916596 | C/T | intron | c.229-672G>A | - | - |
| *CDH9* | 26930921 | C/A | intron | c.229-14997G>T | - | rs898718041 |
| *CDH9* | 26952652 | -/AG | intron | c.228+35453_228+35454insCT | - | rs1554000026 |
| *GHR* | 42428387 | G/A | intron | c.-12+4432G>A | - | rs564009075 |
| *GHR* | 42608108 | C/T | intron | c.71-20930C>T | - | - |
| *ISL1* | 51386480 | A/C | intron | c.219-1010A>C | - | rs187728145 |
| *MTREX* | 55349186 | T/- | intron | c.1241-371del | - | rs369214186 |
| *MTREX* | 55351842 | -/T | intron | c.1431+829dup | - | rs754738083 |
| *OXCT1* | 41836374 | A/G | intron | c.732+4077T>C | - | - |

| | | | | | | |
|---|---|---|---|---|---|---|
| *OXCT1* | 41837667 | A/G | intron | c.732+2784T>C | - | - |
| *TRIP13* | 901470 | A/G | intron | c.535+39A>G | - | rs371178266 |

## A 2.2 Primer pairs for the PCR- and Sanger-based sequencing experiments

| ***GLH35 NGS variant primers*** | |
|---|---|
| **Name** | **Sequence (5'-3')** |
| PLEKHG4B-Ex11-F | caggaaacatccccagaag |
| PLEKHG4B-Ex11-R | actgctaggacgagaggag |
| ZDHHC11-5'UTR-F | agtgtggcctctttctgacg |
| ZDHHC11-5'UTR-R | ccgttcactctggagatgc |
| SLC12A7-Ex18-F | ctggaagcaggaggacaacc |
| SLC12A7-Ex18-R | gcagaggggacagtgag |
| TERT-Ex13-F | ggcaggcagatgacacagag |
| TERT-Ex13-R | caggagttccaaggtgaagc |
| NDUFS6-Ex1-F | ctgggatgaaaacgggtgac |
| NDUFS6-Ex1-R | agcgacagcacaaccttacc |
| NSUN2-EX9-F | gaaggtggaaggatggtg |
| NSUN2-EX9-R | agatggatggtggtggtg |
| DNAH5-Ex52-F | gaggctgatgctgaaacacc |
| DNAH5-Ex52-R | ctccgtggtgaaagcactg |
| FAM134B-3'UTR-F | caagaccaggaagcagaagc |
| FAM134B-3'UTR-R | agagatggcagtcaatgg |
| MYO10-Ex27-F | aagctcaagggcaccgtag |
| MYO10-Ex27-R | accaggcaacttccagatcc |
| PRDM9-Ex3-F | agccactcgaccagctttc |
| PRDM9-Ex3-R | aggctgaggcaggagaatc |
| PDZD2-5'UTR-F | tgcttggtgcctggaaag |
| PDZD2-5'UTR-R | gcattgtcctgggtgatg |
| C1QTNF3-Ex4-F | gggtttctcaggtccttg |
| C1QTNF3-Ex4-R | ctgaagtgggttgccagag |
| NIPBL-Ex37-F | cctcagactggctgctaatcc |
| NIPBL-Ex37-R | gctcaaggttacccaggaag |
| C5ORF42-Ex20-F | tcccttccttcactgagtcc |
| C5ORF42-Ex20-R | ggataggttcagcccaggag |
| GDNF-5'UTR-Trns-2-F | gtctccaagtccctgctaac |
| GDNF-5'UTR-Trns-2-R | ggtagttcccacccttcgtc |
| RICTOR-Ex5-F | gagcagcagggctacgag |
| RICTOR-Ex5-R | ggttgtgtgtgtgtgtgtg |
| DAB2-Ex1-F | gcctagcaagttttcgatcc |
| DAB2-Ex1-R | taacctcccacagacacctg |
| TTC33-Ex3-F | ccgtccagcaaaatccac |
| TTC33-Ex3-R | acaggaggctgagacaggag |
| MRPS30-Ex1-F | tacagcaagagggcaaggac |
| MRPS30-Ex1-R | gcttctcgtctaccgactcc |
| ISL1-Ex4-F | cctgctttgtgtgctgagg |
| ISL1-Ex4-R | gtcttctccggctgcttg |
| DHX29-Ex14-F | aaaggtggtgtcccctgttc |
| DHX29-Ex14-R | acttgccctatcctgagcac |
| SLC12A7-Ex3-F | cgcaggaggagatggacag |

| ***GLH35 NGS variant primers*** | |
|---|---|
| **Name** | **Sequence (5'-3')** |
| NUP155-Ex22-R | ccctccaaagccaaaatactc |
| PRKAA1-Ex4-F | cccccagaactcataatcctc |
| PRKAA1-Ex4-R | aggggcttttgcataccac |
| HEATR7B2-Ex34-F | tagtccaggggtgggaaag |
| HEATR7B2-Ex34-R | tgaggtgagtggtgtgtgtg |
| GZMA-Ex5-F | gtcaaggttggtcttaactgc |
| GZMA-Ex5-R | acgcacaaatgactctggtg |
| CDC20B-Ex12-F | cactgagggagaaaactgtcc |
| CDC20B-Ex12-R | tgggaagcagagcaagtaaag |
| DDX4-Ex13-F | cttctgagtggaggcatgttac |
| DDX4-Ex13-R | ccaccccaggaatagttcag |
| SLC6A18-Ex5-F | cagccactctgaccacaagg |
| SLC6A18-Ex5-R | ctcacccctgaccaccag |
| EXOC3-Ex1-F | gagttcctcatcctagttcagc |
| EXOC3-Ex1-R | ctgcctccttgcacattacc |
| SLC9A3-Ex16-F | gtgtggggtcatggatgtg |
| SLC9A3-Ex16-R | ggtccagggaggagagagac |
| SLC9A3-Ex12-F | cgtggggagtcagcctaag |
| SLC9A3-Ex12-R | atcaggcacggaggtcac |
| TRIP13-Ex5-F | gctccctcttctcatgtagg |
| TRIP13-Ex5-R | gagctggatctgcttcacag |
| SLC12A7-Ex9-F | gaattggagcctgcttgc |
| SLC12A7-Ex9-R | ctgcttgggagactcaggtg |
| CLPTM1L-Ex14-F | gcatagacgtggcagctc |
| CLPTM1L-Ex14-R | gaagacgcccttgctcac |
| SLC6A3-Ex13-F | tgagggtgctggtaggtgag |
| SLC6A3-Ex13-R | gcacacgtctccactgtcac |
| MArCh6-Ex24-F | ggttggcagattcagttgtg |
| MArCh6-Ex24-R | gcatatggagtgagacatcc |
| MArCh6-Ex25-F | ctgccccagtccattcttc |
| MArCh6-Ex25-R | ccttgctaatcccacctc |
| ROPN1L-Ex3-F | cctgtcatgcatttggag |
| ROPN1L-Ex3-R | ataataggcctggagcag |
| DROSHA-Ex5-F | cacggtgatttggcatgg |
| DROSHA-Ex5-R | caggtagagcccagagatgc |
| PDZD2-Ex5-F | ccctggcctttgaggtaac |
| PDZD2-Ex5-R | gcactggtgtcttttcccatc |
| RAI14-Ex2-F | cgggagcaggctttaatttg |
| RAI14-Ex2-R | tcaacatgggagtggagagg |
| SPEF2-Ex2-F | gctgtcatcataggttgtgtg |
| SPEF2-Ex2-R | tgcactccagtctcagtaacag |
| IL7R-Ex8-F | tgtgtctctctggtgccatc |
| IL7R-Ex8-R | cccattcttgccactctcc |

| | |
|---|---|
| SLC12A7-Ex3-R | cagggcagctcttctgac |
| LIFR-Ex19-F | ccgcagatgaagctggag |
| LIFR-Ex19-R | ggaccaccctcctccattag |
| DHX29-Ex11-F | gcactgtccttggaatcagc |
| DHX29-Ex11-R | cttcggggttgggtacagac |
| ANKRD55-Ex2-F | ctgaggaagttgacctgacc |
| ANKRD55-Ex2-R | tagcctgctctttccgtacc |
| ANKRD55-Ex4-F | agggggagatggagagattc |
| ANKRD55-Ex4-R | gccacagagggatggattag |
| ITGA2-Ex26-F | ttcccagacccctacaagtg |
| ITGA2-Ex26-R | ggagaaaagcagcgtcctg |
| FYB-Ex2-F | gcttgaccaccagagacc |
| FYB-Ex2-R | cgcttttggacctgactcc |
| BRD9-Ex3-F | gccagggtctctgtgaggag |
| BRD9-Ex3-R | ctccacctccaccttcttcc |
| ITGA1-Ex14-F | ggcaagactataaggaaagag |
| ITGA1-Ex14-R | tatgcacactacatatacac |
| ZDHHC11-3'UTR-F | agccacctgcttaactgtgc |
| ZDHHC11-3'UTR-R | tgggctaggctgaaaaactc |
| Med10-Ex4-F | tggagtgttgatgtgaagagc |
| Med10-Ex4-R | gcactctgaaagccagttgac |
| DNAH5-Ex54-F | gttccaggttgacttttcag |
| DNAH5-Ex54-R | aatacccatccccaatagcac |
| CDH18-Ex3-F | tgctaaatgccatcacacac |
| CDH18-Ex3-R | gcaacttactgtagcaaacg |
| CDH18-Ex4-F | tttggggcatgggataag |
| CDH18-Ex4-R | tcagaagacaatagctggagtg |
| CDH12-Ex8-F | ccctggtatgatggctaagtc |
| CDH12-Ex8-R | tgcggtatcacctccacatac |
| IL7R-Ex5-F | ttgctgttgactcctttacg |
| IL7R-Ex5-R | acttgctcccacactttgac |
| C5orf42-Ex40-F | cccagaaatcaagaagctg |
| C5orf42-Ex40-R | ctgaggttgacaaaccctagc |
| NUP155-Ex22-F | atgggggctaggaagaaaac |

| | |
|---|---|
| RICTOR-Ex35-F | atgctgagatgctgatcctc |
| RICTOR-Ex35-R | gtcagtgggtgatataggtg |
| DAB2-Ex13-F | atacactgccatctgcaagc |
| DAB2-Ex13-R | aaatccacatccccaagg |
| DAB2-Ex8-F | cttcctgacctaaacagc |
| DAB2-Ex8-R | acctcctccttctaatcc |
| ITGA1-Ex12-F | ggagaaagcagttgtggtag |
| ITGA1-Ex12-R | gccctgttcccagttgttac |
| ITGA2-Ex13-F | aggaactgtgctctctgtcttc |
| ITGA2-Ex13-R | ggggacatcctcaaaaatg |
| PPAP2A-Ex3-F | gctgggattacaggcatgag |
| PPAP2A-Ex3-R | aacactcacctgccttcc |
| IL31RA-Ex7-F | gattccagttccttgaccac |
| IL31RA-Ex7-R | gaagggaaggaggatggaag |
| DNAH5-Ex5-F | gcaatttcacaacaatgagca |
| DNAH5-Ex5-R | gagccctgtccacgttaaga |
| MYO10-5'UTR-F | aggtcgctgggctgtagtc |
| MYO10-5'UTR-R | acggcagcctttgtctctt |
| NPR3-3'UTR-F | ctaaaagtggcagaacaaatca |
| NPR3-3'UTR-R | tttccaacaccaaaaagcatc |
| C1QTNF3-3'UTR-F | tgacccttggacaacaaaga |
| C1QTNF3-3'UTR-R | cacttcaaacataccaaagcaa |
| PRLR-3'UTR-F | tgggaaactaggtatttgcttc |
| PRLR-3'UTR-R | tgccattctcaaaacattgg |
| LIFR-3'UTR-F | ccaatcagagccatccctaa |
| LIFR-3'UTR-R | tccaaaagtaggggaaaggtg |
| SNX18-3'UTRA-F | tggtggtacaaggacggttt |
| SNX18-3'UTRA-R | atgcactggcctttcactct |
| SNX18-3'UTRb-F | aaacccaagagaaaagccatt |
| SNX18-3'UTRb-R | tcggcatttaagaatctgctg |
| SNX18-3'UTRC-F | ttgccatagccattttcca |
| SNX18-3'UTRC-R | tcctgatgtcactaatggaagc |
| IL6ST-3'UTR-F | tggagtcataaaatggcaagt |
| IL6ST-3'UTR-R | tcatcacacgacccatcaac |

| ***GLH35 coding region gaps coverage*** | |
|---|---|
| **Name** | **Sequence (5'-3')** |
| SLC9A3-Ex1-F | ggggatcggactgtagcc |
| SLC9A3-Ex1-R | gagtctcgctctgctggac |
| TERT Ex1,2-F | agcccctcccctttccttt |
| TERT Ex1,2-R | gtgaaccagcacgtcgtc |
| IRX4-Ex1-F | cgctgctcactttgttaccc |
| IRX4-Ex1-R | cagtcgtaggtcggacactc |
| IRX4-Ex2-F | cgagcatcctggtgtgga |
| IRX4-Ex2-R | cctactggggcaaaagtgg |
| IRX2-Ex3-F | gcccttctacggcaactaca |
| IRX2-Ex3-R | cgccctcactctgctaattc |
| IRX1-Ex1-F | gctcaaggggatgccaatc |
| IRX1-Ex1-R | cgggtgagacagaggagaag |
| KIAA0947-Ex1-F | gccgggtagcgtttcttt |
| KIAA0947-Ex1-R | caagagctacggttgggaag |
| ANKRD33B-Ex1A-F | tgcctggctctgaactctg |

| ***GLH35 coding region gaps coverage*** | |
|---|---|
| **Name** | **Sequence (5'-3')** |
| NADK2-Ex1-F | gcacgtcctggcacttaac |
| NADK2-Ex1-R | gaagagtcgtcccgagaggt |
| FYB-Ex1-Trns3-F | tggccagctatacaaatgctc |
| FYB-Ex1-Trns3-R | tgactgtgaagcaccctcaa |
| GHR-Ex1-Trns2-F | tgtacgtggaggggtttactc |
| GHR-Ex1-Trns2-R | cagcctctggcagttctgt |
| SEPP1-Ex1-Trns3-F | caactctctgccaatgggtta |
| SEPP1-Ex1-Trns3-R | ccaccacgtcctgctaattt |
| ITGA1-Ex1-F | agcgtggagcgcatttag |
| ITGA1-Ex1-R | gggcgctctctcttcctatc |
| SNX18-Ex1-F | tcggagaacccaggagagat |
| SNX18-Ex1-R | gtcgtcccagtcatcatcg |
| MCIDAS-Ex1-F | ggttgccaggttgggttt |
| MCIDAS-Ex1-R | ggatcctcgtacaccgacac |
| MCIDAS-Ex2,3-F | aagccggagaggaaggtg |

| | | | |
|---|---|---|---|
| ANKRD33B-Ex1A-R | tcgtcctcgaaagggtagaa | MCIDAS-Ex2,3-R | gacccgagaggagcttttg |
| ANKRD33B-Ex1b-F | gaccccgctgactacgaa | MCIDAS-Ex4-F | tgcatagtgcctgccaga |
| ANKRD33B-Ex1b-R | gctgacacctcgactgctc | MCIDAS-Ex4-R | tgctggagagtcgttgacag |
| TRIO-Ex1-F | cggagtcctcgtctatgtgg | MCIDAS-Ex5-F | cctacacccagagtggcatt |
| TRIO-Ex1-R | cgcatggaagcagtggtc | MCIDAS-Ex5-R | gacctcatcttctcccttgc |
| MARCH11-Ex1A-F | aactccgtcccactgatcc | MCIDAS-Ex6-F | tccccagttcctgatgagac |
| MARCH11-Ex1A-R | agctcgtccgctcccctg | MCIDAS-Ex6-R | gtcagtattcgcgggctct |
| MARCH11-Ex1b-F | gcacaaccatgagctttgag | MCIDAS-Ex7-F | aagtggacgccatcctga |
| MARCH11-Ex1b-R | gtggtgctgaaagcttgacc | MCIDAS-Ex7-R | ggtttctaaatccttccctgct |
| BASP1-Ex1-F | gaccccgaaggagagtgag | MCIDAS-3'UTR-F | gtatagcccctcccccttgtc |
| BASP1-Ex1-R | gccttgggtgtggaactagg | MCIDAS-3'UTR-R | caccccagttttatctgtgc |
| GOLPH3-Ex1-F | gccctgacctgtttcctcat | SLC38A9-Ex1-Trns3-F | gggcaggaattcactcagac |
| GOLPH3-Ex1-R | acctacctggagctcacctg | SLC38A9-Ex1-Trns3-R | atgctgcctgtaccccact |
| ZFR-Ex2-F | gcgcctctctctttctctcc | DDX4-Ex1-Trns4-F | cctgcaagctgtttggtcac |
| ZFR-Ex2-R | tacacctccgccgtctctac | DDX4-Ex1-Trns4-R | gattacaggcgtgagccact |
| TARS-Ex4-Trns3-F | tgttgcctaccttcgttgag | SETD9-Ex6-Trns2-F | ctggaaggtctagctaatgcag |
| TARS-Ex4-Trns3-F | tgacagacaatgctaacctcag | SETD9-Ex6-Trns2-R | gggtttgtgctaaatcggtag |
| AMACR-Ex1-F | attgggagggcttcttgc | DEPDC1B-Ex1-F | agggtgggcaacagatttc |
| AMACR-Ex1-R | ggaggcaaaaatctggaagg | DEPDC1B-Ex1-R | ggaagcgaagactgacaagc |

| ***CDC20B gene primers*** | |
|---|---|
| **Name** | **Sequence (5'-3')** |
| CDC20B-Ex1-F | ccaataagaagccggcgag |
| CDC20B-Ex1-R | gaccagggcttagagtttcg |
| CDC20B-Ex2-F | gtgtaatctgacaagcacaag |
| CDC20B-Ex2-R | tcagaccagctcaaacttcct |
| CDC20B-Ex3-F | gggaatttttacttttgcatca |
| CDC20B-Ex3-R | cctttccttttgtttttgtgtca |
| CDC20B-Ex4-F | tgggctttttcactctgtcc |
| CDC20B-Ex4-R | agcaactcttcccttacatttg |
| CDC20B-Ex5-F | ttgtggtgttggctttgaca |
| CDC20B-Ex5-R | agagaatgggacttggaggt |
| CDC20B-Ex6-F | ccccggctctttgtactgta |
| CDC20B-Ex6-R | atttcccacaccatgccaac |
| CDC20B-Ex7-F | ttccattttccagctaacca |
| CDC20B-Ex7-R | cccatacttcaagagccaca |

| ***CDC20B gene primers*** | |
|---|---|
| **Name** | **Sequence (5'-3')** |
| CDC20B-Ex8-F | cagcagtaattggctctcagg |
| CDC20B-Ex8-R | cctgagagggtgggatcata |
| CDC20B-Ex9-F | agaatcctccaggcagaaca |
| CDC20B-Ex9-R | gggttgtctttcaggagagtg |
| CDC20B-Ex10-F | tagatagcaggcccaggaga |
| CDC20B-Ex10-R | agagcttgttgggggcatag |
| CDC20B-Ex11-F | gaggttgggtcatggacagt |
| CDC20B-Ex11-R | gctctctgaggtagggtgga |
| CDC20B-3'UTRA-F | tgtcttctccatggggtttt |
| CDC20B-3'UTRA-R | tgaggaagcacagagcagaa |
| CDC20B-3'UTRb-F | aagcatgcatattgctctctc |
| CDC20B-3'UTRb-R | caaaagacagaacagcgaactg |
| CDC20B-3'UTRC-F | gcccaggacttgaagttagg |
| CDC20B-3'UTRC-R | gggggcatagggaataggta |

| ***CDC20B cDNA primers*** | |
|---|---|
| **Name** | **Sequence (5'-3')** |
| CDC20B-CDNA-1-F | acggaagaggagatgctgtg |
| CDC20B-CDNA-1-R | ctccatctttgcagcctttc |
| CDC20B-CDNA-2-F | gcctcaaaagggcagaaatg |

| ***CDC20B cDNA primers*** | |
|---|---|
| **Name** | **Seqquence (5'-3')** |
| CDC20B-CDNA-2-R | cttgcttgtggcgaagtgt |
| CDC20B-CDNA-3-F | atcctcagcagtgggtcaag |
| CDC20B-CDNA-3-R | aaaacccacctgacctgga |

## A 2.3 Primer pairs used for cloning and site-directed mutagenesis reactions (SDM)

| ***GLH35 luciferase primers*** | |
|---|---|
| **Name** | **Sequence (5'-3')** |
| PRLR-3UTR-LuC-SpeI-F | ctagactagttgggaaactaggtatttgc |
| PRLR-3UTR-LuC-MluI-R | ctgacgcgttgccattctcaaaacattg |
| SNX18-3UTRA-LuC-SpeI-F | ctagactagttggtggtacaaggacgg |

| | |
|---|---|
| SNX18-3UTRA-LuC-MluI-R | ctgacgcgtatgcactggcctttcactc |
| SNX18-3UTRb,C-LuC-SpeI-F | ctagactagtaaacccaagagaaaagcc |
| SNX18-3UTRb,C-LuC-MluI-R | ctgacgcgttcctgatgtcactaatg |
| LMBRD2-3UTR-LuC-SpeI-F | ctagactagtgccaggagcagagataatttg |
| LMBRD2-3UTR-LuC-MluI-R | ctgacgcgttccaaagtaaacccaaag |

| ***GLH135 minigene primers*** | |
|---|---|
| **Name** | **Sequence (5'-3')** |
| IL7R-Ex5-MG-EcoRI-FP | gtgtgaattcttaaagataatacactgtg |

| ***SDM primers*** | |
|---|---|
| **Name** | **Sequence (5'-3')** |
| **CDC20B patient variant SDM primers** | |
| CDC20B-Gln94His-F | ttcctttggggaagagcattcaaccacctacctc |
| CDC20B-Gln94His-R | gaggtaggtggttgaatgctcttccccaaaggaa |
| CDC20B-Arg200Gly-F | gctgcaaagatggagtcggagatgaatccttccat |
| CDC20B-Arg200Gly-R | atggaaggattcatctccgactccatctttgcagc |
| CDC20B-Ile212Lys-F | tggagtatggaatcgtttttatcaccagaacttttaagatggaagg |
| CDC20B-Ile212Lys-R | ccttccatcttaaaagttctggtgataaaaacgattccatactcca |
| CDC20B-Trp363Arg-F | tgtgtgctctgaagcggtcaccggatggc |
| CDC20B-Trp363Arg-R | gccatccggtgaccgcttcagagcacaca |
| CDC20B-Ala509Thr-F | gtgttttctgctgcaactgatgggacggcct |
| CDC20B-Ala509Thr-R | aggccgtcccatcagttgcagcagaaaacac |
| **SDM primers to generate certain truncation variants** | |
| CDC20B-Phe486Ter-F | tgtccaggtcaggtgggttttaaggccacagggg |
| CDC20B-Phe486Ter-R | cccctgtggccttaaaacccacctgacctggaca |
| CDC20B-Tyr98Ter-F | aagagcagtcaaccacctaactcccagaagcttc |
| CDC20B-Tyr98Ter-R | gaagcttctgggagttaggtggttgactgctctt |
| CDC20B-Lys222Ter-F | catactccaaccagaggtgtagattcatattactggtct |
| CDC20B-Lys222Ter-R | agaccagtaatatgaatctacacctctggttggagtatg |
| **Minigene assay SDM primers** | |
| IL7R-Ex5-30var-SDM F | caacacctcttttcccatcttaagaatgtaactgcactc |
| IL7R-Ex5-30var-SDM R | gagtgcagttacattcttaagatgggaaaagaggtgttg |
| **Luciferase assay SDM primers** | |
| PRLR-3'UTR-SDM-F | tctagcagcccctctctgtttttcccttcacttga |
| PRLR-3'UTR-SDM-R | tcaagtgaagggaaaaacagagaggggctgctaga |
| SNX18-3'UTRa-SDM-F | tggctggatacaagcctgctatggatgccttttt |
| SNX18-3'UTRa-SDM-R | aaaaaggcatccatagcaggcttgtatccagcca |
| SNX18-3'UTRb-SDM-F | agttgtttatatgctgtaactaacgcagcgagactggt |
| SNX18-3'UTRb-SDM-R | accagtctcgctgcgttagttacagcatataaacaact |
| SNX18-3'UTRc-SDM-F | aatgttatgattagtgcagatcagccgtgtttatctttttaaaaaataaacagg |
| SNX18-3'UTRc-SDM-R | cctgtttatttttttaaaaagataaacacggctgatctgcactaatcataacatt |
| LMBRD2-3'UTR-SDM-F | tacttgctatttacctgctggtctaccttacaaagccattg |
| LMBRD2-3'UTR-SDM-R | caatggctttgtaaggtagaccagcaggtaaatagcaagta |

| ***CDC20B cloning/sub-cloning primers*** | |
|---|---|
| **Name** | **Sequence (5'-3')** |
| **pcDNA3.1(+)-CDC20B** | |
| CDC20B-NheI-FP | gctggctagcgccaccatggagtggaaactggagcgcac |
| CDC20B-XhoI-RP | tctagactcgagctagtagcaattccatacagagg |
| | |
| **pEGFP-N1-CDC20B** | |

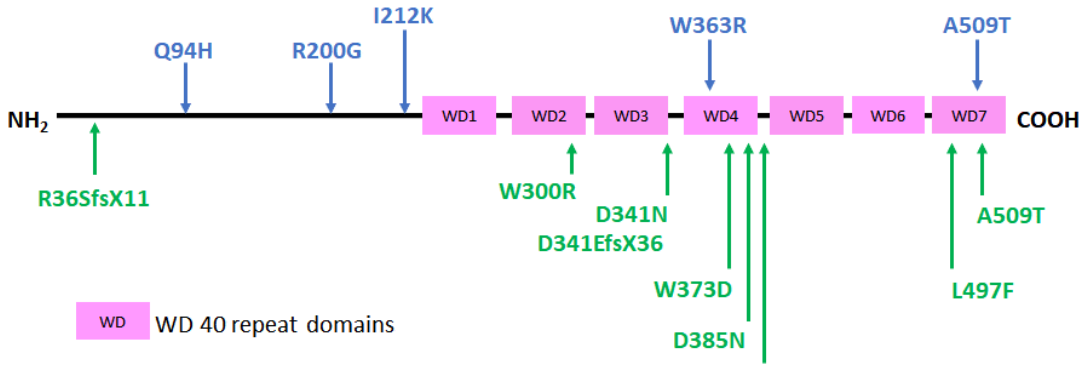
| | |
|---|---|
| CDC20B-XhoI-FP | gctgctcgagcgccaccatggagtggaaactggagcgcac |
| CDC20B-SacII-RP | tctagaccgcgggtagcaattccatacagagg |
| | |
| **pCMV6-AC-GFP (Turbo)** | |
| CDC20B-AsisI-FP | gctggcgatcgcgccaccatggagtggaaactggagcgcac |
| CDC20B-MluI-RP | tctagaacgcgtgtagcaattccatacagagg |
| | |
| **pEGFP-C2-CDC20B** | |
| CDC20B-BglII-FP | gctgagatctgccaccatggagtggaaactggagcgcac |
| CDC20B-SalI-RP | tctagagtcgacctagtagcaattccatacagagg |
| | |
| **pEGFP-C2-CDC20B-2** | |
| CDC20B-XhoI-FP-2 | gctgctcgagcagccaccatggagtggaaactggagcgcac |
| CDC20B-XbaI-RP | ctagtctagactagtagcaattccatacagagg |
| | |
| **p3X-FLAG-CDC20B** | |
| CDC20B-NotI-FP | gctggcggccgccgccaccatggagtggaaactggagcgcac |
| CDC20B-EcoRI-RP | tctagagaattcctagtagcaattccatacagagg |
| | |
| **pCAG-EGFP-C1-CDC20B** | |
| CDC20B-KpnI-RP | cggggtaccctagtagcaattccatacagagg |
| | |
| ***Primers for subcloning CDC20B domains in the p3X-FLAG vector*** | |
| **Reverse primers for cloning only the N-term. of protein in p3X-FLAG** | |
| CDC20B-EcoRI-RP-Nterm | tctagagaattccacctctggttggagtatggaatcgt |
| | |
| **Forward primer for cloning only the C-term. of protein in p3X-FLAG** | |
| CDC20B-NotI-FP-Cterm | gctggcggccgccgccaccatgaagattcatattactggtct |
| | |
| **3x-FLAG-CDC20B-C-term. SDM primer** | |
| 3x-CDC20B-Cterm-ins-SDM-F | gcttgcggccgc**c**gccaccatgaag |
| 3x-CDC20B-Cterm-ins-SDM-R | cttcatggtggcggcggccgcaagc |
| | |
| **Primer to clone CDC20B WD domains 5,6,7 in p3X-FLAG** | |
| CDC20B-NotI-WD5,6,7-FP | gctggcggccgccgccaccatgaaagtcataacccagtctac |
| | |
| ***Primers for subcloning CDC20B domains in the EGFPN1 vector*** | |
| **Reverse primer for cloning first 100 amino acids of protein in EGFPN1 vector** | |
| CDC20B-SacII-100aa-RP | tctagaccgcggtgggaggtaggtggttgac |
| | |
| **Reverse primer for cloning first 395 amino acids of protein in EGFPN1** | |
| CDC20B-SacII-395aa-RP | tctagaccgcggcagcggttggccctgtgcac |
| | |
| **Forward primer for cloning WD-5,6,7 of protein in EGFPN1** | |
| CDC20B-XhoI-396aa-FP | gctgctcgagcgccaccatgaaagtcataacccagtctac |
| | |
| **Forward primer for cloning 101 to 519 aa of protein in EGFPN1** | |
| CDC20B-XhoI-101aa-FP | gctgctcgagcgccaccatggaagcttccggatcagtgc |
| | |
| **Reverse primer for cloning only the N-term of protein in EGFPN1** | |
| CDC20B-SacII-RP-Nterm | tctagaccgcggcacctctggttggagtatggaatcgt |
| | |

| Forward primer for cloning only the C-term of protein in EGFPN1 | |
|---|---|
| CDC20B-XhoI-FP-Cterm | gctgctcgagcgccaccatgaagattcatattactggtct |

### A 2.4 *CDC20B* rare pathogenic variants identified in the Epi25 consortium study

| Chr position (Grch38) | cDNA position | Amino acid position | Consequence | Allele count Case | Allele number Case | MAF gnomAD v2 |
|---|---|---|---|---|---|---|
| 54468431 | c.106_110delAGAAG | p.Arg36SerfsTer11 | loss of function | 2 | 18340 | 0.003 |
| 54442450 | c.355+6G>A | _ | splice region | 1 | 18326 | _ |
| 54423176 | c.898T>C | p.Trp300Arg | missense | 1 | 18338 | 0.00001 |
| 54423079 | c.989+6A>C | _ | splice region | 4 | 18338 | 0.00001 |
| 54420825 | c.1021G>A | p.Asp341Asn | missense | 1 | 18340 | 0.00003 |
| 54420823 | c.1022dupA | p.Asp341GlufsTer36 | loss of function | 2 | 18340 | 0.000003 |
| 54420728 | c.1118G>A | p.Gly373Asp | missense | 1 | 18340 | _ |
| 54420693 | c.1153G>A | p.Asp385Asn | missense | 5 | 18340 | 0.00001 |
| 54420665 | c.1181C>T | p.Pro394Leu | missense | 1 | 18340 | 0.00004 |
| 54416383 | c.1216-5T>C | _ | splice region | 1 | 18338 | 0.00001 |
| 54410152 | c.1460-6C>T | _ | splice region | 1 | 18334 | 0.000004 |
| 54410115 | c.1491G>T | p.Leu497Phe | missense | 1 | 18336 | 0.000004 |
| 54410081 | c.1525G>A | p.Ala509Thr | missense | 6 | 18336 | 0.0002 |

### A 2.5 CDC20B protein schematic with the position of the patient variants




***CDC20B* pathogenic variants in epilepsy cases:** Variants identified in our study are represented on top in blue. Variants reported in the Epi25 consortium study have been represented in green (https://epi25.broadinstitute.org/).

## A 2.6 CDC20B protein sequence (NCBI, FASTA format)

CDC20B protein [Homo sapiens] – Human
GenBank: AAI36573.1

>gi|223459660|gb|AAI36573.1| CDC20B protein [Homo sapiens]
MEWKLERTAPRRVRTEEEMLWESIMRVLSKDLKQKRSQDSANVLDSVNATYSDFKSNFAKRLSAEVPVAS
SPITTRWQQSQTRALSSDSFGEEQSTTYLPEASGSVLKTPPEKETLTLGSCKEQLKTPSKGISETSNSAL
HFCKAPHAMDRDWKESVASKGQKCLKQLFVTQNVVQQANGKMQLCEQSECVWKGCKDGVRDESFHLKSSG
DINDSILQPEVKIHITGLRNDYYLNILDWSFQNLVAIALGSAVYIWNGENHNGIENIDLSLTCNYISSVS
WIKEGTCLAVGTSEGEVQLWDVVTKKRLRNMLGHLSVVGALSWNHFILSSGSRLGRVYHHDVRVAQHHVG
TLRHKQAVCALKWSPDGRLLSSGCSDGLLTIWPHDPGASAQGQPLKVITQSTAVKAMDWCPWQSGVLAIG
GGMKDGRLHILDINAGKSIQTPSTNSQICSLIWLPKTKEIATGQGTPKNDVTVWTCPTVSRSGGFFGHRG
RVLHLALSPDQTWVFSAAADGTASVWNCY

PREDICTED: cell division cycle protein 20 homolog B isoform X1 [Macaca fascicularis] - Crab-eating macaque

NCBI Reference Sequence: XP_005556954.1

>gi|544437150|ref|XP_005556954.1| PREDICTED: cell division cycle protein 20 homolog B isoform X1 [Macaca fascicularis]
MEWKLERTAPRRVRTEEEMLWESIMRVLSKELKQKKSQSSTKVFDSVNATYSDFKSHFAKRLSAEVPVAS
SPITTRWRQSQTRALSSDSFGEEQSTTYLPEASGSVLKTTPEKETWTPGSCKEPPKAPSKGISETSNSAL
HFCKAPHTMDRDWKENVASKGQKCLKQLFVTQKVVQQANGKMQLCEQSQCVWKGCKDGVRDESFHLKRSS
DINDSILQPEVKIHLTGLRNDYYLNILDWSFQNLVAIALGSAVYIWNGENCNGIETIDLSLTCNYISSVS
WIKEGTCLAVGTSEGEVQLWDVVTKKRLRNMLGHLSVVGALSWNHFILSSGSRLGRVYHHDVRVAQHHVG
TLRHKQAVCALKWSPDGRLLSSGCSDGLLTIWPHDPGASAQGQPLKVIPQSTAVKAMDWCPWQSGVLAIG
GGMKDGCLHILDINAGKSIQTPSTNSQICSLIWLPKTKEIATGLGTPKNDVTVWTCPTVSRSGGFFGHRG
RVLHLALSPDQTRVFSAAADGTASVWNCY

PREDICTED: cell division cycle protein 20 homolog B isoform 2 [Pan troglodytes] – Chimpanzee

NCBI Reference Sequence: XP_001147536.1

>gi|114600277|ref|XP_001147536.1| PREDICTED: cell division cycle protein 20 homolog B isoform 2 [Pan troglodytes]
MDWKLERTAPRRVRTEEEVLWESIMRVLSKDLKQKRSQDSANVLDSVNATYSDFKSNFAKRLSAEVPVAS
SPITTRWQQSQTRALSSDSFGEEQSTTYLPEASGSVLKTPPEKETLTLESCKEQLKTPSKGISETSNSAL
RFCKAPHAMDRDWKESVASKGQKCLKQLFVTQNVVQQANGKMQLCEQSQCVWKGCKDGVRDESFHLKRSG
DINDSILQPEVKIHITGLRNDYYLNILDWSFQNLVAIALGSAVYIWNGENDNGIENIDLSLTCNYISSVS
WIKEGTCLAVGTSEGEVQLWDVVTKKRLRNMLGHLSVVGALSWNHFILSSGSRLGRVYHHDVRVAQHHVG
TLRHKQAVCALKWSPDGRLLSSGCSDGLLTIWPHDPGASAQGQPLKVIPQSTAVKAMDWCPWQSGVLAIG
GGMKDGCLHILDINAGKSIQTPSTNSQICSLIWLPKTKEIATGQGTPKNDVTVWTCPTVSRSGHRGRVLH
LALSPDQTRVFSAAADGTASVWNCY

PREDICTED: cell division cycle protein 20 homolog B isoform 1 [Gorilla gorilla gorilla] - Western lowland gorilla

NCBI Reference Sequence: XP_004058891.1

>gi|426384693|ref|XP_004058891.1| PREDICTED: cell division cycle protein 20 homolog B isoform 1 [Gorilla gorilla gorilla]
MEWKLERTATRRVRTEEEMLWESIMRVLSKDLKQKRSQDSANVLDSVNATYSDFKSNFAKRLSAEVPVAS
SPITTRWQQSQTRALSSDSFGEEQSTTYLPEASGSVLKTPPEKETLTLGSCKEQLKTPSKGISETSNSAP
HFCKAPHAMDRDWKESVASKGQKCLKQLFVTQNVVQQANGKMQLCEQSQCVWKGCKDGVRDESFHLKRSG
DINDSILQPEVKIHITGLRNDYYLNILDWSFQNLVAIALGSAVYIWNGENHNGIENIDLSLTCNYISSVS
WIKEGTCLAVGTSEGEVQLWDVVTKKQLRNMLGHLSVVGALSWNHFILSSGSRLGRVYHHDVRVAQHHVG
TLCHKQAVCALKWSPDGGLLSSGCSDGLLTIWPHDPGASAQGQPLKVIPQSTAVKAMDWCPWQSGVLAIG
GGMKDGRLHILDINAGKSIQTPSTNSQICSLIWLPKTKEIATGQGTPKNDVTVWTCPTVSRSGHRGRVLH
LALSPDQTRVFSAAADGTASVWNCY

PREDICTED: cell division cycle protein 20 homolog B isoform [Macaca mulatta] - Rhesus macaqueNCBI Reference Sequence: XP_001097843.2

>gi|297294276|ref|XP_001097843.2| PREDICTED: cell division cycle protein 20 homolog B isoform 2 [Macaca mulatta]
MEWKLERTAPRRVRTEEEMLWESIMRVLSKELKQKKSQSSTKVFDSVNATYSDFKSHFAKRLSAEVPVAS
SPITTRWRQSQTRALSSDSFGEEQSTTYLPEASGSVLKTTPEKETWTPGSCKEPPKTPSKGISETSNSAL
HFCKAPHTMDRDWKENVASKGQKCLKQLFVTQKVVQQANGKMQLCEQSQCVWKGCKDGVRDESFHLKRSS
DIKESILQPEVKIHLTGLRNDYYLNILDWSFQNLVAIALGSAVYIWNGENCNGIETIDLSLTCNYISSVS
WIKEGTCLAVGTSEGEVQLWDVVTKKRLRNMLGHLSVVGALSWNHFILSSGSRLGRVYHHDVRVAQHHVG
TLRHKQAVCALKWSPDGRLLSSGCSDGLLTIWPHDPGASAQGQPLKVIPRSTAVKAMDWCPWQSGVLAIG
GGMKDGCLHILDIDAGKSIQTPSTNSQICSLIWLPKTKEIATGLGTPKNDVTVWTCPTVSRSGHRGRVLH
LALSPDQTRVFSAAADGTASVWNCY

PREDICTED: cell division cycle protein 20 homolog B isoform 1 [Pan paniscus] - Bonobo

NCBI Reference Sequence: XP_003827408.1

>gi|397514257|ref|XP_003827408.1| PREDICTED: cell division cycle protein 20 homolog B isoform 1 [Pan paniscus]
MDWKLERTAPRRVRTEEEVLWESIMRVLSKDLKQKRSQDSANVLDSVNATYSDFKSNFAKRLSAEVPVAS
SPITTRWQQSQTRALSSDSSGEEQSTTYLPEASGSVLKTPPEKETLTLESCKEQLKTPSKGISETSNSAL
RFCKAPHAMDKDWKESVASKGQKCLKQLFVTQNVVQQANGKMQLCEQSQCVWKGCKDGVRDESFHLKRSG
DINDSILQPEVKIHITGLRNDYYLNILDWSFQNLVAIALGSAVYIWNGENHNGIENIDLSLTCNYISSVS
WIKEGTCLAVGTSEGEVQLWDVVTKKRLRNMLGHLSVVGALSWNHFILSSGSRLGRVYHHDVRVAQHHVG
TLRHKQAVCALKWSPDGRLLSSGCSDGLLTIWPHDPGASAQGQPLKVIPQSTAVKAMDWCPWQSGVLAIG
GGMKDGRLHILDINAGKSIQTPSTNSQICSLIWLPKTKEIATGQGTPKNDVTVWTCPTVSRSGHRGRVLH
LALSPDQTQVFSAAADGTASVWNCY

PREDICTED: cell division cycle protein 20 homolog B isoform 1 [Papioanubis] – Olive Baboon

NCBI Reference Sequence: XP_003899721.1

>gi|402871548|ref|XP_003899721.1| PREDICTED: cell division cycle protein 20 homolog B isoform 1 [Papioanubis]
MEWKLERTAPRRVHTEEEMLWESIMRVLSKELKQKKSQSSTKVFDSVNATYSDFKSHFAKRLSAEVPVAS
SPITTRWRQSQTRALSSDSFGEEQSTTYLPEASGSVLKTTPEKETWTLGSCKEPPKTPSKGISETSNSAL
HFCKAPHAMDRDWKENVASKGQKCLKQLFVTQKVVQQANGKMQLCEQSQCVWKGCKDGVRDESFHLKRSS
DINDSILQPEVKIHLTGLRNDYYLNILDWSFQNLVAIALDSAVYIWNGENCNGIETIDLSLTCNYISSVS
WIKEGTCLAVGTSEGEVQLWDVVTKKRLRNMLGHLSVVGALSWNHFILSSGSRLGRVYHHDVRVAQHHVG
TLRHKQAVCALKWSPDGRLLSSGCSDGLLTIWPHDPGASAQGQPLKVIPQSTAVKAMDWCPWQSGVLAIG
GGMKDGCLHILDINAGNSIQTPSTNSQICSLIWLPKTKEIATGLGTPKNDVTVWTCPTVSRSGHRGRVLH
LALSPDQTRVFSAAADGTASVWNCY

PREDICTED: cell division cycle protein 20 homolog B isoform 1 [Callithrix jacchus] – Common Marmoset – New Age Monkey

NCBI Reference Sequence: XP_002745017.1

>gi|296194591|ref|XP_002745017.1| PREDICTED: cell division cycle protein 20 homolog B isoform 1 [Callithrix jacchus]
MEWKLERTAPRRVRTEEEVLWESIMRVLAKDLKQKRSQGSAKVFDSVSATYFDFKSNFAKRLSAEVPVAS
SPITTRWRQSQTRALSADSFGEEESITYLPEASGSVLKTTPEKDTLTPGSGKEPPKTPIKGSSETNNSAL
HFCKAPHAMDRHWKENVAPKGQKCLKQLFVTQKVVQQANGKMQLCEQSQCVWKGCKDGVRDESFHLKRSS
DINYSILQPEVKIHLTGLRNDYYLNLLDWNYQNLVAIALGSAVYIWNGENRNVIENIDLSLTCNYISSVS
WIKDGTCLAVGTSEGEVQLWDAVTKKQLRNMLGHLSVVGSLSWNHFILSSGSRLGHVYHHDVRVAQHHVG
TLHHKQAVCALKWSPDGRLLSSGCSDGLLTIWPHDPGASAHGQPLKVITQSTAVKAMDWCPWQSGVLAIG
GGMKDGRLHILDINAGKSIQTPSTNSQICSLIWLPKTKEITTGQGAPKNDVTVWTCPTLSRSGHRGRVLH
LALSPDQTQVFSAAADGTASIWSCY

PREDICTED: cell division cycle protein 20 homolog B isoform 1 [Saimiri boliviensisboliviensis] – Black Capped Monkey

NCBI Reference Sequence: XP_003925904.1

>gi|403267587|ref|XP_003925904.1| PREDICTED: cell division cycle protein 20 homolog B isoform 1 [Saimiri boliviensisboliviensis]
MEWKLERTAPRRVLTEEEMLWESIMRVLAKDLKQKRSQGSAKVFDSVNATYFDFKTNFAKRLSAEVPAAS
SPITTRWRQSQARALSADSFGEEESITYLPEASGSVLKTTPEKETLTLGSCKEPPKTPIKGSSETNNSAL
HFCKAPHAMDRDWKENVASKGQKCLKQLFVTQKVVQQANGKMQLCEQSQCAWKGCKDGVRGESFHLKRSS
DINYSILQPEVKIHLTGLRNDYYLNILDWNYQNLVAIALGSAVYIWNGKNHNGIENIDLSLTCNYISSVS
WIKEGTCLAVGTSEGEVQLWDVVTKKRLRNVLGHLSVVGTLSWNHFILSSGSRLGHVYHHDIRVAQHHVG
TLRHKQAVCALKWSPDGRLLSSGCSDGLLTIWPHDPGASAHGEPLKVITQSTAVKAMDWCPWQSGVLAIG
GGMKDGRLHILDINTGKSIQTPSTNSQICSLIWLPKTKEITTGQGTPKNDVTVWTCPTLSRSGHRGRVLH
LALSPDQTRVFSAAADGTASVWSCY

cell division cycle protein 20 homolog B [Mus musculus]

NCBI Reference Sequence: NP_001268416.1

>gi|527498264|ref|NP_001268416.1| cell division cycle protein 20 homolog B [Mus musculus]

MEWKLQRTARRKIRTEEEMLWENIMRVLANGMKQQRNQGSPKELDSVAVTYSSFKSNFVKRLSAEIPVAS
SPITTRWQLSPARDPESSSSVEEGPPSHTPESLASGLKITPAADTLTLRSHNKNSPKTLSKGSSEVNNST
LRFCKTPLAGDRGWKENLATKGQRCLNQPFSTQKGAQQIDGKMHLCEESRCVRTGCRFGARDEFYLRRFS
GVYHSTCQPEVKIHLTGLRNDYYLNTLDWSSQNLVAVALGTSVYIWNGQNHSWIENIDLSVCCHYVSSVT
WMREGSCLAVGTSEGEVQLWDAITKKQLRNLHGHLSVVGALSWNHCTLSSGSRLGRVHHHDVRVAQHRVG
TLYHKEAVCSLKWSPDGRLLSSGCNDGLLTIWPHDPGAGVQGLPLKVIPQSTAVKAMEWCPWQSEVLAVG
GGVKDGCLHVLDINTGKNIQTPSTQSQICSLIWLPKTKEIATGQGAPKNDVALWTCPTLFRSGGFFGHRD
RVLHLSLSPDQTRLFSAAADGTACVWKCC

PREDICTED: cell division cycle protein 20 homolog B isoform X1 [Chinchilla lanigera] – Chinchilla - Rodent

NCBI Reference Sequence: XP_005392765.1

>gi|533157617|ref|XP_005392765.1| PREDICTED: cell division cycle protein 20 homolog B isoform X1 [Chinchilla lanigera]
MEWKLERTARRRIRTEEEMLWENVMRVLAQNMKQRRSPGSPKVLDSVTTPYSSFKSTIVKRLSAEVPVAS
SPLTTRWQQSQIRDSETYSFGDEHSVSNISESVGTVLKITPEKASLTLGSYKEIPKTPVEGSSETNNSAL
HFCTASHALDRDGKENVSKGQKCLNQVFSAQKVIQKINGKMQLCEESQCIWKGCRDGVRDESFHLQKVRD
VNCPTWQPEVKIHLTGLRNDYYLNILDWSFQNLVAIALGSTVYIWNGENHSEVENIDLSLTCNYISSVSW
IKKGTCLAVGTSEGEVQLWDVVTKKRLRNMQGHLSVVGALSWNHCILSSGSRLGRVYHHDVRAAQHQVGT
LCHKQAVCALKWSPDGRLLSSGCSDGLLTIWPHDLGAHAQGQSVKVISQSTAVKAIDWCPWQSKVLAIGG
GMKDGCLHVLDISSGKSIQTPSTNSQICSLLWLPKTQEIATGQGAPQNDVTLWTCPTLSRSSGFLGHKGR
VLHLALSPDQTRVFSAAADGTACVWKYRRSPSAS

PREDICTED: cell division cycle protein 20 homolog B [Heterocephalusglaber] – Naked Mole Rat - Rodent

NCBI Reference Sequence: XP_004848735.1

>gi|512973274|ref|XP_004848735.1| PREDICTED: cell division cycle protein 20 homolog B [Heterocephalusglaber]
MEWKLERTARRRIRTEEEMLWENVMRVLAEDMKQQRSRGSPKLFDAGTTPYSNFKSNIVKRLSDEVPVAS
SPLSTRWQQSQIRDLEAYSFDESDISESVRSVLKIMPEKGTLTLGSYKEPLKTPVKGSSQTSNSALHFCK
ASRALERGGKENVASKAQNCLNELFSAQKVIQQINGKMQLCEESQCIWEGCRDGVRNESFHLKKFRDVNY
PICQPEVKIHLTGLRNDYYLNVLDWSFQNLVAVALGSAVYIWNGENHRVEKIDLSLTCNYVSSVSWIKKG
TCLAVGTSEGEVQLWDVVTEKRLRNMQGHLSVVGALSWNHYILSSGSRLGRVYHHDVRAAQHQVGTLCHK
QAVCALKWSPDGRLLSSGCSDGLLTIWPHDPGAHAQGQSLKVISQSTAVKAIDWCPWKSEVLAIGGGMKD
GCLHILDINSGKSIQTPSTNSQICSLIWLPKTKEIATGQGAPKNDVILWTCPTLSRPSGFLGHRGRVLHL
ALSPDQTRVFSAAADGTACVWKCHQSPRPS

PREDICTED: cell division cycle protein 20 homolog B [Ictidomystridecemlineatus] – Thirteen Lined Ground Squirrel

NCBI Reference Sequence: XP_005319731.1

>gi|532067774|ref|XP_005319731.1| PREDICTED: cell division cycle protein 20 homolog B [Ictidomystridecemlineatus]
MEWKLERTARRRIRTEEEMLWENVMKVLSNDMKQKRSQVSPKVLDSVTTTYSNFKSNFLKRLSAELPVAS

SPIATRWQQRQTGDLACSFGAELSTTDIPESLQSVLGTPEKDTLTPGIYKEPLKFPTKEGYERNNSALCV
CEAPHALDRAVKENVASKSQKCPNQLFSSQKVVQKTDGKMQLCDKSQCIWKGCQYGVGDELFHLKRFCDK
NYPICQPEVKIHLTGLRNDYYLNILDWSFKNVVAVALGSAVYIWNGENHNRVENIDLGLTCNYVSSVSWI
KEGTCLAVGTSDGEVQLWDVVTKKRLRNMLGHLSVVGALSWNHYLLSSGSRLGHVYHHDVRIAQHHVGTL
CHKQAVCALKWSLDGRLLSSGCSDGLLTIWPYDPGASARGQPLKVLPQSTAVKAMDWCPWQSAVLAVGGG
MKDGCLHILDINTGKSIQTPSTNSQICSLIWLPKTKEIATGQGTPKNDVTLWTCPTLSRSGGFFGHRGRV
LHLALSPDQTQVFSAAADGTATVWKCCQSPNPS

PREDICTED: LOW QUALITY PROTEIN: cell division cycle 20B [Mesocricetus auratus] – Golden Hamster

NCBI Reference Sequence: XP_005139343.1

>gi|524922544|ref|XP_005139343.1| PREDICTED: LOW QUALITY PROTEIN: cell division cycle 20B [Mesocricetus auratus]
MVVSSSLITTRWQYPATPSPPDGSVQQPHHHQMAVSSSPITTRWQPNQARDLEACSSAKEGYSSNIPESP
SVLKITPAVDTQTLHESNFRLAVNLNILDWNSQNLVAITHRSSIHIIWNGQNHSRNENIDLSLFSIYASS
VPWVREGSCLAVGISDGEAQVLGHLLTMFLSRLGVVHHYDVWMAQHYVETLHQKEAVCALKWSRDGRLLP
GGCKDGLLAIWLHDPGASVXGPALKVIPQFTAVKVCSLIXVPKTKXIATGQGASKNDMALWTCSSLNCIS
LCDPSLGTHPKGLALGEYCNGISWLSESAEDKANDSGASVVV

PREDICTED: cell division cycle protein 20 homolog B [Cavia porcellus] – Guinea Pig

NCBI Reference Sequence: XP_003470297.1

>gi|348569022|ref|XP_003470297.1| PREDICTED: cell division cycle protein 20 homolog B [Cavia porcellus]
MEWKLERTARRRIRTEEEMLWENIMRVLAHNMKQQRSRSSPKVFDSVTTPYSNFKENIVKRLSAEVPIAS
SPLTTRWQQSQSKDLETYSFGDEYSVTDISQSVGTVLKIIPEQGTLTLGSHKEPLKTPVKGSSESNNSSL
QFCKASHALERGGKENVASKGQKCLNQLFSAQKVIQQINGKMQLCQESQCVWKGCRDGVRDKSFHLKKLR
DVNYPIWQPEVKIHLTGLRNDYYLNILDWSFQNLVAVALGSAVYIWNGENHNGVEKIDLSLTCNYVSSVS
WIKKGTCLAIGTSEGEVQLWDVVAKKRLRSMQGHLSVVGSLSWNHYILSSGSRLGRVYHHDVRAAQHQVG
TLCHRKAVCALKWSPDGRLFSTGCSDGLLTLWPHDPGTHAQGQPLKVISHLTAIKAIDWCPWQSEVLAIG
GGMKDGCLHILDISSGKSLQTPSINSQICSLIWLPKTKEIATGQGAPQNDVTLWSCPTLSRASVFLGHRG
RVLHLALSPDQTRVFSAAADGTACVWKCHQSPSSS

PREDICTED: LOW QUALITY PROTEIN: cell division cycle 20B [Octodon degus] - Degu

NCBI Reference Sequence: XP_004649047.1

>gi|507615600|ref|XP_004649047.1| PREDICTED: LOW QUALITY PROTEIN: cell division cycle 20B [Octodon degus]
MVEAQEPPVKALAMERNSHNLGIFYFCIVEATADLKSLCCPAGAGFRYQPALALGERRRPDAVPARRQGA
ATSSRGLGSRLRRGPARRRQPLGACGDSADEDARDRSSAPGEPGCEAQRPTRPRRSSGGTLGVRSAASRL
NVSLVGSYTELLKTPVKGSSEAHHSGLHFCKAYRALDRXGKESIAFKGQKCLKQLLAAQKVVQQXQGNMQ
LCEDSQCAWEVLNILDWSFQNLVAIALGSTVCIWNGEDHSGVENVDLSLTCSYISSVSWIRQGTXLAVGT
SEGEVRLWDVVTKKRLCNLRGHLSVVGALSWNYRVLRSGSGLGRVYHHEVRAAQHRVSPLPXQQAVCALK
WSPDGRLLSSGCSGGLLQYVWPQDPGCSAQGADLSMLLSETLKALDWCPWKPEVLAVGGGMEDGCLHVLD
ISSGKSMQTPSTNSQICSLIXLPKTMEIATGQGAPWNDVALWTCPTLCRSSGFLSRDVQEHGTDMRWGLC

AVPSSAEGRSAGDTFELPHSRWSRPHPGRPASRVFHRILPWFTPCSLLPTHTCVYAGQCTVKALSILCL

PREDICTED: cell division cycle protein 20 homolog B isoform X1 [Capra hircus] - Goat
NCBI Reference Sequence: XP_005694763.1

>gi|548511825|ref|XP_005694763.1| PREDICTED: cell division cycle protein 20 homolog B isoform X1 [Capra hircus]
MVVLGDKTPEVCRLRVKTGLFLGNKTKSGRGEATLEPVGQTPPEMEWKLERAARRRVRTEEEILWEDVMR
VLDKDMKQRRGQVSPKVFNVVNTTYSNFTSHLAKRLSAEVPVASSPITARWRQSQAGDVAAPSFGEERST
TDLPETVLKMTPEKETLTPGSCKEPLKTPNKESSGTNNSAFQFGKAPHAVDRGRKDKVASKDLKCLNQLF
STQNVAQQVNGSMQFCEGSQCTWKGCSNGIRDGYFQRLSDINYPILQPEVKIHLTGLRNDYYLNILDWNF
QSLVAIALGSSVYIWNGDNHHRTENMCFSLPCNYVSSVSWMTEGTCLAVGTSEGEIQLWDVVTKKRLRNM
LGHLSVVGALSWNHYILSSGSRLGRVYHHDIRVAQHHVGTLQHKQAVCGLKWAPGGRLLSSGCSDGLLTI
WPHDPGANAQSHPLKVIHQSTAVKAIDWCPWQTEVLAVGGGMKDGRLHILDINTGQSIQTPSTDSQICSL
TWLPKTKEIATGQGSPKNDVTVWTCPGLARSRGFFDHRGRVLHLALSPDQTRVFSAAADGTACIWNCC

PREDICTED: cell division cycle protein 20 homolog B isoform X3 [Canis lupus familiaris] - Dog
NCBI Reference Sequence: XP_535236.3

>gi|345793816|ref|XP_535236.3| PREDICTED: cell division cycle protein 20 homolog B isoform X3 [Canis lupus familiaris]
MEWKLERTARRRVRTEDEMLWENVMRVLAKDLRQKRSPGSPKIFDEMKITYSSFKSNFVKRLSAEVPVAS
SPIATRWRQNQARDPAACSFGEEFSTTYLPEASGSVLEKTPEKETLVLGSYQEPLKTPIKGSFETNNSAL
RFCKAPHALDRGWNENIASKGQKCLNQLFLTQKVVQWVSGQMQLCEPSQCVWKGCRNGVGDESFHLKRFG
DIDYSILQPEVKIHLSGLRNDYYLNILDWNFRNLVAVALGSSVFIWTGENNVIENIDLSLNCSYISSVSW
IKDGNCLAVGTSEGEVQLWDVVTKKRLRNMLGHLSVVGALSWNHCILSSGSRLGRVYHHDVREAQHHVGS
LHHKQAVCALKWSPDGRLLSSGCSDGLLAIWPHDPGVRTQTQPLKVISQPTAVKAMDWCPWQSAVLAVGG
GMRDGHLRILDINTGRSIQTPSTNSQICSLLWLPKTKEIATGQGTPQNDVTTWACPGLARSGGFFGHRGR
VLHLALSPDQTRVFSAAADGTACVWNCCRSALL

cell division cycle protein 20 homolog B [Bos taurus] - Cow
NCBI Reference Sequence: NP_001193911.1

>gi|332801029|ref|NP_001193911.1| cell division cycle protein 20 homolog B [Bos taurus]
MEWKLERAARRRVRTEEEILWEDVMRVLDKDMKQRRGQVSPKVFNVVNTTYSNFTSNLAKRLSAEVPVAS
SPITTRWRQSQAGDVTALSFGEEHSTADLPETVLKMTPEKETLTPGSYKEPLKTPNKGSSGTNNSAFQFG
KAPHALNRGWKDKVASKDPKCLNQLFSTQNVAQQVNGSMQFCERSQCTWKGCRDGVRDGYFQRFSDINYS
ILQPEVKIHLTGLRNDYYLNILDWNFQNLIAIALGSSVYVWNGDNHNRIENMYFSLPCNYVSSVSWMTEG
TCLAVGTSEGEIQLWDLVTKKRLRNMLGHLSVVGALSWNHYILSSGSRLGRVYHHDVRVAQHHVGTLQHT
QAVCGLKWAPGGRLLSSGCSDGLLTIWPYDPGANAQGPPLKVIHQSTAVKAIDWCPWQTEVLAVGGGMKD
GRLHILDISTGQSIQTPSTDSQICSLTWLPKTKEIATGQGSPKNDVTMWTCPGLARSRGFFDHRGRVLHL
ALSPDQMRVFSAAADGTACIWNCC

PREDICTED: cell division cycle protein 20 homolog B [Pantholopshodgsonii] – Tibetan Antelope

NCBI Reference Sequence: XP_005958424.1

>gi|556725277|ref|XP_005958424.1| PREDICTED: cell division cycle protein 20 homolog B [Pantholopshodgsonii]

MEWKLERAARRRVRTEEEILWEDVMRVLDKDMKQRRGQVSPKVFNVVNTTYSNFTSNLAKRLSAEVPVAS
SPITTRWRQSQAGDVTAPSFGAERSTTDLPETVLKMTPEKETLTPGSCKEPLKTPNKESSGTNNSAFQFG
KAPHAVDRGWKDKVASKDPKCLNQLFSTQNVAQQVNGSMQFCERSQCTWKGCSDGIRDGYFQRFSDINYP
ILQPEVKIHLTGLRNDYYLNILDWNFQSLVAIALGSSVYIWNGDNHHRIENMCFTLPCNYVSSVSWMTEG
TCLAVGTSEGEIQLWDVVTKKRLRNMLGHLSVIGALSWNHYILSSGSRLGRVYHHDVRVPQHHVGTLQHK
QAVCGLKWAPGGRLLSSGCSDGLLTIWPHDPGANAQSHPLKVIHQSTAVKAIDWCPWQTEVLAVGGGMKD
GRLHILDINTGQSIQTPSTDSQICSLTWLPKTKEIATGQGSPKNDVTVWTCPSLARSRGYFDHRGRVLHL
ALSPDQTRVCSAAADGTACIWNCC

PREDICTED: cell division cycle protein 20 homolog B [Bos mutus] - Yak

NCBI Reference Sequence: XP_005887437.1

>gi|555950469|ref|XP_005887437.1| PREDICTED: cell division cycle protein 20 homolog B [Bos mutus]

MEWKLERAARRRVRTEEEILWEDVMRVLDKDMKQRRGQVSPKVFNVVNTTYSNFTSNLAKRLSAEVPVVS
SPITTRWRQSQAGDVTALSFGEEHSTADLPETVLKMTPEKETLTPGSYKEPLKTPNKGSSGTNNSAFQFG
KAPHALNRGWKDKVASKDPKCLNQLFSTQNVAQQVNGSMQFCERSQCTWKGCRDGVRDGYFQRFSDINYS
ILQPEVKIHLTGLRNDYYLNILDWNFQNLIAIALGSSVYIWNGDNHNRIENMYFSLPCNYVSSVSWMTEG
TCLAVGTSEGEIQLWDVVTKKRLRNMLGHLSVVGALSWNHYILSSGSRLGRVYHHDVRVAQHHVGTLQHT
QAVCGLKWAPGGRLLSSGCSDGLLTIWPYDPGANAQGPPLKVIHQSTAVKAIDWCPWQTEVLAVGGGMKD
GRLHILDISTGQSIQTPSTDSQICSLTWLPKTKEIATGQGSPKNDVTMWTCPGLARSRGFFDHRGRVLHL
ALSPDQMRVFSAAADGTACIWNCC

PREDICTED: cell division cycle protein 20 homolog B [Sus scrofa] - Pig

NCBI Reference Sequence: XP_005674286.1

>gi|545889948|ref|XP_005674286.1| PREDICTED: cell division cycle protein 20 homolog B [Sus scrofa]

MRVLAKDMKQKRGQGSPKVFDVVNTTYSNFKSNFAKRLSAEVPVASSPITTRWRQSQTRDVAAHSFGEEH
STTDLAETVLRMTPEKETLSPGSYKEPLKNPNKGSFETSNSAFHFCKASHALDRGWKENVASTGQKCLNQ
LFSTQKVAQQVNGKMQLCEQPQCIWKGCNDGVKDGYFYIQSFSDTNYSILQPEVKIHLXXFRNGYYLNIL
DWNFQNLVAIALGSSVHIWNGESHSGIENIDLSLTCNYVSSVSWIEKGNCLAVGTSEGEVQLWDVVTKKR
LRNMLGHLSVVGALSWNHCILSSGSRLGRVYHHDVRVAQHRVGTLHHKQAVCALKWSPDGRLLSSGCSDG
LLTIWPHDPGVTAQGHALKVIPQPTAVKAVDWCPWKSEVLAVGGGMKDGHLHILDINTGQSIQTPSTNSQ
ICSLVWLPKTKEIATGQGSPKNDVTVWACPALARSRGFFGHRGRVLHLALSPDQTKVFSAAADGIACVWN
CC

PREDICTED: cell division cycle protein 20 homolog B [Equus caballus] - Horse

NCBI Reference Sequence: XP_001494074.2

>gi|338718862|ref|XP_001494074.2| PREDICTED: cell division cycle protein 20 homolog B [Equus caballus]
MEWKLERTAPRRVRTEEEMLWENVMRVLAKDMKQKRSQGSSKVFDAVNTTYSNFKSNFVKRLSAEVPVAS
SPVTTRWRQSQTRDLAAYSFEEEHSTTELPEASASVLKITPENQTLILGSYEEPRNTPIKGNFETNYSAL
RFGKAPHALDRGWKENIASKGQKCLNQLFSTQKVAHQVNGKMQLCEQSQCVWKGCRDGVRDESFYLKRFS
DINSSILQPEVKIHLTGLRNDYYLNILDWNFQNLVAIALGSSVCIWNGENHNRIENMDLNLTCNYVSSVS
WIKEGSCLAVGTSEGEVQLWDVVTKKRLRNMLGHLSVVGALSWNHCILSSGSRLGRVYHHDVRVAQHHVG
TLHHKQAVCALKWSFDGRLLSSGCSDGLLTIWPHDPGVSAQGQPLKVIPQPTAVKAMNWCPWQSAVLAVG
GGMKDGHLHILDINTWRSIQTPSTNSQICSLIWLPKTKEIATGQGTPKNDVTVWACPALAKSGGFFGHRG
RVLHLALSPDQTRVFSAAADGTACVWNCC

PREDICTED: cell division cycle protein 20 homolog B [Ceratotheriumsimumsimum] - White rhinoceros
NCBI Reference Sequence: XP_004422869.1
>gi|478497061|ref|XP_004422869.1| PREDICTED: cell division cycle protein 20 homolog B [Ceratotheriumsimumsimum]
MEWKLERTAPRRVRTEEEMLWENVMRVLAKDMKQKRRQGSSKVFDEVNPTYSDFKSNFLKRLSAEVPVAS
SPVTTRWRQSQTRDFAAYSFEEEPSTTDLPETSGSVLKITPEEETLILGSYKEPLKAPIKGSFETSNSTL
LFCKASHALDRGWKENIALKGQKCLNQLFSTQKVAHQVNGKMQLCERSQCAWKCCKDGVGDESFHLKRFG
DVNYSILQPEVKIHLTGLRNDYYLNILDWNFQNLVAIALGSSVYVWNGENHNGIEHIDLSLTCNYVSSVS
WVKEGTCLAVGTSEGEVQLLDVATKKRLRNMLGHLSVVGALSWNHCILSSGSRLGRVYHHDVRVAQHHVG
TLHHKQAVCALKWSLDGRLLSSGCSDGLLTIWPRDSGVSAQGQPLKVIPQPTAVKAMDWCPWQSAVLAVG
GGMKDGRLHILDINTGQSIQTPSTNSQICSLIWLPKTREIATGQGTPENDVTVWACPALARSGGFFGHRG
RVLHLALSPDQTRVFSAAADGTARVWNCC

PREDICTED: cell division cycle protein 20 homolog B [Felis catus] - Cat
NCBI Reference Sequence: XP_003981039.1
>gi|410948641|ref|XP_003981039.1| PREDICTED: cell division cycle protein 20 homolog B [Felis catus]
MEWKLERTAPRRVRTEEEMLWENVMRVLAKDMKQKRSQGSPKIFDGVNTTYSNFKSNFVKRLSAEVPVAS
SPITTRWRQSQTRDLEVDSFEEERSTTYLPEASGSVLKITPEKETLILGSYKEPLKTPIKGNFETNNSAL
HFCKAPHALGRGWNENVASKGQKCLNQLFLTQKVAQWVSGKMQLCEKSQCVWKGCRDGVRDESFHLKRFN
GINYSILQPEVKFHLTGLRNDYYLNILDWNFQNLVAIALGSSVYIWNGENHNKIENIDFTLNCNYVSSVS
WIKEGNCLAVGTSEGEVQLWDVVTKKRLRNMLGHLSVVGALSWNQYILSSGSRLGRVYHHDVRVARHHVG
TLHHQQAVCALKWSPDSRLLSSGCSDGLLTIWPGDPSAKAQVQPLKVIPQPTAVKAMDWCPWQSAVLAVG
GGMKDGHLHILDINTGQTIQTPNMNSQICSLIWLPKTKEIASGQGTPKNDVTVWACPGLARSAGFFGHRG
RVLHLALSPDQTRVFSAAADGTACVWNCCESSLSPSVSSSFLTLK

PREDICTED: cell division cycle protein 20 homolog B [Ovis aries] - Sheep
NCBI Reference Sequence: XP_004017037.1
>gi|426246511|ref|XP_004017037.1| PREDICTED: cell division cycle protein 20 homolog B [Ovis aries]
MIVLGDKTPEVCRLRVKTGLFLGDKTKSGRGEATLEPVGQTPPEMEWKLERAARRRVRTEEEILWEDVMR

VLDKDLKQRRGQVSPKVFNVVNTTYSNFTSHLAKRLSAEVPVASSPITARWRQSQAGDVAAPSFGEERST
TDLPETVLEMTPEKETLTPGSCKEPLKTPNKESSGTNNSAFQFGKAPHAVDRGQKDKVASKDPKSLNQLF
STQNVAQQVNGSMQFCEGSQCTWKGCSNGIRDGYFQRFSDINYPILQPEVKIHLTGLRNDYYLNILDWNF
QSLVAIALGSSVYIWNGDNHHRIENMCFSLPCNYVSSVSWMTEGTCLAVGTSEGEIQLWDVVTKKRLRNM
LGHLSVVGALSWNHYILSSGSRLGRVYHHDIRVAQHHVGTLQHKQAVCGLKWAPGGRLLSSGCSDGLLTI
WPHDPGANAQSHPLKVIHQSTAVKAIDWCPWQTEVLAVGGGMKDGRLHILDINTGKSIQTPSTDSQICSL
TWLPKTKEIATGQGSPKNDVTVWTCPGLARSRGFFDHRGRVLHLALSPDQTRVFSAAADGTACIWNCC

PREDICTED: cell division cycle protein 20 homolog B isoform X6 [Gallus gallus] - Chicken

NCBI Reference Sequence: XP_004937294.1

>gi|513229162|ref|XP_004937294.1| PREDICTED: cell division cycle protein 20 homolog B isoform X6 [Gallus gallus]

MSVLGKDPEWGREQNLQKVAEALDSDRTTYSGFKSSIKRKLKAGAPVASSPVVTRWQQRHAGGLAAVPRR
ALSAELCALRSAPLVGARGRTPVTNQMLKKPSIMKSAVTLSSRHTAAGLEGRNDRKTEIRKKRKCTTQLL
MIQETGYSMAESDMKICENTSCIWKGCREGMQNEFRTVTIMKPSVTLEPELRLHITGLRNDYYLNILDWN
LENLIAVALGSAAYIWNGRTLQGIESIELNSSSKYISSLAWIKEGTCLAVGTSDGEVQLWDIERKRRLRS
MFGHLSVVGALSWNHYILSSGSRLGSIHHHDVRVAQHHIGTLCQNKQSICSLKWSLTNQLLASGSSDGTV
NIWHSDPGVNVKSQPLKTIPHSSAVKAMNWCPWQSNVLATGGGMKDGILRVWDINHEKLLQSAATDSQIC
SLLWLPKTSELMTGQGLPENQIKIWQHPALISSSELYGHKGRVLHMALSPDQRRLFSVAADGIACLWKCH
EYEESNEKK

PREDICTED: cell division cycle protein 20 homolog B [Columba livia] – Rock Dove

NCBI Reference Sequence: XP_005500956.1

>gi|543719467|ref|XP_005500956.1| PREDICTED: cell division cycle protein 20 homolog B [Columba livia]

MECLQERRVPGSEGTTYSRFKRSIKRKLAARAAVVSSPIVTRWQQPHRRGGARERRRACCLTHLCPAGGP
PLLSTAGKAPVTPQVPKKLSVTKTLVTLHSSHAAAGLEGRKNRRAEMKKRWKRPSQLPVVQEVRYSIAEY
GDVKICENTNCIWKGCQEGTQNEFRAVTIMKPSIPLEPEVRLHITGLRNDYYLNILDWSLENLIAVAVGP
AVHIWNGETLQAIESIGLNSSSKYISSLAWIKEGTCLAVGTSDGEVQLWDIETKKRLRNMFGHLSVVGAL
SWNRYILSSGSRLGSIHHHDVRVAQHHVGTLCQNKQSICSLKWSLTNQLLASGSSDGILNIWPGDPGVKL
QSQPLKTIPHSSAVKAVNWCPWQSKILATGGGMKDGVLRVWDINQEKIIQSAATDSQICSLLWLPKTSEV
VTGQGLPGNQMKIWKYPMLISSSELYGHEGRVLHIALSPDQSRLFSVAADGIASLWKCHETGESKHNSKG
FF

PREDICTED: LOW QUALITY PROTEIN: cell division cycle 20B [Pseudopodoces humilis] – Ground Tit

NCBI Reference Sequence: XP_005534726.1

>gi|543376025|ref|XP_005534726.1| PREDICTED: LOW QUALITY PROTEIN: cell division cycle 20B [Pseudopodoces humilis]

MLQQVWKEEITGELKQKCERRGCQEKTQNELRSVTTMKLSTALEQELRLHITGCCHDYXLNMLGWSLENL
IAITVGPAAHIWNGXTLQAIESIDLNSSSKCFMSRLDKKXMCLATCTSDGEVQLWNIETKXRLRNMFGHL
FVIGXWNHCILSSGSLLGSIHHYDVQHSVGILCPKTPSICCLKLSLANLLLAIWSRGGVLNIWPRKPGVK
LQSQTVQTILPVIAVKSMTWCPWQTKVLATGGGMKNGILHVWHMNCEKITQSAIADSQKDFKLHWISSLP

WLPSQLMTAQGLPENQMKIWKYPMLVNSLELYGHKGRVLHITLSPAQNSLIFVAVYGMSCL

PREDICTED: LOW QUALITY PROTEIN: cell division cycle 20B [Geospiza fortis] – Medium Ground Finch

NCBI Reference Sequence: XP_005431751.1

>gi|543263317|ref|XP_005431751.1| PREDICTED: LOW QUALITY PROTEIN: cell division cycle 20B [Geospiza fortis]

MVPEVGSSVAESGDVKFWENTSCIWKGCQEGTQNELRNVTTMKPSTPLEQEGFILLTGAIIKLIPPMLIF
SSFLLELNMLGWSLAANVGPAAHMWNGQTPDLKAVEGIDLNPSSKHISCLAWIKKGICLAVWISDGELQL
WNTETRKRLRNMFGHLSVIXGWNHCILSSGSLLGFAHHREVQVAQHYVGSLCQNKQIIXSLKFFLANLLL
AIWSRGGTLNIWPRNPGVKLQSQTLKTNFCDGLAQAVMTWSPWQSKVLATXGGMKDGILCVWHMNCEKII
QSAIADSQCFQRAILSASQFKGNKGRVLHITPSSDHSRFIFVAVEGMSCLWKQHGSGESMHSNKAF

PREDICTED: LOW QUALITY PROTEIN: cell division cycle 20B [Zonotrichiaalbicollis] – White Throated Sparrow

NCBI Reference Sequence: XP_005497997.1

>gi|542164204|ref|XP_005497997.1| PREDICTED: LOW QUALITY PROTEIN: cell division cycle 20B [Zonotrichiaalbicollis]

MVPEVGYSIAESDDVKVWENTSCVWKGCQEGTQNELRSVTTMKPSTLLEQEVRLHTTDWCNDXVINMLAW
SLENLIPIPVGPAAHIWNGQTLDLKAVEGIDLNSCSKHIXCLAWMKXGISLAFWMSDXEVQLWNIETRKR
LRNMFGHLSVIXGWNHCILSSGSLLGFALHSEVRSAQHHVGILCQNKQIICSLKFSLANMLLSVKLLKAG
HVLAYVPRSASAEVLTYEVLVKM

PREDICTED: cell division cycle protein 20 homolog B [Falco cherrug] – Saker Falcon

NCBI Reference Sequence: XP_005438741.1

>gi|541969095|ref|XP_005438741.1| PREDICTED: cell division cycle protein 20 homolog B [Falco cherrug]

MLGKDCDISVFVWVFPVNWGHVRLDLCMAVRPVPDSGRITYSGFKSSMKRRLAARAAVASSPIVTRWQQL
HARGRSPARLGASAGPQRGRAPGETPVTPQVPKKLSVTKTVVTLPSSHAALGLEGSSNMRAGMEKKQICS
SQLPVLQEVGYSIAENGDMKVCENTSCIWKGCQEGTHSDFRAVTTMKPSILWEPEARLHISGLRNDYYLN
ILDWSLENLVAVAVGPAAHIWKGGALQTIERIYLNSNSKYISSLAWIKEGTCLAIGTSDGEVQLWDIETK
KRLRNMFGHLSVVGALSWNHHILSSGSRLGSIHHHDVRVAQHHVGTLRHNKKSICSLKWSLTDQLLASGS
SDGILNIWPGDPGVKLQLQPLKTIPHSSAVKAMNWCPWQSKVLATGGGMKDRILRVWDINCEKIIQSAAT
DSQVCSLLWLPQTSELMTGQGLPGNQMKIWKYPMLINSSELYGHKGRVLHIALNPDQNRLFSVAADGIAC
LWKCCEPGQSKHSSKAFSSDYLSSFL

PREDICTED: cell division cycle protein 20 homolog B [Melopsittacus undulatus] – Pet Parakeet

NCBI Reference Sequence: XP_005152152.1

>gi|527268118|ref|XP_005152152.1| PREDICTED: cell division cycle protein 20 homolog B [Melopsittacus undulatus]

MYFFLVFSIEVTGCQHPFPWPVLCGDQSPSGADPVAEGTTYSGFKNAMRRKLAARAPVASSPLVTRWQQR
HAGGTAEARGAIRLISAPGTTAVTSQVPKELSSIKTVVPLASRHAAAGLEGRNIRRAGMEKKQKCPSQLP
VVQDVGYSIAERGDMKVCENTNCTWKGCQEGNQNEFRAVTKMKLSVPMELEVRLHITGLRNDYYLNILDW

SLENLIAIAVGPAAHIWNGGTLQAIGSIDLNSSSKYISSLAWITEGTCLAIGTSDGEVQLWDIETKKRLR
NMFGHLSVVGALSWNHYILSSGSRLGSIHHHDVRVAQHHVGTLCHKKQSICSLKWSLTNQLLASGSSDGI
LNIWPSDPGVKLQSQPLKTIPHSSAVKAMNWCPWQSKVLATGGGMKDGILRVWDINREKIIQSAATDSQI
CSLLWLPKTSELMTGHGLPGHQMKIWKYPILINSSELYGKYFIHKGRVLHTALSPDQSRLFSVAADGTSC
LWKCHEPRESKHSSKAFSEDLPADDRDIYLPQGIRKHFSDDCVFPILLLCKLFLDWGSSEDFARYFEYGL
AVFMFISHGAWAKVFVSLLREAGKERPQELPAAVAEAARESSPGNQMAVAEQQEEAGEERREAVAEQQDQ
TAEEPSPVSKPNPAATERVGRARVVVSERAVCGLDFNHDTI

PREDICTED: cell division cycle protein 20 homolog B [Falco peregrinus] – Duck Hawk

NCBI Reference Sequence: XP_005242064.1

>gi|529444705|ref|XP_005242064.1| PREDICTED: cell division cycle protein 20 homolog B [Falco peregrinus]
MLGKDCDISVFVWVFPVNWGLVRLDLCMAVRPVPDSERITYSGFKSSMKRRLAARAAVASSPIVTRWQQL
HARGRSPARLGASAGPQRGRAPGETPVTPQVPKKLSVMKTVVTLPSSHAALGLEGSSNMRAGMEKKQICS
SQLPVLQEVGYSIAENGDMKVCENTSCIWKGCQEGTHSDFRAVTTMKPSILWEPEVRLHISGLRNDYYLN
ILDWSLENLVAVAVGPAAHIWKGGALQTIERIYLNSNSKYISSLAWIKEGTCLAIGTSDGEVQLWDIETK
KRLRNMFGHLSVVGALSWNHHILSSGSRLGSIHHHDVRVAQHHVGTLRHNKKSICSLKWSLTDQLLASGS
SDGILNIWPGDPGVKLQLQPLKTIPHSSAVKAMNWCPWQSKVLATGGGMKDRILRVWDINCEKIIQSAAT
DSQVCSLLWLPQTSELMTGQGLPGNQMKIWKYPMLINSSELYGHKGRVLHIALNPDQNRLFSVAADGIAC
LWKCCEPGQSKHSSKAFSNDYLSSFL

PREDICTED: cell division cycle protein 20 homolog B [Alligator sinensis] – Chinese Alligator

NCBI Reference Sequence: XP_006017361.1

>gi|557263746|ref|XP_006017361.1| PREDICTED: cell division cycle protein 20 homolog B [Alligator sinensis]
MKLFFIQAYVPFHSAVGLRVIWQVKMSYHFVKTSKFVQSIGVERRSSVRLNLMRWLPGGKHSSQGSNNRS
GVALPQGNDLCVTLSTLEIVPTPKTIRVEPGGRQQLIRVVPGITPVLRRGSPASRLTPDFGKTTYSRFKS
SIVKKLTSSVPVASSPISTRWQQAHVKSEAEQTFTECWPINLTPSDGSELTRTPEEITKIPVTSPMLIKT
SATEPVINMNARYHKAYREDSQGRNKRPAESTNEQKCLTQILMVQEYGVSFPENGRMKICENMECIWEGC
QDGIQNEFTLHRVSLDTEPFLPLEPEVRFHITGFRNDYYLNTLDWNRENLIAVALGSCAYIWNGETHQSI
ESIDLNSSAKYISSIAWIKEGTCLAVGTSDGEVQLWDIETKKRLRNMVGHLSVVGAMSWNHYILSSGSRL
GFVHHHDVRVAQHHVGTLCQNKQSICSLKWSLNNKLLASGSSDGILNIWPSDPGATVHCKPVQTMPHPSA
VKAMNWCPWQSEILATGGGLKDGLLRVWDISSGKTIKTTETKSQICSLFWLPNTNELVTGHGLPKNQIKI
WKYPMLANSVELFGHKGRVLHVALSPDCSRFISLAADGVACVWKYFQSSETRHRRDDIQGDYSRTLW

PREDICTED: cell division cycle protein 20 homolog B isoform X1 [Pelodiscus sinensis] – Chinese Soft Shelled Turtle

NCBI Reference Sequence: XP_006139607.1

>gi|558232057|ref|XP_006139607.1| PREDICTED: cell division cycle protein 20 homolog B isoform X1 [Pelodiscus sinensis]
MEWKLERSASRKVKTDEDMLWEKVMKTLGKELKQIRKQSLQKLTKTPDSAKPTYSRFKSCIVKKLTSNIP
VASSPIITRWQQTHTRSQAEQTFTECQPINLFPADGSVQVRTPEKISTIPVFFHVPIKNSTKEPDIITNA
PASCHETTYLQGKNKRRAEIAKEQKCLTQLVMLQENGYRISENDDIKICGNTECVWKGCQNGIHNEFTAH

RISTNIKSLVPLEPEVRLHITGLRNDYYLNLLDWNQGNLIAIALGSAAHIWNGESHQSTQRVNLNSSSKY
ISSVAWIKEGTCLAVGTSDGEVQLWDIETKKRLRNMFGHLSVVGAMSWNHYILSSGSRLGLIHHHDVRIA
QHHVGTLCHSKQSICSLKWSPDSKLLASGSSDGLLNIWPNDPGLTVHCKPLKAIPQASAVKAMDWCPWQS
EILATGGGMKDGFLRVWDISKGKIIRTTDTKSQICSLLWLPKTNELMTGQGIPKNQMKIWKYPTLTNSTE
LHGHKGRVLHIALSPDHCRIFSLAADGLACVWKYG

PREDICTED: cell division cycle protein 20 homolog B isoform X2 [Xenopus (Silurana) tropicalis] – Western Clawed Frog

NCBI Reference Sequence: XP_004910413.1

>gi|512809821|ref|XP_004910413.1| PREDICTED: cell division cycle protein 20 homolog B isoform X2 [Xenopus (Silurana) tropicalis]
MDWKLEKISSRRVRAEEDMMWEGVMKTLGKDLRWSRKYSMQKMNKTPDGRSSYSRFKRSIMKRRACHVPF
ASSPVSTKWCQNMDQIQGTGLSSSPSEELCQTPEPLVLLPYLTEIPITAIPDCAQKPPSLKDTSAPKRTQ
KTSKKLLSQTRLKFLQEQNFLQTVEDFKRGSCIRHDSNIKICEKDTCMWKGCLQTKKNKSLGQKTQKVHI
CSHTILQPDIRLHIIGLHNDYYLNLLDWNSENLVAIGLKSTAYIFSGENRTVTQKIHLSCPATYVSSVSW
ISSGTCLAIGTSSGEVQLWDIETQKRLRNMLGHMSVVGALSWNNHILSSGSRLGHIHHHDVRIAEHHIGT
LQHKQGICSLKWSPCGNKLASGSSDGDLKIWPCDPGETKLKSPLLNMPHPTAVKAMNWCPWLSDTLAVGG
GMTDGLIRIWDTNSGKNIHSANTNSQICSLLWLPQTKELLTGHGPSRNQMTIWQYPSLLKMNDYYGHKGR
VLHLALSPDQRRIFSAAANGTANIWKYGN

PREDICTED: cell division cycle protein 20 homolog B [Latimeria chalumnae] - West Indian Ocean coelacanth

NCBI Reference Sequence: XP_005998514.1

>gi|556985654|ref|XP_005998514.1| PREDICTED: cell division cycle protein 20 homolog B [Latimeria chalumnae]
MESSGYGHQGSSDESSFQCLFANRHAECSISSQPETRLDITGLKDDYYLNLLDWNSENLVAIAVDAAAHI
LNGDTQVMHGKIDLQSDFKYISSVAWLKDRSCLAIGTSDGEVQLWDVETKKRLRNMLGHMSVVGALSWNS
YILSSGSRLGLIHHHDVRIAQHHVGSLRHKQSICALKWSGNGKYLASGSNNGLLNIWPNDPGVTVQHQPL
LAFAHTTAVKAMNWCPWQANILAVGGGIKDGRLRTWDLNSGKCVRIVETKSQICSLLWLPKTNELLTAHG
HPRNQISLWKYPSLKKTANLYGHRGRILHLAMSPDQSRICSVAADGTACIWKCLY

## Appendix II

### A 3.1 Rare gene variants identified at the 2q33-q36 locus (SCT135 family)

| Gene name | Genomic position (Grch38) | Allele | Consequence | cDNA change | Amino acid change | rs ID |
|---|---|---|---|---|---|---|
| *RAPH1* | 203430502 | G/A | downstream | _ | _ | rs993751252 |
| *RAPH1* | 203481759 | T/- | intron | c.732+7825del | _ | rs757903816 |
| *RAPH1* | 203522658 | T/C | intron | c.-1+12453A>G | _ | rs569076111 |
| *FN1* | 215405565 | G/A | intron | c.2986+673C>T | _ | rs1177012037 |
| *FN1* | 215419921 | G/C | intron | c.1676-536C>G | _ | rs548810098 |
| *PARD3B* | 204619869 | T/C | intron | c.121-66312T>C | _ | - |
| *PARD3B* | 204637730 | -/T | intron | c.121-48439dup | _ | rs796534134 |
| *PARD3B* | 204674440 | A/G | intron | c.121-11741A>G | _ | rs185339992 |
| *PARD3B* | 204706035 | A/- | intron | c.222+19755del | _ | - |
| *PARD3B* | 204740180 | -/T | intron | c.222+53911dup | _ | rs547815624 |
| *PARD3B* | 204803142 | -/G | intron | c.222+116862dup | _ | - |
| *PARD3B* | 204811849 | C/A | intron | c.222+125567C>A | _ | rs553140066 |
| *PARD3B* | 204857324 | C/A | intron | c.223-107828C>A | _ | rs568295831 |
| *PARD3B* | 204874866 | A/T | intron | c.223-90286A>T | _ | rs532403294 |
| *PARD3B* | 204928791 | T/A | intron | c.223-36361T>A | _ | rs557483257 |
| *PARD3B* | 204997823 | A/G | intron | c.394+32500A>G | _ | - |
| *PARD3B* | 205013743 | G/A | intron | c.395-33838G>A | _ | - |
| *PARD3B* | 205036626 | G/A | intron | c.395-10955G>A | _ | rs981689283 |
| *PARD3B* | 205049541 | G/A | intron | c.504+1851G>A | _ | rs1339859281 |
| *PARD3B* | 205083990 | C/T | intron | c.505-20436C>T | _ | rs1331976154 |
| *PARD3B* | 205095341 | C/T | intron | c.505-9085C>T | _ | rs1309017480 |
| *PARD3B* | 205168320 | -/GAGAGAGAGTGT | intron | c.1621-3890_1621-3889insAGAGAGAGTGTG | _ | rs371904121 |
| *PARD3B* | 205171677 | G/A | intron | c.1621-534G>A | _ | rs1221160245 |
| *PARD3B* | 205185326 | G/A | intron | c.1925-438G>A | _ | rs1185486899 |
| *PARD3B* | 205188791 | AAAAAAAA/- | intron | c.2024+2933_2024+2940del | – | rs68034154 |
| *PARD3B* | 205241354 | AAGATC/- | intron | c.2141-4425_2141-4420del | – | rs1399724320 |
| *PARD3B* | 205242733 | C/T | intron | c.2141-3045C>T | _ | rs534150255 |
| *PARD3B* | 205291604 | G/A | intron | c.2186-8926G>A | _ | rs574128236 |
| *PARD3B* | 205403388 | C/T | intron | c.2741+2265C>T | _ | rs959585731 |
| *PARD3B* | 205423144 | A/G | intron | c.2741+22021A>G | _ | - |
| *PARD3B* | 205426263 | A/G | intron | c.2741+25140A>G | _ | rs565448402 |
| *PARD3B* | 205489711 | A/G | intron | c.2742-10185A>G | _ | - |
| *PARD3B* | 205496858 | C/G | intron | c.2742-3038C>G | _ | - |
| *PARD3B* | 205507206 | TTTTTTTTT/- | intron | c.2877+7192_2877+7201del | – | rs71410819 |
| *PARD3B* | 205509529 | C/T | intron | c.2877+9498C>T | _ | rs546012821 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *PARD3B* | 205557745 | -/T | intron | c.2957+4349dup | _ | rs535255748 |
| *ACSL3* | 222928665 | -/AAAAC | intron | c.1466-194_1466-193insACAAA | – | rs1320841235 |
| *SLC4A3* | 219627530 | CGG/- | upstream | _ | _ | rs1184272855 |
| *NRP2* | 205734187 | T/C | intron | c.1146+6141T>C | _ | rs375677863 |
| *NRP2* | 205770813 | T/A | intron | c.2425+4010T>A | _ | - |
| *NRP2* | 205772253 | G/A | intron | c.2425+5450G>A | _ | - |
| *MOB4* | 197545689 | T/C | intron | c.259-2647T>C | _ | rs546609847 |
| *MOB4* | 197552917 | C/- | 3' UTR | c.*2271del | _ | rs1459274442 |
| *ASIC4* | 219530229 | G/- | intron | c.964-1524del | _ | rs562663240 |
| *CTDSP1* | 218401413 | G/A | intron | c.68-151G>A | _ | rs1412406093 |
| *UNC80* | 209826987 | -/ATCT | intron | c.2463+965_2463+968dup | – | rs567993337 |
| *UNC80* | 209887016 | C/G | intron | c.4102-1079C>G | _ | rs560422875 |
| *UNC80* | 209950605 | C/G | intron | c.7074-3495C>G | _ | - |
| *PAX3* | 222209724 | AAAAAAAAAAAAAAAAAAAAAAAAAAAA/- | intron | c.1174-8282_1174-8255del | – | rs546468709 |
| *PAX3* | 222255297 | T/C | intron | c.587-23014A>G | _ | rs865854385 |
| *PAX3* | 222264004 | C/T | intron | c.586+30163G>A | _ | rs182522407 |
| *PAX3* | 222302941 | GT/- | upstream | | _ | - |
| *PAX3* | 222303511 | G/A | upstream | | _ | - |
| *DAW1* | 227882442 | C/T | intron | c.41-2909C>T | _ | rs542200987 |
| *DAW1* | 227917146 | TCTG/"TCTGTCTG""" | intron | _ | _ | - |
| *CYP20A1* | 203297084 | T/C | 3' UTR | c.*176T>C | _ | - |
| *FBXO36* | 229927472 | A/G | intron | c.96+4863A>G | _ | rs536688627 |
| *FBXO36* | 229929700 | T/C | intron | c.96+7091T>C | _ | rs532069168 |
| *FBXO36* | 229980794 | G/A | intron | c.205+4445G>A | _ | - |
| *FBXO36* | 229995901 | T/- | intron | c.206-839del | _ | rs1212028671 |
| *FAM126B* | 201053181 | G/A | intron | c.-128-7588C>T | _ | rs533715460 |
| *FAM126B* | 201075182 | AAA/- | upstream | | _ | rs752371143 |
| *FAM117B* | 202658562 | A/G | intron | c.601+22774A>G | _ | rs531701685 |
| *FAM117B* | 202665186 | -/T | intron | c.601+29411dup | _ | rs1052100943 |
| *FAM117B* | 202715553 | C/G | intron | c.754-9364C>G | _ | - |
| *FAM117B* | 202724054 | TTT/- | intron | c.754-844_754-842del | – | - |
| *ABCA12* | 215012302 | G/A | intron | c.1957-167C>T | _ | rs917286812 |
| *ABCA12* | 215026709 | T/G | intron | c.1180+111A>C | _ | rs987081478 |
| *ABCA12* | 215055746 | T/C | intron | c.318-1082A>G | _ | - |
| *ABCA12* | 215103150 | T/G | intron | c.163+8447A>C | _ | - |
| *ABCA12* | 215103782 | GTTTA/- | intron | c.163+7815_163+7819del | – | rs1006208950 |
| *ANKRD44* | 197098319 | C/T | downstream | _ | _ | - |
| *ANKRD44* | 197286575 | G/A | intron | c.27+24003C>T | _ | rs1030931805 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *ARPC2* | 218226922 | TTCA/- | intron | c.109+966_109+969del | – | - |
| *PIKFYVE* | 208292594 | T/C | intron | c.620+3776T>C | _ | rs768362084 |
| *PIKFYVE* | 208296014 | T/- | intron | c.621-2616del | _ | rs996291100 |
| *PIKFYVE* | 208306160 | G/C | downstream | _ | _ | rs181223471 |
| *SLC16A14* | 230033225 | C/T | downstream | _ | _ | rs577681437 |
| *SLC16A14* | 230058028 | -/A | intron | c.259+1065dup | _ | rs35641768 |
| *C2CD6* | 201552267 | G/T | intron | c.978-4657C>A | _ | rs550587835 |
| *C2CD6* | 201604015 | T/G | intron | c.318+596A>C | _ | rs527710406 |
| *SGO2* | 200549392 | A/T | intron | c.473+6728A>T | _ | rs540874338 |
| *SGO2* | 200589016 | A/C | downstream | _ | _ | rs550101234 |
| *CCNYL1* | 207752035 | -/A | intron | c.759+928dup | _ | rs999012414 |
| *KANSL1L* | 210062044 | T/C | intron | c.1755+13508A>G | _ | rs576559715 |
| *KANSL1L* | 210109023 | G/A | intron | c.1231-4722C>T | _ | rs532648832 |
| *KANSL1L* | 210115533 | C/T | intron | c.1231-11232G>A | _ | - |
| *KANSL1L* | 210118642 | T/- | intron | c.1230+10389del | _ | - |
| *KANSL1L* | 210130349 | T/C | intron | c.1089-1177A>G | _ | rs547296260 |
| *TMEM237* | 201626436 | T/C | intron | c.1014-289A>G | _ | rs533285967 |
| *KCTD18* | 200497641 | T/C | intron | c.661+112A>G | _ | rs548684633 |
| *KCTD18* | 200500513 | A/G | intron | c.373-1429T>C | _ | rs531743722 |
| *KCTD18* | 200501760 | C/T | intron | c.373-2676G>A | _ | - |
| *KCTD18* | 200509966 | T/C | 5' UTR | c.-414A>G | _ | - |
| *KCTD18* | 200509973 | C/G | 5' UTR | c.-421G>C | _ | - |
| *KCTD18* | 200513132 | G/C | upstream | _ | _ | - |
| *KCTD18* | 200513133 | G/C | upstream | _ | _ | - |
| *KCTD18* | 200513134 | G/C | upstream | _ | _ | - |
| *SGPP2* | 222443411 | C/T | intron | c.219+18590C>T | _ | - |
| *SGPP2* | 222482881 | G/A | intron | c.378+8155G>A | _ | rs867019125 |
| *SGPP2* | 222519762 | T/G | intron | c.379-2005T>G | _ | rs1281731079 |
| *SGPP2* | 222549841 | G/A | intron | c.649-8506G>A | _ | rs1163832157 |
| *METTL21A* | 207624749 | A/C | intron | c.-30+93T>G | _ | rs539918048 |
| *RFTN2* | 197588600 | G/- | intron | c.1233+7391del | _ | - |
| *RFTN2* | 197620188 | A/- | intron | c.929-2267del | _ | rs1159201273 |
| *RFTN2* | 197622644 | C/G | intron | c.929-4723G>C | _ | - |
| *RFTN2* | 197638863 | T/C | intron | c.439-4866A>G | _ | rs566193361 |
| *RFTN2* | 197679446 | A/- | upstream | _ | _ | rs1171961772 |
| *CDK15* | 201848573 | C/T | intron | c.792+1099C>T | _ | rs533741825 |
| *CDK15* | 201881497 | C/A | intron | c.1045+1330C>A | _ | rs538745850 |
| *CDK15* | 201885224 | G/A | intron | c.1045+5057G>A | _ | rs1264600139 |
| *DNER* | 229468823 | G/C | intron | c.1261+8317C>G | _ | - |
| *DNER* | 229544777 | G/A | intron | c.993+2170C>T | _ | rs1009673106 |
| *DNER* | 229544778 | C/A | intron | c.993+2169G>T | _ | - |
| *DNER* | 229576148 | T/C | intron | c.847+9710A>G | _ | rs746745413 |
| *DNER* | 229649476 | T/C | intron | c.277-57588A>G | _ | rs562909102 |
| *DNER* | 229652796 | G/A | intron | c.277-60908C>T | _ | - |

| | | | | | | |
|---|---|---|---|---|---|---|
| *DNER* | 229709066 | G/A | intron | c.276+5082C>T | _ | rs564475129 |
| *ICA1L* | 202806646 | -/A | intron | c.910+5099dup | _ | rs796219614 |
| *CREB1* | 207587468 | C/T | intron | c.882-9446C>T | _ | - |
| *PPIL3* | 200873195 | -/T | intron | c.360-1675dup | _ | rs533187117 |
| *PPIL3* | 200875561 | C/T | intron | c.359+1358G>A | _ | rs545932195 |
| *MYL1* | 210305968 | G/A | upstream | _ | _ | rs571094159 |
| *MYL1* | 210305968 | G/A | intron | c.133-3453C>T | _ | rs571094159 |
| *MYL1* | 210310809 | C/T | intron | c.132+4102G>A | _ | rs559871087 |
| *MOGAT1* | 222670969 | C/T | upstream | _ | _ | rs563785930 |
| *CASP8* | 201232360 | C/G | upstream | _ | _ | rs570559698 |
| *MPP4* | 201665067 | TTTT/- | intron | c.1052-966_1052-963del | – | - |
| *CASP10* | 201189785 | C/A | intron | c.441+1986C>A | _ | rs559545059 |
| *CASP10* | 201232360 | C/G | downstream | | _ | rs570559698 |
| *RHBDD1* | 226883510 | G/A | intron | c.566+16192G>A | _ | rs571084977 |
| *RHBDD1* | 226902323 | A/G | intron | c.567-4470A>G | _ | rs991870599 |
| *RHBDD1* | 226942243 | TT/- | intron | c.856+27904_856+27905del | – | rs11449035 |
| *RHBDD1* | 226950718 | G/A | intron | c.856+36367G>A | _ | rs530279674 |
| *RHBDD1* | 226981435 | -/A | intron | c.857-13982dup | _ | rs201620422 |
| *SPHKAP* | 227975202 | T/C | downstream | | _ | rs546399230 |
| *SPHKAP* | 228005459 | G/T | intron | c.4449-9765C>A | _ | rs570041801 |
| *SPHKAP* | 228043912 | G/A | intron | c.247-16369C>T | _ | rs536127499 |
| *SPHKAP* | 228090287 | C/T | intron | c.246+18545G>A | _ | rs539288434 |
| *SPHKAP* | 228174303 | -/T | intron | c.32+7263dup | _ | rs76289406 |
| *SPHKAP* | 228181663 | G/C | upstream | _ | _ | rs547462934 |
| *SLC19A3* | 227679225 | G/T | downstream | _ | _ | rs184431804 |
| *SLC19A3* | 227708874 | C/T | intron | c.-2-6554G>A | _ | rs563376551 |
| *SLC19A3* | 227713316 | C/T | intron | c.-3+4627G>A | _ | rs531711796 |
| *WNT10A* | 218876530 | T/G | upstream | | _ | rs1043802370 |
| *PGAP1* | 196834398 | G/C | 3' UTR | c.*6836C>G | _ | rs567059943 |
| *TM4SF20* | 227357595 | C/T | downstream | | _ | - |
| *NHEJ1* | 219102710 | -/A | intron | c.589-24505dup | _ | rs879703835 |
| *NHEJ1* | 219103038 | A/C | intron | c.589-24832T>G | _ | - |
| *NHEJ1* | 219150787 | C/G | intron | c.391-2992G>C | _ | rs538334205 |
| *CARF* | 202953692 | G/A | intron | c.428-313G>A | _ | - |
| *SPAG16* | 213342378 | TA/- | intron | c.644+2119_644+2120del | – | rs562141851 |
| *SPAG16* | 213346065 | G/A | intron | c.645-4463G>A | _ | - |
| *SPAG16* | 213442715 | G/A | intron | c.943-47248G>A | _ | rs72939102 |
| *SPAG16* | 213492949 | TC/- | intron | c.1070+2870_1070+2871del | – | rs372933104 |
| *SPAG16* | 213520088 | AG/- | intron | c.1070+30018_1070+30019del | – | rs35798649 |
| *SPAG16* | 213579512 | A/G | intron | c.1070+89422A>G | _ | - |
| *SPAG16* | 213584809 | T/C | intron | c.1070+94719T>C | _ | rs902766004 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *SPAG16* | 213588524 | C/T | intron | c.1070+98434C>T | _ | - |
| *SPAG16* | 213605268 | -/T | intron | c.1070+115194dup | _ | rs34856658 |
| *SPAG16* | 213619694 | C/T | intron | c.1070+129604C>T | _ | - |
| *SPAG16* | 213688048 | A/G | intron | c.1071-174437A>G | _ | - |
| *SPAG16* | 213696990 | GGAG/- | intron | c.1071-165497_1071-165494del | _ | - |
| *SPAG16* | 213839809 | C/T | intron | c.1071-22676C>T | _ | rs1032461064 |
| *SPAG16* | 213841883 | C/A | intron | c.1071-20602C>A | _ | - |
| *SPAG16* | 213843600 | C/T | intron | c.1071-18885C>T | _ | rs576334834 |
| *SPAG16* | 213874861 | G/C | intron | c.1214+12233G>C | _ | rs1033740987 |
| *SPAG16* | 213916622 | C/G | intron | c.1215-13338C>G | _ | - |
| *SPAG16* | 213936481 | G/- | intron | c.1400+6336del | _ | - |
| *SPAG16* | 213962753 | A/C | intron | c.1400+32608A>C | _ | - |
| *SPAG16* | 213966000 | T/A | intron | c.1400+35855T>A | _ | rs963945712 |
| *SPAG16* | 214059260 | -/TGTA | intron | c.1527+45184_1527+45185insGTAT | – | rs1553706252 |
| *SPAG16* | 214065561 | -/T | intron | c.1528-42633dup | _ | rs528752885 |
| *SPAG16* | 214076920 | A/C | intron | c.1528-31276A>C | _ | rs1434566396 |
| *SPAG16* | 214176810 | TG/- | intron | c.1720+27558_1720+27559del | – | rs141462328 |
| *SPAG16* | 214189081 | T/C | intron | c.1720+39815T>C | _ | rs369511287 |
| *SPAG16* | 214194228 | G/C | intron | c.1720+44962G>C | _ | rs376546548 |
| *SPAG16* | 214298175 | ATACACACACACACAC/"ACACACACACACAC""" | intron | – | – | - |
| *SPAG16* | 214321663 | C/A | intron | c.1721-88477C>A | _ | rs868085594 |
| *SPAG16* | 214355292 | A/C | intron | c.1721-54848A>C | _ | - |
| *SPAG16* | 214380529 | A/G | intron | c.1721-29611A>G | _ | - |
| *SPAG16* | 214390754 | G/A | intron | c.1721-19386G>A | _ | rs533549464 |
| *SPAG16* | 214394274 | G/A | intron | c.1721-15866G>A | _ | rs898822006 |
| *TNS1* | 217855558 | TC/- | intron | c.1055-6471_1055-6470del | – | rs3838562 |
| *TNS1* | 217870346 | A/C | intron | c.1054+10552T>G | _ | rs570160788 |
| *TNS1* | 217890215 | A/G | intron | c.491+747T>C | _ | - |
| *TNS1* | 217893400 | -/CGCGCA | intron | c.342+38_342+39insTGCGCG | – | rs58297965 |
| *PNKD* | 218347215 | -/A | downstream | | _ | rs954482127 |
| *RNF25* | 218671266 | -/T | intron | c.41+663dup | _ | rs35556340 |
| *TMBIM1* | 218295333 | A/T | upstream | | _ | rs541151171 |
| *XRCC5* | 216136357 | A/- | intron | c.1114-717del | _ | rs767749732 |
| *EEF1B2* | 206157963 | -/TTCT | upstream | | _ | rs539046435 |
| *USP37* | 218458275 | -/AC | intron | c.2644-1116_2644-1115dup | – | - |
| *USP37* | 218505608 | G/T | intron | c.1025+4371C>A | _ | - |

| | | | | | | |
|---|---|---|---|---|---|---|
| *ALS2* | 201747462 | -/T | intron | c.1816-715dup | _ | rs35304479 |
| *NYAP2* | 225446216 | -/TCTGTCTCTCTC | intron | c.221+37118_221+37119insGTCTCTCTCTCT | _ | rs1422980177 |
| *NYAP2* | 225655937 | ACATACACACAT/- | downstream | _ | _ | rs1245298448 |
| *4-Mar* | 216348069 | C/T | intron | c.516+21676G>A | _ | rs796100931 |
| *4-Mar* | 216371874 | -/G | 5' UTR | c.-1615dup | _ | - |
| *HECW2* | 196226395 | A/G | intron | c.3918-525T>C | _ | rs574670136 |
| *HECW2* | 196381227 | T/C | intron | c.293-37463A>G | _ | rs1186968513 |
| *HECW2* | 196401901 | G/A | intron | c.292+31231C>T | _ | rs577326564 |
| *HECW2* | 196439030 | C/T | intron | c.-35-5572G>A | _ | rs948169563 |
| *HECW2* | 196573870 | -/A | intron | c.-36+18593dup | _ | rs1052835076 |
| *MFF* | 227324771 | T/C | upstream | | _ | rs375972413 |
| *MFF* | 227328321 | AAAAAC/- | intron | c.-252-362_-252-357del | _ | rs1182644669 |
| *MFF* | 227357595 | C/T | 3' UTR | c.*478C>T | _ | - |
| *C2orf83* | 227610047 | -/AAAAAAAAAAAAAAAAAAAAAAAAA | downstream | _ | _ | - |
| *DNAH7* | 195756076 | C/T | intron | c.11586+57G>A | _ | rs747717290 |
| *DNAH7* | 195768464 | G/A | intron | c.11433+3196C>T | _ | - |
| *DNAH7* | 195889961 | T/C | intron | c.5047-980A>G | _ | rs531106848 |
| *DNAH7* | 195915714 | T/C | intron | c.3936-5519A>G | _ | - |
| *DNAH7* | 196031306 | C/T | intron | c.399-3259G>A | _ | rs533983080 |
| *PECR* | 216051281 | -/A | intron | c.603+167dup | _ | rs147574036 |
| *PECR* | 216056519 | G/A | intron | c.506+2376C>T | _ | rs1018250395 |
| *PECR* | 216056524 | C/T | intron | c.506+2371G>A | _ | - |
| *MREG* | 215944002 | C/A | 3' UTR | c.*861G>T | _ | - |
| *PID1* | 229089964 | T/C | intron | c.271-63856A>G | _ | - |
| *PID1* | 229115835 | C/G | intron | c.270+39983G>C | _ | rs558364146 |
| *PID1* | 229124677 | -/A | intron | c.270+31140dup | _ | rs938961830 |
| *PID1* | 229148050 | C/T | intron | c.270+7768G>A | _ | rs532416011 |
| *PID1* | 229153991 | A/G | intron | c.270+1827T>C | _ | - |
| *PID1* | 229203168 | C/T | intron | c.124-47204G>A | _ | - |
| *PID1* | 229209754 | A/T | intron | c.123+52929T>A | _ | - |
| *PID1* | 229217258 | A/C | intron | c.123+45425T>G | _ | - |
| *PID1* | 229228062 | T/C | intron | c.123+34621A>G | _ | rs781045636 |
| *PID1* | 229254165 | C/G | intron | c.123+8518G>C | _ | COSV67426841 |
| *PID1* | 229267929 | A/- | intron | c.-89+3085del | _ | rs965881707 |
| *PID1* | 229269576 | C/T | intron | c.-89+1438G>A | _ | - |
| *INO80D* | 205989595 | C/T | downstream | | _ | rs562399034 |
| *INO80D* | 206010064 | AC/- | intron | c.1543-270_1543-269del | _ | rs144790541 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *INO80D* | 206046771 | A/- | intron | c.965-159del | _ | - |
| *INO80D* | 206046775 | A/G | intron | c.965-163T>C | _ | - |
| *IKZF2* | 213107335 | G/A | intron | c.139+40373C>T | _ | - |
| *IKZF2* | 213119875 | T/C | intron | c.139+27833A>G | _ | rs900728913 |
| *NOP58* | 202277122 | C/T | intron | c.123-828C>T | _ | rs1434511508 |
| *NOP58* | 202285102 | C/T | intron | c.434+621C>T | _ | rs867305301 |
| *STK36* | 218671266 | -/T | upstream | _ | _ | rs35556340 |
| *SPATS2L* | 200301105 | G/A | upstream | _ | _ | - |
| *SPATS2L* | 200308628 | ACAC/- | intron | c.-73+2480_-73+2483del | – | - |
| *SPATS2L* | 200344384 | C/T | intron | c.-23+14904C>T | _ | rs530409983 |
| *SPATS2L* | 200344520 | C/T | intron | c.-23+15040C>T | _ | rs367817593 |
| *SPATS2L* | 200390868 | A/C | intron | c.39+1585A>C | _ | rs578231219 |
| *SPATS2L* | 200407046 | C/A | intron | c.40-5265C>A | _ | - |
| *SPATS2L* | 200474104 | T/- | intron | c.1281+1063del | _ | rs559587714 |
| *PNKD* | 218295333 | A/T | intron | c.236+23784A>T | _ | rs541151171 |
| *PNKD* | 218303577 | TTT/- | intron | c.236+32044_236+32046del | – | rs59067154 |
| *PNKD* | 218315038 | -/TCTTTCTTTCTTTCTTTCTTTCTTTC | intron | c.237-24744_237-24743insCTTTCTTTCTTTCTTTCTTTCTTTCT | – | - |
| *OBSL1* | 219551155 | G/A | intron | c.5684-313C>T | _ | rs1165309466 |
| *SATB2* | 199338398 | T/C | intron | c.1174-9488A>G | _ | rs550820581 |
| *SATB2* | 199366062 | C/T | intron | c.700+2543G>A | _ | rs562615265 |
| *SATB2* | 199395088 | -/T | intron | c.347-13269dup | _ | rs548201582 |
| *SATB2* | 199403409 | G/A | intron | c.347-21589C>T | _ | rs1559033419 |
| *SATB2* | 199437675 | T/C | intron | c.170-4161A>G | _ | rs1013512106 |
| *SATB2* | 199446469 | G/A | intron | c.169+9400C>T | _ | - |
| *SATB2* | 199461522 | -/A | intron | c.-140-991dup | _ | rs751358455 |
| *SATB2* | 199466944 | T/C | intron | c.-141+4070A>G | _ | rs539141538 |
| *TRAK2* | 201403793 | T/C | intron | c.287-2699A>G | _ | - |
| *TRAK2* | 201404547 | T/C | intron | c.286+2856A>G | _ | rs556781523 |
| *PIKFYVE* | 208306160 | G/C | intron | c.1636+1147G>C | _ | rs181223471 |
| *PIKFYVE* | 208314367 | C/T | synonymous | c.1770C>T | p.Asn590%3D | rs61752185 |
| *PIKFYVE* | 208322158 | T/C | intron | c.2190+1799T>C | _ | - |
| *PIKFYVE* | 208336395 | T/C | intron | c.4520+195T>C | _ | rs140875159 |
| *PIKFYVE* | 208353406 | C/T | intron | c.5845-492C>T | _ | rs190726596 |
| *FASTKD2* | 206766380 | G/A | intron | c.-50-264G>A | _ | - |
| *FASTKD2* | 206770731 | AA/- | intron | c.882-436_882-435del | – | rs549037266 |
| *DOCK10* | 224770741 | C/T | intron | c.6205-96G>A | _ | - |
| *DOCK10* | 225001197 | T/C | intron | c.123+41055A>G | _ | rs547264930 |
| *DOCK10* | 225012881 | A/T | intron | c.123+29371T>A | _ | rs1398096086 |
| *DOCK10* | 225014090 | T/A | intron | c.123+28162A>T | _ | rs796399862 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *DOCK10* | 225019623 | G/A | intron | c.123+22629C>T | _ | rs1357877146 |
| *DOCK10* | 225023040 | -/T | intron | c.123+19211dup | _ | rs879839691 |
| *BZW1* | 200807620 | G/C | upstream | | _ | - |
| *BZW1* | 200823713 | -/A | 3' UTR | c.*1538dup | _ | rs1199092908 |
| *SMARCAL1* | 216437785 | -/TATA | intron | c.1645-623_1645-620dup | _ | rs3836032 |
| *SMARCAL1* | 216473366 | T/- | intron | c.2245-1890del | _ | rs74269760 |
| *SF3B1* | 197395632 | T/G | intron | c.3539+424A>C | _ | rs756320802 |
| *SF3B1* | 197431451 | A/C | intron | c.28+3521T>G | _ | - |
| *DNPEP* | 219382864 | G/A | intron | c.936+267C>T | _ | rs1015924721 |
| *ICOS* | 203965940 | G/A | downstream | _ | _ | rs537126499 |
| *ICOS* | 203965957 | C/G | downstream | _ | _ | rs932105844 |
| *ICOS* | 203965961 | A/G | downstream | _ | _ | rs1200900961 |
| *ICOS* | 203965962 | G/A | downstream | _ | _ | - |
| *ICOS* | 203965968 | C/T | downstream | _ | _ | - |
| *ICOS* | 203965969 | C/G | downstream | _ | _ | - |
| *ICOS* | 203965989 | A/G | downstream | _ | _ | rs950737459 |
| *VIL1* | 218458275 | -/AC | downstream | _ | _ | - |
| *WNT6* | 218876530 | T/G | downstream | _ | _ | rs1043802370 |
| *PLCL1* | 197868706 | T/- | intron | c.240+63377del | _ | rs959622220 |
| *PLCL1* | 197958156 | A/G | intron | c.241-125602A>G | _ | rs568219403 |
| *PLCL1* | 197973461 | C/T | intron | c.241-110297C>T | _ | - |
| *PLCL1* | 197985209 | A/T | intron | c.241-98549A>T | _ | - |
| *PLCL1* | 198041026 | G/A | intron | c.241-42732G>A | _ | rs561236256 |
| *PLCL1* | 198043568 | A/G | intron | c.241-40190A>G | _ | rs1331849303 |
| *PLCL1* | 198113766 | C/A | intron | c.3105+9830C>A | _ | rs527854471 |
| *CD28* | 203743063 | -/T | downstream | | _ | rs761945786 |
| *LANCL1* | 210465855 | G/C | intron | c.199+6104C>G | _ | rs778234896 |
| *TUBA4A* | 219257649 | AAAAAAAA/- | upstream | _ | _ | rs1158116997 |
| *SPEG* | 219446378 | C/T | intron | c.815+1217C>T | _ | - |
| *ABI2* | 203392292 | A/C | intron | c.560+1149A>C | _ | - |
| *ABI2* | 203392295 | A/C | intron | c.560+1152A>C | _ | - |
| *ABI2* | 203392298 | A/C | intron | c.560+1155A>C | _ | - |
| *ABI2* | 203392326 | CAACAACAACAA/- | intron | c.560+1190_560+1201del | – | rs150870167 |
| *ABI2* | 203430502 | G/A | 3' UTR | c.*3150G>A | _ | rs993751252 |
| *FARSB* | 222569954 | C/G | 3' UTR | c.*1917G>C | _ | - |
| *FARSB* | 222610976 | A/C | intron | c.1462+2835T>G | _ | - |
| *FARSB* | 222632840 | -/A | intron | c.715+358dup | _ | rs780438070 |
| *CNOT9* | 218600828 | AA/- | downstream | | _ | - |
| *GPR1* | 206175509 | A/C | 3' UTR | c.*671T>G | _ | rs536364692 |
| *GPR1* | 206179377 | TTT/- | intron | c.-28-2102_-28-2100del | – | - |
| *GPR1* | 206220145 | A/T | upstream | _ | _ | - |
| *GPR1* | 206220158 | G/T | upstream | _ | _ | - |

| | | | | | | |
|---|---|---|---|---|---|---|
| *ERBB4* | 211413305 | -/AAAAAAAAAAA | intron | c.3135+7135_3135+7136insTTTTTTTTTTT | _ | rs768289370 |
| *ERBB4* | 211415146 | T/C | intron | c.3135+5295A>G | _ | rs546521707 |
| *ERBB4* | 211510447 | G/A | intron | c.2487+51456C>T | _ | rs1385389801 |
| *ERBB4* | 211511993 | T/G | intron | c.2487+49910A>C | _ | - |
| *ERBB4* | 211512957 | T/C | intron | c.2487+48946A>G | _ | rs1011933747 |
| *ERBB4* | 211514852 | G/C | intron | c.2487+47051C>G | _ | - |
| *ERBB4* | 211538319 | C/T | intron | c.2487+23584G>A | _ | rs537421077 |
| *ERBB4* | 211632631 | TTG/- | intron | c.1947-2037_1947-2035del | – | rs911282360 |
| *ERBB4* | 211635752 | G/T | intron | c.1947-5158C>A | _ | - |
| *ERBB4* | 211698322 | A/T | intron | c.1489+3645T>A | _ | rs879628859 |
| *ERBB4* | 211957031 | T/C | intron | c.235-9415A>G | _ | rs538643805 |
| *ERBB4* | 212065032 | GTGTGTTGTG/- | intron | c.234+59720_234+59729del | – | rs1350369666 |
| *ERBB4* | 212116272 | TATATTAGATTGATAAATACTAGTCTATT/- | intron | c.234+8480_234+8508del | – | rs1219355699 |
| *ERBB4* | 212192047 | -/GTTATATGTTATATATGTTATATGTTATATTATGTTATGTTATGTTATATGTTATATATGTTATAA | intron | c.83-67145_83-67144insTTATAACATATAATATATAACATATATAACATATAACATATAACATATAACATATAACATATATAACATAAC | – | - |
| *ERBB4* | 212211629 | -/AAACC | intron | c.83-86727_83-86726insGGTTT | – | rs148929858 |
| *ERBB4* | 212264710 | T/G | intron | c.83-139807A>C | _ | rs373721395 |
| *ERBB4* | 212265829 | C/G | intron | c.83-140926G>C | _ | rs547430478 |
| *ERBB4* | 212327728 | TTTTTTC/- | intron | c.83-202825_83-202819del | – | rs1376674292 |
| *ERBB4* | 212329240 | -/A | intron | c.83-204338dup | _ | rs963480896 |
| *ERBB4* | 212394394 | C/T | intron | c.82+144055G>A | _ | - |
| *PTH2R* | 208419957 | A/- | intron | c.76-8239del | _ | rs1418612034 |
| *PTH2R* | 208447365 | G/A | intron | c.853+2478G>A | _ | - |
| *NDUFS1* | 206110067 | T/C | downstream | | _ | - |
| *NDUFS1* | 206125472 | AC/- | intron | c.2092+1067_2092+1068del | – | rs373368439 |
| *NDUFS1* | 206157963 | -/TTCT | intron | c.-5+1377_-5+1378insAGAA | – | rs539046435 |
| *AGFG1* | 227502182 | A/G | intron | c.261+10542A>G | _ | - |

| | | | | | | |
|---|---|---|---|---|---|---|
| *AGFG1* | 227507602 | -/AAAAAAAA | intron | c.262-12329_262-12322dup | – | - |
| *EPHA4* | 221545467 | -/C | intron | c.823+18263_823+18264insG | | rs1440429863 |
| *EPHA4* | 221545468 | T/A | intron | c.823+18263A>T | | rs2680855 |
| *EPHA4* | 221566951 | -/AGAG | intron | c.159+1766_159+1767insCTCT | | - |
| *TRIP12* | 229767930 | C/T | intron | c.5783-180G>A | | rs543800792 |
| *TRIP12* | 229857877 | G/A | intron | c.901+895C>T | _ | rs550103015 |
| *TRIP12* | 229878297 | G/A | intron | c.98+1685C>T | _ | - |
| *TRIP12* | 229878307 | A/G | intron | c.98+1675T>C | _ | - |
| *TRIP12* | 229878314 | A/G | intron | c.98+1668T>C | _ | - |
| *TRIP12* | 229878320 | -/GG | intron | c.98+1661_98+1662insCC | – | - |
| *TRIP12* | 229878328 | A/T | intron | c.98+1654T>A | _ | - |
| *TRIP12* | 229878330 | A/G | intron | c.98+1652T>C | _ | - |
| *STK17B* | 196157793 | G/A | intron | c.123-1142C>T | _ | rs545812634 |
| *STK17B* | 196161144 | G/C | intron | c.122+2118C>G | _ | rs554913101 |
| *STK17B* | 196164877 | A/T | intron | c.-44-1450T>A | _ | rs563427446 |
| *CLK1* | 200852191 | T/- | downstream | | _ | rs34007574 |
| *ATIC* | 215336737 | A/G | intron | c.1098+613A>G | _ | rs1186194306 |
| *ATIC* | 215338653 | A/- | intron | c.1099-123del | _ | rs532282971 |
| *ADAM23* | 206457427 | A/T | intron | c.432+11903A>T | _ | - |
| *ADAM23* | 206466589 | C/T | intron | c.433-14643C>T | _ | rs910071009 |
| *ADAM23* | 206523233 | G/C | intron | c.510-7652G>C | _ | rs1043631215 |
| *ADAM23* | 206556043 | C/T | intron | c.934-1384C>T | _ | - |
| *ADAM23* | 206608566 | ATT/- | intron | c.2360-1328_2360-1326del | – | rs545632632 |
| *KLF7* | 207078821 | C/A | 3' UTR | c.*2392G>T | _ | rs540095240 |
| *KLF7* | 207081984 | C/T | intron | c.858-720G>A | _ | rs564244373 |
| *KLF7* | 207126811 | AA/- | intron | c.103-2407_103-2406del | – | rs35099385 |
| *CUL3* | 224476437 | G/A | intron | c.2175+1763C>T | _ | rs567463315 |
| *CUL3* | 224479626 | G/A | intron | c.2030-1281C>T | _ | - |
| *CUL3* | 224511266 | A/G | intron | c.883+88T>C | _ | rs188803950 |
| *CUL3* | 224526808 | C/T | intron | c.378+8720G>A | _ | - |
| *CUL3* | 224585358 | G/C | 5' UTR | c.-349C>G | _ | rs1020886904 |
| *SUMO1* | 202211967 | -/A | intron | c.166-1162dup | _ | rs1157149339 |
| *NDUFB3* | 201075182 | AAA/- | intron | c.-3+3134_-3+3136del | – | rs752371143 |
| *NDUFB3* | 201085924 | -/T | downstream | | _ | rs879723391 |
| *DES* | 219420899 | A/T | missense | c.969A>T | p.Glu323Asp | - |
| *CPS1* | 210575641 | A/G | intron | c.237-705A>G | __ | rs957479748 |
| *CPS1* | 210645644 | G/A | intron | c.3142-2219G>A | | rs887005019 |
| *ACADL* | 210192812 | T/C | synonymous | c.1191A>G | p.Pro397%3D | rs544440334 |
| *MREG* | 216030730 | T/- | intron | c.-68+2953del | _ | rs71047960 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *SLC19A3* | 227708874 | C/T | upstream | _ | _ | rs563376551 |
| *NYAP2* | 225655937 | ACATACACACAT/- | intron | c.1932+4394_1932+4405del | _ | rs1245298448 |
| *NYAP2* | 225681946 | G/A | intron | c.1933-16371G>A | _ | rs954124071 |
| *KIAA2012* | 202114952 | G/A | intron | c.1759+1506G>A | _ | - |
| *KIAA2012* | 202120549 | C/T | intron | c.1760-4665C>T | _ | rs563098661 |
| *KIAA2012* | 202141117 | T/A | intron | c.1905+2609T>A | _ | rs569758981 |
| *KIAA2012* | 202185846 | C/T | intron | c.2207+1003C>T | _ | - |
| *KIAA2012* | 202199922 | TT/- | intron | c.3405-2481_3405-2480del | _ | rs71025287 |
| *ANKRD44* | 196995812 | A/G | intron | c.2749-351T>C | _ | rs528804910 |
| *ANKRD44* | 197028379 | C/T | intron | c.1651-3112G>A | _ | rs1290966101 |
| *ANKRD44* | 197098319 | C/T | intron | c.1100+1497G>A | _ | - |
| *MAP2* | 209570865 | G/A | intron | c.-171-9171G>A | _ | rs1267688410 |
| *FTCDNL1* | 199824664 | G/A | intron | c.212-4907C>T | _ | rs558259520 |
| *TUBA4B* | 219257649 | AAAAAAAA/- | intron | c.12+4240_12+4247del | _ | rs1158116997 |
| *TUBA4B* | 219259106 | C/G | intron | c.12+5687C>G | _ | rs553998533 |
| *TUBA4B* | 219259124 | -/AATAAATA | intron | c.12+5745_12+5752dup | _ | rs58291199 |
| *TUBA4B* | 219262685 | -/T | intron | c.13-3825dup | _ | rs993144498 |
| *CCDC150* | 196737344 | C/A | downstream | | _ | - |
| *CCDC150* | 196693586 | G/A | intron | c.514-1460G>A | _ | rs553964936 |
| *FTCDNL1* | 199780630 | C/T | intron | c.212-19795G>A | _ | rs1336697362 |
| *TRIP12* | 229927472 | A/G | upstream | | _ | rs536688627 |
| *VWC2L* | 214505634 | T/C | intron | c.391-70038T>C | _ | rs529240222 |
| *VWC2L* | 214543140 | A/T | intron | c.391-32532A>T | _ | rs114741519 |
| *MDH1B* | 206758380 | A/G | intron | c.-159-1009T>C | _ | rs776766328 |
| *MDH1B* | 206766380 | G/A | upstream | | _ | - |
| *PID1* | 229228062 | T/C | splice region, 5' UTR | c.-118A>G | _ | rs781045636 |
| *METTL21A* | 207587468 | C/T | intron | c.260-5522G>A | _ | - |
| *KCTD18* | 200509966 | T/C | intron | c.-681+45A>G | _ | - |
| *KCTD18* | 200509973 | C/G | intron | c.-681+38G>C | _ | - |
| *DNPEP* | 219404068 | C/T | upstream | _ | _ | rs528794126 |
| *EPHA4* | 221578799 | C/T | upstream | _ | _ | - |
| *TMEM198* | 219551155 | G/A | downstream | _ | _ | rs1165309466 |
| *SPATS2L* | 200308628 | ACAC/- | upstream | _ | _ | - |
| *BARD1* | 214801901 | C/T | intron | c.159-4784G>A | _ | rs941133188 |
| *TUBA4A* | 219259106 | C/G | upstream | | _ | rs553998533 |
| *TUBA4A* | 219259124 | -/AATAAATA | upstream | _ | _ | rs58291199 |
| *MFF* | 227328321 | AAAAAC/- | upstream | _ | _ | rs1182644669 |
| *BMPR2* | 202410897 | A/G | intron | c.76+33347A>G | _ | rs528082988 |
| *BMPR2* | 202424376 | -/A | intron | c.77-40420dup | _ | rs55785585 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *BMPR2* | 202430953 | -/T | intron | c.77-33855dup | _ | rs1248490113 |
| *BMPR2* | 202470773 | AAA/- | intron | c.418+3098_418+3100del | _ | - |
| *BMPR2* | 202502383 | AGAC/- | intron | c.419-11336_419-11333del | _ | rs895918851 |
| *BMPR2* | 202561288 | T/- | 3' UTR | c.*1346del | _ | rs1267490196 |
| *HSPE1-MOB4* | 197545689 | T/C | intron | c.463-2647T>C | _ | rs546609847 |
| *HSPE1-MOB4* | 197552917 | C/- | 3' UTR | c.*2271del | _ | rs1459274442 |
| *OBSL1* | 219551155 | G/A | downstream | _ | _ | rs1165309466 |
| *SATB2* | 199466944 | T/C | upstream | _ | _ | rs539141538 |
| *SATB2* | 199461522 | -/A | upstream | _ | _ | rs751358455 |
| *AOX1* | 200589016 | A/C | intron | c.45+2863A>C | _ | rs550101234 |
| *AOX1* | 200623149 | GAGA/- | intron | c.2002-711_2002-708del | _ | rs1371086457 |
| *AOX1* | 200675003 | C/T | downstream | | _ | - |
| *SPHKAP* | 228181663 | G/C | 5' UTR | c.-65C>G | _ | rs547462934 |
| *NBEAL1* | 203034161 | TT/- | intron | c.52-7590_52-7589del | _ | rs1166429847 |
| *NBEAL1* | 203063286 | -/A | intron | c.516-5095dup | _ | rs1014461622 |
| *NBEAL1* | 203188894 | C/T | intron | c.6736+305C>T | _ | rs376059753 |
| *NBEAL1* | 203191320 | A/- | intron | c.6834+946del | _ | rs200659518 |
| *C2orf80* | 208176926 | C/T | intron | c.366+3819G>A | _ | rs1369897624 |
| *C2orf80* | 208176928 | T/C | intron | c.366+3817A>G | _ | rs112610733 |
| *C2orf80* | 208186788 | A/G | intron | c.41+158T>C | _ | rs548221565 |
| *DYTN* | 206672262 | C/A | intron | c.981-6233G>T | _ | rs544480461 |
| *DYTN* | 206714944 | -/T | intron | c.19+3316dup | _ | rs373615386 |
| *PLEKHM3* | 207858013 | TATTT/- | intron | c.2108+3092_2108+3096del | _ | rs577314868 |
| *PLEKHM3* | 207861771 | -/A | intron | c.1951-510dup | _ | rs548566889 |
| *PLEKHM3* | 207865650 | TGT/- | intron | c.1951-4388_1951-4386del | _ | - |
| *PLEKHM3* | 207866360 | T/C | intron | c.1951-5098A>G | _ | rs762602890 |
| *PLEKHM3* | 207906305 | C/T | intron | c.1950+2209G>A | _ | rs76792136 |
| *PLEKHM3* | 207918797 | G/T | intron | c.1887-10220C>A | _ | rs79954436 |
| *PLEKHM3* | 207921035 | C/T | intron | c.1886+9891G>A | _ | rs115965627 |
| *PLEKHM3* | 207929452 | T/C | intron | c.1886+1474A>G | _ | rs75773720 |
| *PLEKHM3* | 207945669 | G/C | intron | c.1692+698C>G | _ | - |
| *PLEKHM3* | 208029591 | C/T | upstream | | _ | rs116698317 |
| *TYW5* | 199941268 | T/C | intron | c.304-1135A>G | _ | - |
| *TYW5* | 199943135 | C/T | intron | c.303+630G>A | _ | rs543434169 |
| *CYP27A1* | 218807149 | C/T | intron | c.256-2428C>T | _ | rs976474940 |
| *CXCR1* | 218158726 | T/C | downstream | _ | _ | rs552515431 |
| *SLC11A1* | 218401413 | G/A | downstream | _ | _ | rs1412406093 |
| *COL4A4* | 227013564 | A/T | intron | c.4217-1267T>A | _ | rs566858863 |
| *COL4A4* | 227051629 | T/C | intron | c.2969-471A>G | _ | rs561425659 |
| *COL4A4* | 227110718 | T/C | intron | c.594+960A>G | _ | rs1008510680 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *COL4A4* | 227156483 | G/A | intron | c.-102+7524C>T | _ | rs552648978 |
| *COL4A3* | 227190998 | TCT/- | intron | c.87+26189_87+26191del | _ | - |
| *COL4A3* | 227223565 | C/G | intron | c.88-14403C>G | _ | - |
| *TNS1* | 217957826 | G/T | intron | c.111+20939C>A | _ | - |
| *TNS1* | 217957831 | G/T | intron | c.111+20934C>A | _ | - |
| *TNS1* | 217966321 | -/CA | intron | c.111+12443_111+12444insTG | _ | rs1163661293 |
| *TNS1* | 218009624 | C/T | intron | c.157-18568G>A | _ | rs531800696 |
| *TNS1* | 218021124 | C/T | intron | c.156+12696G>A | _ | rs561786150 |
| *SCYGR3* | 227610047 | -/AAAAAAAAAAAAAAAAAAAAAAAAA | downstream | _ | _ | _ |
| *BMPR2* | 202561288 | T/- | downstream | | _ | rs1267490196 |
| *CCDC195* | 224709233 | C/A | intron | c.482+740G>T | _ | - |
| *MREG* | 215944002 | C/A | intron | c.*752+109G>T | _ | - |
| *STK17B* | 196157793 | G/A | upstream | _ | _ | rs545812634 |
| *STK17B* | 196161144 | G/C | upstream | _ | _ | rs554913101 |
| *HSPE1-MOB4* | 197552917 | C/- | downstream | _ | _ | rs1459274442 |
| *DOCK10* | 224770741 | C/T | downstream | _ | _ | - |
| *MFF* | 227357595 | C/T | downstream | _ | _ | - |
| *DNPEP* | 219382864 | G/A | downstream | _ | _ | rs1015924721 |
| *GPR1* | 206175509 | A/C | downstream | _ | _ | rs536364692 |
| *KLF7* | 207078821 | C/A | downstream | _ | _ | rs540095240 |
| *STRADB* | 201403793 | T/C | intron | c.28+15718T>C | _ | - |
| *STRADB* | 201404547 | T/C | intron | c.28+16472T>C | _ | rs556781523 |
| *RAPH1* | 203392292 | A/C | downstream | _ | _ | - |
| *RAPH1* | 203392295 | A/C | downstream | _ | _ | - |
| *RAPH1* | 203392298 | A/C | downstream | _ | _ | - |
| *RAPH1* | 203392326 | CAACAACAACAA/- | downstream | _ | _ | rs150870167 |
| *RAPH1* | 203430502 | G/A | intron | c.1776+14366C>T | _ | rs993751252 |
| *AGFG1* | 227507602 | -/AAAAAAAA | upstream | _ | _ | - |
| *RFTN2* | 197620188 | A/- | upstream | _ | _ | rs1159201273 |
| *RFTN2* | 197622644 | C/G | upstream | _ | _ | - |
| *ABI2* | 203430502 | G/A | downstream | _ | _ | rs993751252 |
| *TNS1* | 217890215 | A/G | upstream | _ | _ | - |
| *BZW1* | 200823713 | -/A | downstream | _ | _ | rs1199092908 |
| *PIKFYVE* | 208314367 | C/T | synonymous | c.1602C>T | p.Asn534%3D | rs61752185 |
| *PIKFYVE* | 208336395 | T/C | downstream | _ | _ | rs140875159 |
| *TNS1* | 217893400 | -/CGCGCA | downstream | _ | _ | rs58297965 |
| *STK17B* | 196161144 | G/C | downstream | _ | _ | rs554913101 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *METTL21A* | 207624749 | A/C | upstream | _ | _ | rs539918048 |
| *MOB4* | 197552917 | C/- | downstream | _ | _ | rs1459274442 |
| *RPL37A* | 216536326 | T/A | intron | c.215+36295T>A | _ | rs531431590 |
| *PPIL3* | 200873195 | -/T | downstream | _ | _ | rs533187117 |
| *PPIL3* | 200875561 | C/T | downstream | _ | _ | rs545932195 |
| *MREG* | 215944002 | C/A | downstream | _ | _ | - |
| *AOX1* | 200675003 | C/T | intron | c.487-1875C>T | _ | - |
| *TNS1* | 217890215 | A/G | downstream | _ | _ | - |
| *SPATS2L* | 200390868 | A/C | upstream | _ | _ | rs578231219 |
| *SLC16A14* | 230058028 | -/A | downstream | _ | _ | rs35641768 |
| *CUL3* | 224511266 | A/G | downstream | _ | _ | rs188803950 |
| *MAIP1* | 199976156 | T/C | intron | c.627+16276T>C | _ | rs555192058 |
| *TRIP12* | 229857877 | G/A | downstream | _ | _ | rs550103015 |
| *CARF* | 202953692 | G/A | downstream | _ | _ | - |
| *SPATS2L* | 200390868 | A/C | downstream | _ | _ | rs578231219 |
| *C2orf80* | 208186788 | A/G | upstream | _ | _ | rs548221565 |
| *TUBA4A* | 219257649 | AAAAAAAA/- | intron | c.-42-5419_-42-5412del | – | rs1158116997 |
| *TUBA4A* | 219259106 | C/G | intron | c.-42-6876G>C | _ | rs553998533 |
| *TUBA4A* | 219259124 | -/AATAAATA | intron | c.-42-6902_-42-6895dup | – | rs58291199 |
| *TUBA4A* | 219262685 | -/T | intron | c.-42-10456dup | _ | rs993144498 |
| *PGAP1* | 196834398 | G/C | downstream | _ | _ | rs567059943 |
| *FASTKD2* | 206770731 | AA/- | downstream | _ | _ | rs549037266 |
| *TRIP12* | 229767930 | C/T | downstream | _ | _ | rs543800792 |
| *SGO2* | 200509966 | T/C | upstream | _ | _ | - |
| *SGO2* | 200509973 | C/G | upstream | _ | _ | - |
| *SGO2* | 200513132 | G/C | intron | c.-120+2866G>C | _ | - |
| *SGO2* | 200513133 | G/C | intron | c.-120+2867G>C | _ | - |
| *SGO2* | 200513134 | G/C | intron | c.-120+2868G>C | _ | - |
| *CDK15* | 201792180 | T/A | intron | c.-150+1520T>A | _ | - |
| *SPATS2L* | 200474104 | T/- | downstream | _ | _ | rs559587714 |
| *RHBDD1* | 226902323 | A/G | upstream | _ | _ | rs991870599 |
| *SLC4A3* | 219627530 | CGG/- | intron | c.-94+66_-94+68del | – | rs1184272855 |
| *TNS1* | 217855558 | TC/- | downstream | _ | _ | rs3838562 |
| *CCDC140* | 222302941 | GT/- | intron | c.-157-803_-157-802del | – | - |
| *CCDC140* | 222303511 | G/A | intron | c.-157-235G>A | _ | - |
| *MARCHF4* | 216348069 | C/T | intron | c.516+21676G>A | _ | rs796100931 |
| *MARCHF4* | 216371874 | -/G | 5' UTR | c.-1615dup | _ | - |

## A 3.2 Primer pairs for the PCR- and Sanger-based sequencing experiments

| *SCT135 NGS variant primers* | |
|---|---|
| **Name** | **Sequence (5'-3')** |
| ACADL-Ex10-F | ctggttttgaagcacctc |
| ACADL-Ex10-R | agattgcacacctcctttcc |
| SLC19A3-Ex2-F | catgcaaaacccagcagag |
| SLC19A3-Ex2-R | tgcttacctccaaaggttgc |
| DNAH7-Ex44-F | ggtgggactgaggagaaaag |
| DNAH7-Ex44-R | acccatttgcacagaccttc |
| STK16-Ex7-F | tgcagaaccaactcagcatc |
| STK16-Ex7-R | gatggaatgggtggtgattc |
| OBSL1-Ex5-F | tggtgttccacggttctg |
| OBSL1-Ex5-R | gcaggatgagtctgtgctg |
| COL4A4-Ex21-F | tctgccattgatcctctgtc |
| COL4A4-Ex21-R | tgtcacctgctgtcgtgtaag |
| CXCR2-Ex1-F | cctatcccagtttcttgagtgg |
| CXCR2-Ex1-R | aggttcagcaggtagacatcag |
| HECW2-Ex9-F | cgttatgggtgagggatgag |
| HECW2-Ex9-R | ggaatgggtctgcctgtc |
| SPATS2L-Ex4-F | ctccccagtggcaggataac |
| SPATS2L-Ex4-R | aacagcagccccaatgag |
| CCNYL1-5'UTR-F | gggaactctgaccatgattg |
| CCNYL1-5'UTR-R | ccctctccgaaatccaactc |
| GPBAR1-5'UTR-I-F | ccgagctggagtagggaaac |
| GPBAR1-5'UTR-I-R | ggctgtaccacctgcacac |
| GPBAR1-5'UTR-IIa-F | tcctcttccctctcctctgg |
| GPBAR1-5'UTR-IIa-R | ccacccagccagctttatc |
| GPBAR1-5'UTR-IIb-F | gctggctggaagaccactac |
| GPBAR1-5'UTR-IIb-R | gcagtaggctcaggaagaagc |
| GPBAR1-Ex1-F | ccctgctctttgccagtc |
| GPBAR1-Ex1-R | gacactgctttggctgcttg |
| AAMP-Ex4-F | gggccagataccagaatcac |
| AAMP-Ex4-R | ggggactttccaggtagcag |
| PAX3-5'UTR-Ex1-F | ccaatcagcgcgtgtctc |
| PAX3-5'UTR-Ex1-R | cctggaagcaccaaaggag |
| PAX3-Ex2-F | tgaaaggaggtctgggtctg |
| PAX3-Ex2-R | cagatgtcagccgttaccc |
| DNAH7-Ex62-F | ggcagctatgtgaatgacttcc |
| DNAH7-Ex62-R | agttctgggcaatgaagc |
| HECW2-Ex11-F | gttccccaaacaatggacag |
| HECW2-Ex11-R | ccagtatcccacagatcaag |
| TRAK2-Ex11-F | gctgagattgaggggactatg |
| TRAK2-Ex11-R | gctggaatgtgcgtatgtg |
| ALS2-Ex4-F | ctccctccctttttactgtg |
| ALS2-Ex4-R | gtttagggctgaggtgcttg |
| CDK15-Ex8-F | ctgccaagaaggccacatc |
| CDK15-Ex8-R | aagcaaccctcgtgagtagg |
| IDH1-Ex6-F | taggcagttggaccctgaac |
| IDH1-Ex6-R | actgagcagccaagggaac |
| UNC80-Ex44-F | tagccagtcttgcagcag |
| UNC80-Ex44-R | gatggtagccaactcccttg |

| *SCT135 NGS variant primers* | |
|---|---|
| **Name** | **Sequence (5'-3')** |
| WDR69-3'UTR-R | gcactgggattacaggtgtg |
| MOB4-Ex6-F | gacacacacttgatggtgctg |
| MOB4-Ex6-R | agactgaactggccttgagc |
| ORC2-3'UTR-a-F | aaggaagaagaggaggcttg |
| ORC2-3'UTR-a-R | aggagaaagctggatcactcac |
| ORC2-3'UTR-b-F | ttttaacactccccgagctg |
| ORC2-3'UTR-b-R | caacagtagccaaggtgagc |
| ORC2-3'UTR-c-F | ttgtctgtgccatgaccttc |
| ORC2-3'UTR-c-R | ggcctcccaaagtgttgtg |
| WDR12-5'UTR-F | agtctactcaacctcgtcttgg |
| WDR12-5'UTR-R | cagaccacaaacatcgactacc |
| CYP20A1-5'UTR-a-F | gccatctcggctcactgtag |
| CYP20A1-5'UTR-a-R | gccagaggaaagagcaatgg |
| CYP20A1-5'UTR-b-F | agtcgcgggcatattctg |
| CYP20A1-5'UTR-b-R | cgctcaccggatagaggtag |
| FASTKD2-Ex7-F | gttccccagacagttcgtg |
| FASTKD2-Ex7-R | cctttccacagcaacacctg |
| RPE-5'UTR-F | agtcaagagccgaggagagg |
| RPE-5'UTR-R | ggagttacccgtccattacg |
| C2orf67-Ex8-F | gagtgaaacagcattaggtagg |
| C2orf67-Ex8-R | agggaacataggagggaaatac |
| CPS1-Ex31-F | ggccaggttatctctcatgc |
| CPS1-Ex31-R | tctccaggtgtctgtgatgc |
| SPAG16-Ex15-F | gaaaatgcttgtctgaaacg |
| SPAG16-Ex15-R | aaattcacctcattgccaggac |
| TNS1-Ex3-F | ggcttctgccctcttctcac |
| TNS1-Ex3-R | gctgatgtcccctacacttc |
| ARPC2-5'UTR-Ex1-F | gcatcacacatacgcaagc |
| ARPC2-5'UTR-Ex1-R | gacaacattggggattagctg |
| CTDSP1-Ex7-F | ctgggcaacagagcaagac |
| CTDSP1-Ex7-R | cgaagaaggggaggaggtc |
| PRKAG3-Ex4-F | actacctggggcgactctc |
| PRKAG3-Ex4-R | cctctgtggctgggaactc |
| ABCB6-Ex17-F | ccttcctgccttgtgtatgc |
| ABCB6-Ex17-R | tctcttaccgtcccctctcc |
| ATG9A-5'UTR-I-F | agcagcgaagaggacaacc |
| ATG9A-5'UTR-I-R | ctcccaacagcggacaac |
| ATG9A-Ex1-F | caaggaggctggtagtggac |
| ATG9A-Ex1-R | aggtgctgtgggataggaac |
| STK16-5'UTR-I-F | acccgtccacccctacac |
| STK16-5'UTR-I-R | ggtgagtgggctcttccag |
| SPEG-Ex20-F | agaggcgtggttaggaggag |
| SPEG-Ex20-R | accagggagcactcattctc |
| SPEG-Ex21-F | tctgtccacctgtcccagtc |
| SPEG-Ex21-R | ccgcaagtaggagaaagcac |
| ACCN4-Ex5-F | gaccacccttctccaactcc |
| ACCN4-Ex5-R | cacactcacacacccacaag |
| OBSL1-Ex8-F | cagttccacccttcataccg |

| | |
|---|---|
| XRCC5-Ex7-F | gccatctcctgctgtttctc |
| XRCC5-Ex7-R | tctcctagcctgccttttacac |
| PRKAG3-UTR3'-I-F | agacgctctctccctcagtc |
| PRKAG3-UTR3'-I-R | cagttttcacagggctgctc |
| CUL3-Ex6-F | aggtggttgaaagggaactc |
| CUL3-Ex6-R | ctcactgggctcaccaatac |
| DOCK10-Ex54-F | tcatgccactgcactctagc |
| DOCK10-Ex54-R | attcttgggggattgaggac |
| DOCK10-Ex28-F | atgagaggccacagagaagg |
| DOCK10-Ex28-R | cttgcatccaccaacatgg |
| COL4A4-Ex46-F | cagggaggtgtcagaaatgg |
| COL4A4-Ex46-R | gtgaatgagccagggttttc |
| WDR69-3'UTR-F | gcgttgactgaaggaagctg |

| | |
|---|---|
| OBSL1-Ex8-R | cctcctctccatccttggtc |
| PAX3-3'UTR-a-F | aaggcaatggtttcacatgg |
| PAX3-3'UTR-a-R | ctggaaaaacgtcacacacc |
| CUL3-3'UTR-a-F | gcctttactccctttgagag |
| CUL3-3'UTR-a-R | caccaccctatacaatccac |
| CUL3-Ex2-F | tctgctccaggtaagattgg |
| CUL3-Ex2-R | gtccagtgtagagcttttctcc |
| COL4A4-Ex26-F | aatcgggataggcaacatcc |
| COL4A4-Ex26-R | gctttccttctctggcaaac |
| AGFG1-Ex10-F | cctgtgtccatgtgttgttc |
| AGFG1-Ex10-R | gatgaagcaggctgtgtctg |
| PID1-3'UTR-a-F | attggtcacggagttcaagg |
| PID1-3'UTR-a-R | ggctgtgctgatttgcatac |

| ***SCT135 coding region gaps coverage*** | |
|---|---|
| **Name** | **Sequence (5'-3')** |
| BZW1-Trns1,2-Ex1-F | ttctggctctttcctcttcg |
| BZW1-Trns1,2-Ex1-R | cggtaacgtgtctcccaatc |
| PPIL3-Ex3-F | ctgtgagaggacacccaaaac |
| PPIL3-Ex3-R | aaaaggccaggtgcagtg |
| TMEM237-Trns1-Ex1-F | gactcgttggtcctggagag |
| TMEM237-Trns1-Ex1-R | cccttagtgattcccagctc |
| ALS2-Ex15-F | ggcacctacatgggcttaac |
| ALS2-Ex15-R | tctctgctctgtgctgaacg |
| SUMO1-Ex1-F | gaaggagctgacaaaactgc |
| SUMO1-Ex1-R | agaagtgggacgacatgagg |
| FAM117B-Ex1a-F | cttcgtcaccccgtcttg |
| FAM117B-Ex1a-R | gctggggctcctcttctc |
| FAM117B-Ex1b-F | acaacggtggctgctgtg |
| FAM117B-Ex1b-R | ctcatcagctcaagcccaag |
| ABI2-Ex1-F | taggagacgccggaagtg |
| ABI2-Ex1-R | attcccacccattcaccag |
| RAPH1-Ex13-F | caccaccaacttctgcatcc |
| RAPH1-Ex13-R | ggcactgaactgtggaagg |
| PARD3B-Ex1-F | ggcagtttcgctttggtg |
| PARD3B-Ex1-R | ttctgcatcccctgagtgtc |
| PARD3B-Ex4-F | ctggggaaagatgagagttg |
| PARD3B-Ex4-R | ccttcctctgcctcaatttc |
| NDUFS1-5'UTR-F | cgggtccaagttgtccttc |
| NDUFS1-5'UTR-R | cctaagtcatcggacactgg |
| ADAM23-Ex1-F | acccctggactcctctgc |
| ADAM23-Ex1-R | cgagaagggtggaaagacag |
| FZD5-Ex-1a-F | ctggaatccgagccctaac |
| FZD5-Ex-1a-R | cggttgtaatccatgcagag |
| FZD5-Ex-1b-F | ctgcgcttcttcctatgctc |
| FZD5-Ex-1b-R | acaggtagcaggctgacagg |
| FZD5-Ex-1c-F | atccgcagcgtcatcaag |
| FZD5-Ex-1c-R | ccctcttccctctctccaag |
| UNC80-Ex1-F | acagtgggaggtgctgaaag |
| UNC80-Ex1-R | gtgcagagggcttgttttcc |
| UNC80-Ex60-F | gacaaacccagcacaaggtg |
| UNC80-Ex60-R | taagacggggcgagaagag |

| ***SCT135 coding region gaps coverage*** | |
|---|---|
| **Name** | **Sequence (5'-3')** |
| CCDC108-Ex6-F | tgcccagcatccctgact |
| CCDC108-Ex6-R | gaacagcccctcgaagaag |
| CCDC108-Ex15-F | cccgtggggatataggattg |
| CCDC108-Ex15-R | aagggcaagtttggggtatc |
| CCDC108-Ex32-F | gtagtggatgaggggagtcg |
| CCDC108-Ex32-R | ggtggagtacctctgcttgc |
| CCDC108-Ex33-F | tctgcgtgcccaggtaag |
| CCDC108-Ex33-R | tgctccctgccatccttc |
| IHH-Ex1-F | gggccggcctatttattg |
| IHH-Ex1-R | gtaattgggggtgagctcct |
| C2orf24-Ex1-F | acgcttgacacctcccttc |
| C2orf24-Ex1-R | ccctacgcccattcctatac |
| FAM134A-Ex1-F | ctgtgtaggcgcagtgtcag |
| FAM134A-Ex1-R | gaaaggggaatggtgctctc |
| ZFAND2B-Ex1-F | gtcttccgactcagccttct |
| ZFAND2B-Ex1-R | cttcccaagcccattacctc |
| ABCB6-Ex5-F | cctttcacatcctggtgctg |
| ABCB6-Ex5-R | cggggtctgttctcttcctc |
| TUBA4A-Ex1-F | ccagcgtgtctgctcaaaac |
| TUBA4A-Ex1-R | aaaaccctcgcacctcctg |
| PTPRN-Ex1-F | gaagagggctacagggatgg |
| PTPRN-Ex1-R | ctcccccaacccatattctc |
| PTPRN-Ex19-F | gtggttggagagaccgagaa |
| PTPRN-Ex19-R | gtggaggagaaggagccagt |
| PTPRN-Ex22-F | gttttctgggtcctgtcctc |
| PTPRN-Ex22-R | acctgtgctctgccctcaag |
| RESP18-Ex1-F | tagaggtcgagcggaggtt |
| RESP18-Ex1-R | cgcctgtactcccagtccta |
| RESP18-Ex2-F | gagtacaggcggtggttctt |
| RESP18-Ex2-R | ctctttccgctagacgctgt |
| RESP18-Ex7-F | actggttgccctggtgct |
| RESP18-Ex7-R | agttgtgctcccctcatcc |
| DNPEP-Ex1-F | acgaggaagcttgacgagg |
| DNPEP-Ex1-R | ggtttcgctttgggtcagg |
| DES-Ex1a-F | caggacagcgggatcttg |
| DES-Ex1a-R | gcgcaccttctcgatgtagt |

| | |
|---|---|
| ACADL-Ex1-F | gtatttgggggctccatagc |
| ACADL-Ex1-R | gctgacacccctttttcctc |
| SPAG16-Ex1-F | tatcttgtccgctcccagag |
| SPAG16-Ex1-R | tggctctcaggaagactgtg |
| BARD1-Ex1-F | ccctgcgagtccctattttg |
| BARD1-Ex1-R | gggaacggaaggaggaaac |
| ATIC-Ex1-F | gtggagtggcctcactttg |
| ATIC-Ex1-R | ggacgctggctttcaatc |
| FN1-Ex1-F | accttcttggaggcgacaac |
| FN1-Ex1-R | cacaaaacttcagccccaac |
| FN1-Ex42-F | tatgtggtgtttgcgctgtg |
| FN1-Ex42-R | agcagttgtatgccaacagg |
| MREG-Ex1-F | gtgccctgggattttgag |
| MREG-Ex1-R | acctccccaactcacacaag |
| March4-Ex1-F | cccacaacacagatccactg |
| March4-Ex1-R | taccactgtccaagctgctg |
| SMARCAL1-Ex14-F | gaaatggggtggagaggaac |
| SMARCAL1-Ex14-R | ctgcaagccaccagtgaac |
| IGFBP2-Ex1-F | ggggaagggagtggtctc |
| IGFBP2-Ex1-R | cccctaaaaccgtccgtaag |
| IGFBP2-Ex2-F | tcatcattacggtccaggtg |
| IGFBP2-Ex2-R | tactgactgccccaaaggtc |
| IGFBP5-Ex1-F | cctcttggcccctttatcc |
| IGFBP5-Ex1-R | aaggaccctccccgactac |
| TNS1-Ex15-F | ctcactaagcgtgccctcac |
| TNS1-Ex15-R | cacaaagctggctgctgac |
| TNS1-Ex18-F | tccaggaagctgtgttccag |
| TNS1-Ex18-R | gccataggcaacagatccag |
| TNS1-Ex19-F | caccagatgatgggtccac |
| TNS1-Ex19-R | cgggaaccacagatccag |
| RUFY4-Ex8-F | tctcgggaatgttgttcagg |
| RUFY4-Ex8-R | tgggaggttctcacaagacc |
| AAMP-Ex9-F | gatggggaaaggggtttg |
| AAMP-Ex9-R | tctggggtagatgctcctg |
| PNKD-Ex1-Trns1-F | gcggagagaacccaaactc |
| PNKD-Ex1-Trns1-R | cctgacctcctgctatcgtc |
| PNKD-Ex1-Trns2-F | ctgtggaccccgatcagc |
| PNKD-Ex1-Trns2-R | gtctcctcgatccctcttcc |
| PNKD-Ex6-Trns1-F | cctggagatgctgtggtaaag |
| PNKD-Ex6-Trns1-R | agatccacctgctgataccc |
| TMBIM1-Ex1-F | ctggacaaggctggaagtg |
| TMBIM1-Ex1-R | cccattctcctgtggtgtg |
| C2ORF62-Ex10-F | taggggctagaaggctccag |
| C2ORF62-Ex10-R | aggtacgacggtaacggttc |
| CTDSP1-Ex1-Trns1-F | gggaaggaaactccatgttg |
| CTDSP1-Ex1-Trns1-R | ttcaaactctctccctgctc |
| CTDSP1-Ex1-Trns3-F | ggtagacccgaagcacgtc |
| CTDSP1-Ex1-Trns3-R | aacacccctcggcacttc |
| CYP27A1-Ex1-F | gcccagagttcagaccaagc |
| CYP27A1-Ex1-R | gctgtcctagacgtggaatctc |
| PRKAG3-Ex6-F | tctcagcacaaggacactgg |
| PRKAG3-Ex6-R | cccaccatcaccaacagc |

| | |
|---|---|
| DES-Ex1b-F | gctgctggacttctcactgg |
| DES-Ex1b-R | acaggtggaggacccttttct |
| SPEG-Ex1-F | cccccagacttgtctccta |
| SPEG-Ex1-R | ggcggaaccagtatccagta |
| SPEG-Ex3-F | ctggggtgtacaaagagcag |
| SPEG-Ex3-R | atgggtggaggctgactg |
| SPEG-Ex4a-F | acccaattcctgtcacaagc |
| SPEG-Ex4a-R | acttgggctgctccagaga |
| SPEG-Ex4b-F | gacaagctgcagttcttcgag |
| SPEG-Ex4b-R | attggtcccgacgcctaac |
| SPEG-Ex4c-F | cctccacccccaagacat |
| SPEG-Ex4c-R | gcagtcttgctctccacaaa |
| SPEG-Ex6-F | aggtttggtctcctgtgtgg |
| SPEG-Ex6-R | tcaaggctgagagtgggaag |
| SPEG-Ex8-F | tgctcccattcaaaccctct |
| SPEG-Ex8-R | gaagcccaccaagattccat |
| SPEG-Ex30a-F | gttccctgaccctctgcat |
| SPEG-Ex30a-R | catactcgccctcacccag |
| SPEG-Ex30b-F | ctccctctcaggaccagga |
| SPEG-Ex30b-R | taggtgtggtggcagaaggt |
| SPEG-Ex30c-F | gagtctccttccctgtctgc |
| SPEG-Ex30c-R | gtatatgccatcctcctcctc |
| SPEG-Ex30d-F | tcgaggccaagttcaagc |
| SPEG-Ex30d-R | ctcctcccacagcctctctc |
| SPEG-Ex41-F | agactcactgtccccattcc |
| SPEG-Ex41-R | gtaagagcccagccagatgt |
| GMPPA-Ex10-F | gtggtcggatggatggag |
| GMPPA-Ex10-R | gcgttcacccttctctgttc |
| ASIC4-Ex1-F | agaatgagctgaggaccctg |
| ASIC4-Ex1-R | ttgcacacaatctcgatcgg |
| ASIC4-Ex7-F | tgctgggtgagactggtgt |
| ASIC4-Ex7-R | ccctgaacctgactttccag |
| CHPF-Ex1-F | tgctggaggggaatcgag |
| CHPF-Ex1-R | cttccttccggagcctgac |
| CHPF-Ex2-F | ctaggacccgctacatcagc |
| CHPF-Ex2-R | ccaacatccctttgcctctc |
| CHPF-Ex4-F | gctaccgacgctttgatcc |
| CHPF-Ex4-R | acgtgggccttgacaggt |
| CHPF-3'UTRa-F | gctgttcctccacttctcca |
| CHPF-3'UTRa-R | cctccttccccaacaactct |
| TMEM198-Ex2-F | agcgggtgctagagacacag |
| TMEM198-Ex2-R | agtcacttggcccccaac |
| OBSL1-5'UTR-F | cagtctgggctcttgtcctc |
| OBSL1-5'UTR-R | aggaccacgcacttgagc |
| OBSL1-Ex1a-F | ctgtgcgggtggtaagtg |
| OBSL1-Ex1a-R | acacgtagacgccggaatc |
| OBSL1-Ex1b-F | acgaagtgtgggacagcag |
| OBSL1-Ex1b-R | ccacctctcctccagtcttg |
| OBSL1-Ex2-F | gcatgaagagcacacagca |
| OBSL1-Ex2-R | ctgagatgcggacaggaatc |
| OBSL1-Ex12-F | tctcagtaccttccccgtca |
| OBSL1-Ex12-R | ggagccagaagcagcaaag |

| | |
|---|---|
| WNT 6-Ex1-F | ccggctctgatttcttctcc |
| WNT 6-Ex1-R | gaacaccccagtctgtcctg |
| WNT6-Ex3-F | ttgcctgagccccacttc |
| WNT6-Ex3-R | actcctgccaacacttcctc |
| WNT6-Ex4a-F | cacctcccattcccaatc |
| WNT6-Ex4a-R | aggcagttctcttcgagctg |
| WNT6-Ex4b-F | gattcgcccgacttctgc |
| WNT6-Ex4b-R | gggagcccagtatccagag |
| WNT6-3'UTR-F | aggggcttgagaggaacg |
| WNT6-3'UTR-R | gaatgccaaggggagatagc |
| WNT10A-Ex1-F | gagtcggagctgtgtgtcg |
| WNT10A-Ex1-R | tcttccagggtcctctaccc |
| WNT10A-Ex4a-F | ggagtgggtttcagaagcag |
| WNT10A-Ex4a-R | gtctggcgcaggatgttg |
| WNT10A-Ex4b-F | tctcccgacttctgcgag |
| WNT10A-Ex4b-R | tccattcattccccacctcc |
| CDK5R2-Ex1a-F | agctgcccatcaggagtg |
| CDK5R2-Ex1a-R | gaaggttctcgcggttgc |
| CDK5R2-Ex1b-F | gtccgccaagaagaagaaag |
| CDK5R2-Ex1b-R | gtttgcaggcgtaatgaagg |
| CDK5R2-Ex1c-F | ggcaacgagatctcctaccc |
| CDK5R2-Ex1c-R | ggaatgggctgaagggaagg |
| FEV-Ex1-F | gatgggacgataagaggggc |
| FEV-Ex1-R | cactctcttcccatgcctga |
| FEV-Ex2-F | cggggccctttgtcaag |
| FEV-Ex2-R | cacactgctcccacctact |
| FEV-Ex3a-F | tttcccgccagcactctt |
| FEV-Ex3a-R | aagctgggactggggtaga |
| FEV-Ex3b-F | gcctctccaaactcaacctc |
| FEV-Ex3b-R | tgaatggggcttctaggagc |
| CRYBA2-Ex1-F | gctgtgtgtggctcgaaac |
| CRYBA2-Ex1-R | atgttggcaggtctctccag |
| CRYBA2-Ex2-F | gtgtcaggggaagggtttg |
| CRYBA2-Ex2-R | atgcagggcttgaacagc |
| CCDC108-Ex4-F | cctcgtttcctgcctcttc |
| CCDC108-Ex4-R | ctcctggttgctctgtcctg |

| | |
|---|---|
| OBSL1-Ex17-F | gggaacacgagacacgcata |
| OBSL1-Ex17-R | cggagagtaccgccacagta |
| OBSL1-Ex18-F | atgtggaggctgggcact |
| OBSL1-Ex18-R | gggaaagaacagggacgag |
| OBSL1-Ex19-F | agcctcgtccctgttctttc |
| OBSL1-Ex19-R | ttagccctctatgccaccag |
| OBSL1-Ex20-F | tggtggcatagagggctaag |
| OBSL1-Ex20-R | cagttccagggctttccag |
| INHA-Ex1-F | gactggggaagactggatga |
| INHA-Ex1-R | cctgcaaacccatgctgt |
| STK11IP-Ex1,2-F | gttttccggtcgtcccttg |
| STK11IP-Ex1,2-R | atgactctcagggcgtgtct |
| STK11IP-Ex14-F | gcttgctcagttctgggttc |
| STK11IP-Ex14-R | ccactcacaaacatcccaag |
| SGPP2-Ex1-F | gcaaggtggaggcagaca |
| SGPP2-Ex1-R | gcatcctggttactggaagg |
| FARSB-Ex1-F | cgggacttcagggtcagta |
| FARSB-Ex1-R | cggagccaaaaccttcag |
| MOGAT1-Ex1-F | agcctctgccttttcctctc |
| MOGAT1-Ex1-R | ctctttgacctgcctgctct |
| AP1S3-Ex1-F | agggaggagaaaggggaaag |
| AP1S3-Ex1-R | ccggcacagactaagcactc |
| WDFY1-Ex1-F | tctcccagccacagactttc |
| WDFY1-Ex1-R | agctaaggggcgcagagt |
| MRPL44-Ex1-F | cgcaagcgtagcctcaag |
| MRPL44-Ex1-R | atcagccgcagacacgac |
| SERPINE2-Trns4-Ex1-F | ccctgaccctgaacctga |
| SERPINE2-Trns4-Ex1-R | cctgaaggtggagtgctgtc |
| FAM124B-Trns2-Ex2-F | ttgggctcaggaatgtcac |
| FAM124B-Trns2-Ex2-R | agctcggccaacatagtgag |
| CUL3-Trns1-Ex1-F | tcactctccggctctcct |
| CUL3-Trns1-Ex1-R | ctgttgggggacttcagc |
| CUL3-Trns3-Ex1-F | actgccattcctcagatgct |
| CUL3-Trns3-Ex1-R | tccttacctccctccaatcc |
| DOCK10-Ex1-F | cgggtggatagagaaggttg |
| DOCK10-Ex1-R | cattaaagcccctgcacatc |

| ***SCT135 non-coding region gaps coverage*** | |
|---|---|
| **Name** | **Sequence (5'-3')** |
| BZW1-Trns1-5'UTR-I-F | gattccttccccaaaccacg |
| BZW1-Trns1-5'UTR-I-R | ctttcaggccctagatccgg |
| BZW1-Trns1-3'UTRa-F | cctacctccctgtatcaagca |
| BZW1-Trns1-3'UTRa-R | tctgttctccatggcttctc |
| BZW1-Trns1-3'UTRb-F | ttgagtacgtggaccattgc |
| BZW1-Trns1-3'UTRb-R | aagagaacgaagccagtcca |
| BZW1-Trns1-3'UTRc-F | ttcctactcggtcatactgga |
| BZW1-Trns1-3'UTRc-R | caccccattcacaacaatcagt |
| BZW1-Trns1-3'UTRd-F | gaaatcgccccttttgaag |
| BZW1-Trns1-3'UTRd-R | ggtggtttacagtttgggagtc |
| CLK1-5'UTR-F | agtgagtggggacaacacg |
| CLK1-5'UTR-R | agccgccatcttacagctc |
| CLK1-3'UTR-F | ttggacagctctctcgaaga |

| ***SCT135 non-coding region gaps coverage*** | |
|---|---|
| **Name** | **Sequence (5'-3')** |
| C2ORF67-5'UTR-I-F | tgccacttccagaacactcc |
| C2ORF67-5'UTR-I-R | aagtccacccggcagtct |
| LANCL1-5'UTR-I-F | tctccgcgtgattctaggtc |
| LANCL1-5'UTR-I-R | actgcggcccaactgcta |
| CPS1-5'UTR-I-Trns3-F | ttccctttgctctctctcg |
| CPS1-5'UTR-I-Trns3-R | tgagtttcatcatgggttgg |
| IKZF2-5'UTR-I-Trns 2-F | acgcgctctcatacacacac |
| IKZF2-5'UTR-I-Trns 2-R | cacagtgcaaaacgagatgc |
| SPAG16-3'UTR-Trns2-F | tgtgagaggcttcattagcaa |
| SPAG16-3'UTR-Trns2-R | gttggtgagaacgtgcagaa |
| VWC2L-5'UTR-I-F | tgggaggtagagagaatgga |
| VWC2L-5'UTR-I-R | ccctccccatattgctacag |
| FN1-5'UTR-F | ggccatcagcatctcttttg |

| | | | |
|---|---|---|---|
| CLK1-3'UTR-R | accgcaataccaagtttgagt | FN1-5'UTR-R | agcctgcctcttgctcttc |
| PPIL3-3'UTR-F | ttgcatcatccttctgcttg | TMEM169-5'UTR-I-II-F | tctgggtctcagggacattc |
| PPIL3-3'UTR-R | ttcagatttggggatacttgc | TMEM169-5'UTR-I-II-R | ccgaaggatgaaaggatctg |
| NIF3L1-5'UTR-I-F | aactacctgggcaacctcaa | MARCH4-5'UTRa-F | agacaacacctggcagcag |
| NIF3L1-5'UTR-I-R | cccgtcctagagtaagcactg | MARCH4-5'UTRa-R | catgatcctcctcctcctc |
| ORC2-5'UTR-I-F | ggcccaagaatccgtttt | MARCH4-5'UTRb-F | agcactcttgcctgagaacc |
| ORC2-5'UTR-I-R | cgagggaaagtcctgagatg | MARCH4-5'UTRb-R | gggcagtggagcagagaata |
| ORC2-5'UTR-II-F | tctcttcccactccatccag | MARCH4-5'UTRc-F | agggacctttgatggaacc |
| ORC2-5'UTR-II-R | gccaccaccacagtcattta | MARCH4-5'UTRc-R | gtggattgagctgggcttag |
| FAM126B-5'UTR-I-F | tccttagacaccgcacacag | MARCH4-5'UTRd-F | caacccctaccacctgagaa |
| FAM126B-5'UTR-I-R | ggccgaaggagaagaaaac | MARCH4-5'UTRd-R | atgtgccccgtggaagtag |
| FAM126B-5'UTR-II-F | atgcaagtgaggctccagat | SMARCAL1-5'UTR-I-F | gttgcagctcccgtttctc |
| FAM126B-5'UTR-II-R | gatgtagccggataatgcaa | SMARCAL1-5'UTR-I-R | taagacccactgatcagccc |
| NDUFB3-5'UTR-I-F | caactgcagcgcaacaac | SMARCAL1-5'UTR-II-F | ccgtgtgttgtgttggaagt |
| NDUFB3-5'UTR-I-R | cctcccatccttctccattt | SMARCAL1-5'UTR-II-R | gataactcacatggcccagc |
| NDUFB3-5'UTR-II-F | ggaagaggtggttggaggat | IGFBP2-3'UTR-F | aagatgtctctgaacgggca |
| NDUFB3-5'UTR-II-R | gtgctggtatccttttccctaag | IGFBP2-3'UTR-R | aggggcaaagctggatgat |
| CFLAR-5'UTR-I-Trns1-F | gaactaggaaggcggaggtt | IGFBP5-5'UTRa-F | tgggaagctcaaattgcagc |
| CFLAR-5'UTR-I-Trns1-R | cctgctgcaatgctgactc | IGFBP5-5'UTRa-R | gagatcggatgcagtggaga |
| CFLAR-5'UTR-I-Trns2-F | ctgaagcggaaattgccta | IGFBP5-5'UTRb-F | tggcagctgagggttagaaa |
| CFLAR-5'UTR-I-Trns2-R | aatccttttctgccagtatcg | IGFBP5-5'UTRb-R | gcacatggagagggctttct |
| CASP10-5"UTR-I-Trns1-F | ggttaacatggccgacaact | TNS1-5'UTR-I-F | gagcactgcaggggttaaag |
| CASP10-5"UTR-I-Trns1-R | gaagccatgaagcaaaggaa | TNS1-5'UTR-I-R | tcagtatcaacagcccagca |
| CASP10-3'UTR-Trns2-F | agagaacagtgtggggtgct | TNS1-5'UTR-II-F | tgagggaatgaagcagggg |
| CASP10-3'UTR-Trns2-R | caaggtgcaaaacagtctgc | TNS1-5'UTR-II-R | cgacaccccagttaccaaca |
| CASP8-5'UTR-I-TrnsE-F | cacagtgccaggaagtgaga | TNS1-5'UTR-III-F | gcctcatcatccctgtctca |
| CASP8-5'UTR-I-TrnsE-R | cgtggaaaagaactgcttgg | TNS1-5'UTR-III-R | gaagcatttccccaaccaca |
| CASP8-5'UTR-II-TrnsE-F | tgcgggtcaagacatcagta | TNS1-5'UTR-V-F | ttttgcaggggttagggtgg |
| CASP8-5'UTR-II-TrnsE-R | aactggcatctgctccactc | TNS1-5'UTR-V-R | ctttcacgtttgcccacact |
| CASP8-5'UTR-I-TrnsC-F | gaaggcactctgctcactca | RUFY4-5'UTR-I-F | agagacgagacaaccacagg |
| CASP8-5'UTR-I-TrnsC-R | ccacccgatctcactctcag | RUFY4-5'UTR-I-R | atgtgtctcctcctctccct |
| CASP8-3'UTR-II-TrnsE-F | gggagggatagagaggggat | RUFY4-5'UTR-II-F | aaggagtgtgatgccaggag |
| CASP8-3'UTR-II-TrnsE-R | ctctgcgagaaacacccaag | RUFY4-5'UTR-II-R | taaatgagggaggtgggtgg |
| ALS2CR12-5'UTR-I-F | ggctgtttctcccaaccttc | RUFY4-3'UTR-F | tgcccatcccactcaatc |
| ALS2CR12-5'UTR-I-R | gccctaacaatagtcaccagtg | RUFY4-3'UTR-R | aagggagggagaccacaaag |
| TRAK2-5'UTR-I-F | ggcttggaactctgcacca | CXCR2-Trns2-5'UTR-I-F | ggaggggagtgatgcaaac |
| TRAK2-5'UTR-I-R | tagcgcccattactttgcac | CXCR2-Trns2-5'UTR-I-R | ggaactgaacaggaaagctg |
| STRADB-5'UTR-I-F | gaagccaaaggagagccact | CXCR2-Trns1-5'UTR-I-F | ggacgcttctcactctgacc |
| STRADB-5'UTR-I-R | caaggctcactccctcacac | CXCR2-Trns1-5'UTR-I-R | tcatcaccccagttctcagg |
| STRADB-3'UTR-F | caccagtttgtagtccctctg | CXCR2-Trns1-5'UTR-II-F | agctccacccctgaattg |
| STRADB-3'UTR-R | gggcttttcttggcaaaac | CXCR2-Trns1-5'UTR-II-R | gtggggagagatgcttagga |
| TMEM237-5'UTR-Trns2-F | atgaagtttgcggggacct | CXCR1-5'UTR-I-F | tcagataatggaccctcaacc |
| TMEM237-5'UTR-Trns2-R | acccacccagagacccaac | CXCR1-5'UTR-I-R | gtcagcctcttcccacaaaa |
| MPP4-3'UTR-F | cccttgatacagcgatccat | ARPC2-3'UTR-F | ccctcacttccctttctcc |
| MPP4-3'UTR-R | tttccttgtaggccgatttg | ARPC2-3'UTR-R | ggtttgcaggttcaagggta |
| ALS2-5'UTR-I-F | ctgctcccaaccactttcac | AAMP-3'UTR-F | ctttccagcctcttcaactg |
| ALS2-5'UTR-I-R | cagagctgggaaaatgcgg | AAMP-3'UTR-R | ctcaacatccagccgattct |
| FZD7-5'UTR-F | aacccaggctgacgagtttt | TMBIM1-5'UTR-I-F | tcttccctgcatttctgtgtc |
| FZD7-5'UTR-R | atgcccttctctccgtggta | TMBIM1-5'UTR-I-R | cgcgctccgaaaagttta |
| BMPR2-5'UTRa-F | tgcaccacaggctagagaca | VIL1-5'UTR-I-F | tctcattgtcaggctgtgga |
| BMPR2-5'UTRa-R | gaaggctgaggagaggtagga | VIL1-5'UTR-I-R | cccagcacttctcacgctat |

| | |
|---|---|
| BMPR2-5'UTRb-F | gagccgcaggaataaaaagc |
| BMPR2-5'UTRb-R | catgtctgtggggaaagcag |
| BMPR2-5'UTRc-F | ctccgcaccctggatatgt |
| BMPR2-5'UTRc-R | gtcttctctcccgtggactg |
| ICA1L-5'UTR-I-Trns1-F | acacctccacgcactggt |
| ICA1L-5'UTR-I-Trns1-R | aacccgggatttggtctc |
| ICA1L-5'UTR-I-Trns2-F | aggcccggagaccaaatc |
| ICA1L-5'UTR-I-Trns2-R | gcctaaggagcaggatctgt |
| ICA1L-5'UTR-II-Trns1-F | agtccatcctggcagctcac |
| ICA1L-5'UTR-II-Trns1-R | cactgccttccacatagctg |
| ICA1L-5'UTR-III-Trns2-F | tgccattattgactcacttgca |
| ICA1L-5'UTR-III-Trns2-R | aaacacaagtgcaggctatatg |
| ICA1L-3'UTR-Trns1-F | tgttatgcttttggggtgct |
| ICA1L-3'UTR-Trns1-R | aatcccagctacttggtaggc |
| WDR12-3'UTR-F | ctgagtggaggagcagacaa |
| WDR12-3'UTR-R | ttgggaggttgagacaggag |
| CARF-5'UTR-I-Trns001-F | gcctctaactgcggctttt |
| CARF-5'UTR-I-Trns001-R | agcaaccctcctgaagacac |
| CARF-5'UTR-I-Trns02-F | cttttcggggtcccacat |
| CARF-5'UTR-I-Trns02-R | gcctaatgtttaaatgaacgcc |
| CARF-5'UTR-II-Trns02-F | gttttaccacaccaggctga |
| CARF-5'UTR-II-Trns02-R | cctggccctgtcaaagtaaa |
| CARF-5'UTR-III-Trns02-F | ctttgtcagctacatgttttgc |
| CARF-5'UTR-III-Trns02-R | ctggacagattcgagatggag |
| NBEAL1-5'UTR-I-F | caacggctgggtaaatcct |
| NBEAL1-5'UTR-I-R | ctcccatttcttgccatgc |
| NBEAL1-3'UTR-F | gcaaggcttttatccctgaa |
| NBEAL1-3'UTR-R | ggcccaactgtcaaaatcac |
| RAPH-5'UTR-F | cgccggaagtgagtggtt |
| RAPH-5'UTR-R | gatccccaacctggcaac |
| RAPH-3'UTR-Trns3-F | gaagccctcctaaacccagt |
| RAPH-3'UTR-Trns3-R | agggccgattcatagactcc |
| NRP2-5'UTR-F | gcggctctgacatccacat |
| NRP2-5'UTR-R | tcagcaaagagggaacaagc |
| NRP2-3'UTR-Trns5-F | caatcccaaccatcctcct |
| NRP2-3'UTR-Trns5-R | tcaaagacaccagaagacaagg |
| INO80D-5'UTR-I-F | gctctgcctccaagatggta |
| INO80D-5'UTR-I-R | cccactcaccctttcattct |
| EEF1B2-5'UTR-F | aagcccggagggactagaat |
| EEF1B2-5'UTR-R | ttttgcattcccacttctcc |
| GPR1-5'UTR-I-Trns1-F | cacagcagccctttctctta |
| GPR1-5'UTR-I-Trns1-R | gaaatgtggggaacagcttc |
| GPR1-5'UTR-II-Trns3-F | tcagagaggtgagtgcttgc |
| GPR1-5'UTR-II-Trns3-R | ttcttagggctggctcttca |
| GPR1-5'UTR-II-Trns1-F | agcaaaagcctacccagaca |
| GPR1-5'UTR-II-Trns1-R | gcattttgtaaacgatcaggag |
| ZDBF2-5'UTR-I-F | gacgtcaaacctccatcagg |
| ZDBF2-5'UTR-I-R | gacacctctcacgcagcag |
| ZDBF2-5'UTR-II-F | ttggaaacatgtcaccgttc |
| ZDBF2-5'UTR-II-R | ggatcatttctttagggtctgg |
| FASTKD2-5'UTR-I-F | ccatggtcgagagagactcag |
| FASTKD2-5'UTR-I-R | atacaaccagccagcctcac |
| USP37-5'UTR-I-F | aaagtgcgcgtgcctactac |
| USP37-5'UTR-I-R | ggtagggaaaacggctcaat |
| USP37-5'UTR-II-F | cctccacctctcaaagtgct |
| USP37-5'UTR-II-R | tcaccatgaatgctcaaagc |
| USP37-5'UTR-III-F | tggctcaagtttctgtagctct |
| USP37-5'UTR-III-R | agcattgccagaccatctaa |
| PLCD4-5'UTR-I-F | gcactgcaactgaggacac |
| PLCD4-5'UTR-I-R | ctagagggcagaggtaggga |
| PLCD4-3'UTR-F | aacaagcaatctgggacctg |
| PLCD4-3'UTR-R | ggtctcctgggaacctcttt |
| ZNF142-5'UTR-I-F | agaagcccacgccactctac |
| ZNF142-5'UTR-I-R | tgctcccttcagtacctcct |
| ZNF142-5'UTR-II-F | acaaaccgcctttccagttc |
| ZNF142-5'UTR-II-R | agcgaaactccgtctcaaaa |
| ZNF142-3'UTR-F | actggacctgagggctga |
| ZNF142-3'UTR-R | gacaaaagggaggcagaggt |
| BCS1L-5'UTR-I-F | gtttttcccgccctccac |
| BCS1L-5'UTR-I-R | ctctcctccaaccccgaac |
| BCS1L-5'UTR-II-F | cagctcacctccctgcat |
| BCS1L-5'UTR-II-R | ggtcttcacctcctgcttca |
| STK36-5'UTR-I-F | cccccaaagttaccagtcct |
| STK36-5'UTR-I-R | tttagccccatctcttctgg |
| TTLL4-5'UTR-I-F | accgcactgaagggactg |
| TTLL4-5'UTR-I-R | cgcgaaaggactcaggtaga |
| TTLL4-5'UTR-II-F | agccacccctgtagcttgt |
| TTLL4-5'UTR-II-R | ggcctcccaaacaaaatagc |
| CYP27A1-5'UTR-F | atttctttgttcagcggccc |
| CYP27A1-5'UTR-R | ctttgggtcgagtgctgagt |
| CYP27A1-3'UTR-F | agacgggggagttgaagagt |
| CYP27A1-3'UTR-R | gaaaccaaggcctctctgtg |
| WNT10A-5'UTR-F | gtggggtagagccctgaga |
| WNT10A-5'UTR-R | agccagcagcagtaggaaga |
| WNT10A-3'UTR-F | cgaagcccagacagttcagt |
| WNT10A-3'UTR-R | ctggaatgcccttacccata |
| FEV-5'UTR-F | agggagggaagcggaaattt |
| FEV-5'UTR-R | gggtcctggagaaacgctt |
| FEV-3'UTR-F | gtgcgccctctttcacttt |
| FEV-3'UTR-R | ttccttggcagttctctgct |
| CRYBA2-5'UTR-F | ggctcaggctggaaagaaag |
| CRYBA2-5'UTR-R | aagtcctcctcgtcccagag |
| NHEJ1-5'UTR-F | acgcgccactacccttctac |
| NHEJ1-5'UTR-R | ttcactcgagccctaccatc |
| ABCB6-5'UTR-F | tcgttgcctgcaccattc |
| ABCB6-5'UTR-R | agaagaagcagggactcagg |
| ANKZF1-5'UTR-I-F | ggagacacaacggccaat |
| ANKZF1-5'UTR-I-R | tctcaaaccccagtctcag |
| GLB1L-5'UTR-F | gcgtagtttcagaggacgtg |
| GLB1L-5'UTR-R | ggttgtacgggatcaggaga |
| GLB1L-3'UTR-F | gtacctctccagccccaagt |
| GLB1L-3'UTR-R | gcactgtggcctttagttcc |
| DNAJB2-5'UTR-I-F | gcttctcgcttggatgga |
| DNAJB2-5'UTR-I-R | agttggaagcggagcacac |

| | |
|---|---|
| KLF7-5'UTR-I-Trns2-F | ttaacgcacagtcgagcaac |
| KLF7-5'UTR-I-Trns2-R | agcgggaaggaaggacagt |
| KLF7-5'UTR-Trns1-F | agcaggaggaggagaaaagg |
| KLF7-5'UTR-Trns1-R | ggggctgtttgtttgtcagt |
| CREB1-5'UTR-I-F | aaggtcttcggcaagttcc |
| CREB1-5'UTR-I-R | cgccgccattattctttg |
| METTL21A-5'UTR-I-Trns2-F | cgcaacgtcactcctcca |
| METTL21A-5'UTR-I-Trns2-R | ctgggtgttagcggtttgat |
| METTL21A-5'UTR-I-Trns1-F | gatcctaagaggccctgtgac |
| METTL21A-5'UTR-I-Trns1-R | actacacaagcggcggaag |
| FZD5-5'UTR-I-F | gctccgtctggatcatgc |
| FZD5-5'UTR-I-R | caccctctctttccaagctg |
| PLEKHM3-5'UTR-I-F | cctcgggcaggaaagact |
| PLEKHM3-5'UTR-I-R | atctcccctcacccattagc |
| CRYGD-5'UTR-F | cgtggtctagcacagcaaag |
| CRYGD-5'UTR-R | gacaaggcgaggtctcacag |
| C2ORF80-3'UTR-F | ccacagctcttgccagtaca |
| C2ORF80-3'UTR-R | aagccatgtgcctaatcagc |
| IDH1-5'UTR-I-F | aattggcgtgtggcgatt |
| IDH1-5'UTR-I-R | ttaccccattcctgccttct |
| IDH1-5'UTR-II-F | gggctgtctggcaggtacta |
| IDH1-5'UTR-II-R | ctgcccccaagaccagtta |
| PIKFYVE-5'UTR-I-Trns2-F | ctggactccttctgcctgag |
| PIKFYVE-5'UTR-I-Trns2-R | gccctggttctcactttcac |
| PIKFYVE-3'UTR-Trns4-F | cacctcctccccatacctttt |
| PIKFYVE-3'UTR-Trns4-R | ggcatttcttatgtgctgct |
| PTH2R-5'UTR-F | agcagcgacatgagatcctt |
| PTH2R-5'UTR-R | ctggctcttacctgggctct |
| MAP2-5'UTR-I-Trns5-F | cactggcctcactttgtcg |
| MAP2-5'UTR-I-Trns5-R | ctcagagaccgaggcaaaag |
| MAP2-5'UTR-II-Trns5-F | cagtgcattttatggagtgagg |
| MAP2-5'UTR-II-Trns5-R | gcaagcatatttctgacacagc |
| MAP2-5'UTR-I-Trns1-F | ccactcgccttatttttcctg |
| MAP2-5'UTR-I-Trns1-R | atgcctggaggtctgtagca |
| MAP2-5'UTR-II-Trns1-F | tgctacagacctccaggcata |
| MAP2-5'UTR-II-Trns1-R | tgcaacataggcaaaagaca |
| MAP2-5'UTR-III-Trns1-F | gccagaagagaaagccacaa |
| MAP2-5'UTR-III-Trns1-R | attgtcaaccaaggcaggag |

| | |
|---|---|
| PTPRN-5'UTR-I-Trns3-F | catgcgagatgaggaaggtg |
| PTPRN-5'UTR-I-Trns3-R | gcttccactgcccagagat |
| PTPRN-3'UTR-F | ctgtccctctttgcctgtgt |
| PTPRN-3'UTR-R | gctctcctgctgtgctctct |
| GMPPA-5'UTR-I-F | tcagcacctctttcctctcc |
| GMPPA-5'UTR-I-R | tcctgaacgtaatccccaag |
| GMPPA-3'UTR-F | atgagaccaggggcacagat |
| GMPPA-3'UTR-R | ccagggctccagattcatta |
| CHPF-5'UTR-I-Trns2-F | ccaaaaaggccgtcaggtag |
| CHPF-5'UTR-I-Trns2-R | ctgaggtggcccttgtaacc |
| TMEM198-5'UTR-I-F | taccggagagggaggagaag |
| TMEM198-5'UTR-I-R | tagtggaggctgggatgaag |
| OBSL1-3'UTR-Trns3-F | ggagttatgcgtcctcttgaa |
| OBSL1-3'UTR-Trns3-R | tgtgtggaagggtgactttct |
| STK11Ip-3'UTR-F | ccaaggtggagacagtgagg |
| STK11Ip-3'UTR-R | aaggtgatgggtgggatagc |
| CCDC140-5'UTR-I-F | tcttcaagtggacggaaagg |
| CCDC140-5'UTR-I-R | caggctcctttggtgcttc |
| CCDC140-5'UTR-II-F | cgccggtacataaagagctg |
| CCDC140-5'UTR-II-R | agcctcttgcctctggttc |
| FARSB-3'UTR-F | gaccatcctttccctgcata |
| FARSB-3'UTR-R | gcaccctcacatattttgaatc |
| ACSL3-5'UTR-I-F | ggcccgaattcgataggc |
| ACSL3-5'UTR-I-R | catcctccctgcctggtc |
| ACSL3-5'UTR-II-F | ccaggaaaaagttggcattc |
| ACSL3-5'UTR-II-R | ctgagttcaaatcccttgcag |
| ACSL3-5'UTR-III-F | ttctaccaaccatgcaacca |
| ACSL3-5'UTR-III-R | ttgaactgttgtgccctttg |
| SCG2-5'UTR-I-F | ttgccacaaaaagacaggtg |
| SCG2-5'UTR-I-R | gcaggttgctattcagttatcc |
| SCG2-3'UTR-F | aggcagatgaaaccaggtca |
| SCG2-3'UTR-R | cctgaattgtctggtgaagatg |
| SERPINE2-5'UTR-I-Trns1-F | agaagtcgcaggcggaga |
| SERPINE2-5'UTR-I-Trns1-R | gcgggaggctgagaaagt |
| FAM124B-5'UTR-Trns1-F | ctataagagccgcccatttg |
| FAM124B-5'UTR-Trns1-R | atgcaatccaggagctggt |
| DOCK10-5'UTR-F | tctcccagcaacctccaaat |
| DOCK10-5'UTR-R | gccgagcttttggggaag |

| ***DES gene sequencing primers*** | |
|---|---|
| **Name** | **Sequence (5'-3')** |
| DES-Ex2,3-F | ccctgtctctcctgcctcta |
| DES-Ex2,3-R | agcaccctcacaccctatt |
| DES-Ex4-F | tataccttggcccctcttcc |
| DES-Ex4-R | ggctctcctcccatggat |
| DES-Ex6-F | acctgaccatctggagttgc |
| DES-Ex6-R | gtttccaggtcttgttgagg |
| DES-Ex7-F | ggatggtctcgatctcctga |
| DES-Ex7-R | ttcttccccagactccaggt |

| ***DES gene sequencing primers*** | |
|---|---|
| **Name** | **Sequence (5'-3')** |
| DES-Ex6-F-n | ccagccccaaagctttctt |
| DES-Ex6-R-n | tttccaggtcttgttgaggt |
| DES-Ex7-F-n | gagacggggtttcactgtgt |
| DES-Ex7-R-n | aacagaggctgcttcctgag |
| DES-Ex8-F-n | ctagggctctgcccaatgt |
| DES-Ex8-R-n | atgggctatgtcgctgttg |
| DES-Ex9a-F-n | cttccatcccaggacacc |
| DES-Ex9a-R-n | cagtctccacaggggctagt |

| | | | |
|---|---|---|---|
| DES-Ex8-F | atgtggccccagatggac | DES-Ex9b-F-n | agcagggtgttgggatactg |
| DES-Ex8-R | ggtgtcctgggatggaaga | DES-Ex9b-R-n | ccatctcagggtgacaggtt |
| DES-Ex9a-F | ggctagcacatggttggact | DES-Ex7-F-nn | ccgatgggagggttcttaac |
| DES-Ex9a-R | gtaagagaccacggggacaa | DES-Ex8-F-nn | actcccagcccctggtatag |
| DES-Ex9b-F | atactgcagggccaggact | DES-Ex8-R-nn | acagctgggtggaagacagt |
| DES-Ex9b-R | ctggctgggtagaaggtcag | DES-Ex9-F-nn | tagtggggcaagagagatcc |

## A 3.3 Primer pairs used for cloning and site-directed mutagenesis reactions (SDM)

| *DES cDNA primers* | |
|---|---|
| **Name** | **Sequence (5'-3')** |
| DES-CCDS-1-F | ctgagttcgcccgtgttc |
| DES-CCDS-1-R | ctcggaaggcagccaaat |
| DES-CCDS-2-F | ggtggaggtgctcactaacc |

| *DES cDNA primers* | |
|---|---|
| **Name** | **Sequence (5'-3')** |
| DES-CCDS-2-R | taggactggatctggtgtcg |
| DES-CCDS-3-F | catcgcggctaagaacattt |
| DES-CCDS-3-R | gatcatcaccgtcttcttgg |

| *DES cloning primers* | |
|---|---|
| **Name** | **Sequence (5'-3')** |
| DES-pcDNA3.1(+) NHEI-F | gctggctagcgccaccatgagccaggcctactcg |
| DES-pcDNA3.1(+) XHOI-R | tctagactcgagttagagcacttcatgctgctg |

| *SCT135 minigene primers* | |
|---|---|
| **Name** | **Sequence (5'-3')** |
| USP37-Ex18-MG-EcoRI-FP | gtgtgaattcacactatgacctatggctg |
| USP37-Ex18-MG-BamHI-RP | tcagtcagggatccgaaacactacccacgtacc |

| *DES SDM primers* | |
|---|---|
| **Name** | **Sequence (5'-3')** |
| DES-Gly65Ser-F | cggggccgggagcctggggtc |
| DES-Gly65Ser-R | gaccccaggctcccggccccg |
| DES-Glu322Asp-F | agcaggagatgatggattaccgacaccagatcc |
| DES-Glu322Asp-R | ggatctggtgtcggtaatccatcatctcctgct |

| *DES SDM primers* | |
|---|---|
| **Name** | **Sequence (5'-3')** |
| DES-Ala337Thr-F | cctgcgagattgacaccctgaagggcact |
| DES-Ala337Thr-R | agtgcccttcagggtgtcaatctcgcagg |
| DES-Leu345Pro-F | gcactaacgattccccgatgaggcagatgcg |
| DES-Leu345Pro-R | cgcatctgcctcatcggggaatcgttagtgc |

## A 3.4 DES protein sequence (NCBI, FASTA format)

desmin [Homo sapiens]

NCBI Reference Sequence: NP_001918.3

>gi|55749932|ref|NP_001918.3| desmin [Homo sapiens]

MSQAYSSSQRVSSYRRTFGGAPGFPLGSPLSSPVFPRAGFGSKGSSSSVTSRVYQVSRTSGGAGGLGSLR
ASRLGTTRTPSSYGAGELLDFSLADAVNQEFLTTRTNEKVELQELNDRFANYIEKVRFLEQQNAALAAEV
NRLKGREPTRVAELYEEELRELRRQVEVLTNQRARVDVERDNLLDDLQRLKAKLQEEIQLKEEAENNLAA
FRADVDAATLARIDLERRIESLNEEIAFLKKVHEEEIRELQAQLQEQQVQVEMDMSKPDLTAALRDIRAQ

YETIAAKNISEAEEWYKSKVSDLTQAANKNNDALRQAKQEMMEYRHQIQSYTCEIDALKGTNDSLMRQMR
ELEDRFASEASGYQDNIARLEEEIRHLKDEMARHLREYQDLLNVKMALDVEIATYRKLLEGEESRINLPI
QTYSALNFRETSPEQRGSEVHTKKTVMIKTIETRDGEVVSEATQQQHEVL

desmin [Mus musculus]

NCBI Reference Sequence: NP_034173.1

>gi|33563250|ref|NP_034173.1| desmin [Mus musculus]
MSQAYSSSQRVSSYRRTFGGAPGFSLGSPLSSPVFPRAGFGTKGSSSSMTSRVYQVSRTSGGAGGLGSLR
SSRLGTTRAPSYGAGELLDFSLADAVNQEFLATRTNEKVELQELNDRFANYIEKVRFLEQQNAALAAEVN
RLKGREPTRVAELYEEEMRELRRQVEVLTNQRARVDVERDNLIDDLQRLKAKLQEEIQLREEAENNLAAF
RADVDAATLARIDLERRIESLNEEIAFLKKVHEEEIRELQAQLQEQQVQVEMDMSKPDLTAALRDIRAQY
ETIAAKNISEAEEWYKSKVSDLTQAANKNNDALRQAKQEMMEYRHQIQSYTCEIDALKGTNDSLMRQMRE
LEDRFASEANGYQDNIARLEEEIRHLKDEMARHLREYQDLLNVKMALDVEIATYRKLLEGEESRINLPIQ
TFSALNFRETSPEQRGSEVHTKKTVMIKTIETRDGEVVSEATQQQHEVL

desmin [Rattus norvegicus]

NCBI Reference Sequence: NP_071976.1

>gi|11968118|ref|NP_071976.1| desmin [Rattus norvegicus]
MSQAYSSSQRVSSYRRTFGGAPGFSLGSPLSSPVFPRAGFGTKGSSSSVTSRVYQVSRTSGGAGGLGSLR
ASRLGTTRAPSYGAGELLDFSLADAVNQEFLATRTNEKVELQELNDRFANYFEKVRFLEQQNAALAAEVN
RLKGREPTRVAELYEEEMRELRRQVEVLTNQRARVDVERDNLIDDLQRLKAKLQEEIQLREEAENNLAAF
RADVDAATLARIDLERRIESLNEEIAFLKKVHEEEIRELQAQLQEQQVQVEMDMSKPDLTAALRDIRAQY
ETIAAKNISEAEEWYKSKVSDLTQAANKNNDALRQAKQEMMEYRHQIQSYTCEIDALKGTNDSLMRQMRE
LEDRFASEASGYQDNIARLEEEIRHLKDEMARHLREYQDLLNVKMALDVEIATYRKLLEGEESRINLPIQ
TFSALNFRETSPEQRGSEVHTKKTVMIKTIETRDGEVVSEATQQQHEVL

desmin [Bos taurus] - cow

NCBI Reference Sequence: NP_001075044.1

>gi|126158909|ref|NP_001075044.1| desmin [Bos taurus]
MSQAYSSSQRVSSYRRTFGGAPSFPLGSPLSSPVFPRAGFGTKGSSSSVTSRVYQVSRTSGGAGGLGALR
ASRLGSTRVPSSYGAGELLDFSLADAVNQEFLTTRTNEKVELQELNDRFANYIEKVRFLEQQNAALAAEV
NRLKGREPTRVAEIYEEELRELRRQVEVLTNQRARVDVERDNLLDDLQRLKAKLQEEIQLKEEAENNLAA
FRADVDAATLARIDLERRIESLNEEIAFLKKVHEEEIRELQAQLQEQQVQVEMDMSKPDLTAALRDIRAQ
YETIAAKNISEAEEWYKSKVSDLTQAANKNNDALRQAKQEMMEYRHQIQSYTCEIDALKGTNDSLMRQMR
ELEDRFASEASGYQDNIARLEEEIRHLKDEMARHLREYQDLLNVKMALDVEIATYRKLLEGEESRINLPI
QTFSALNFRETSPEQRGSEVHTKKTVMIKTIETRDGEVVSEATQQQHEVL

desmin [Sus scrofa] - pig

NCBI Reference Sequence: NP_001001535.1

>gi|48374063|ref|NP_001001535.1| desmin [Sus scrofa]
MSQAYSSSQRVSSYRRTFGGAPSFPLGSPLSSPVFPRAGFGTKGSSSSVTSRVYQVSRTSGGAGGLGPLR
ASRLGATRVPSSSYGAGELLDFSLADAVNQEFLTTRTNEKVELQELNDRFANYIEKVRFLEQQNAALAAE

VNRLKGREPTRVAEIYEEELRELRRQVEVLTNQRARVDVERDNLLDDLQRLKAKLQEEIQLKEEAENNLA
AFRADVDAATLARIDLERRIESLNEEIAFLKKVHEEEIRELQAQLQEQQVQVEMDMSKPDLTAALRDIRA
QYETIAAKNISEAEEWYKSKVSDLTQAANKNNDALRQAKQEMMEYRHQIQSYTCEIDALKGTNDSLMRQM
RELEDRFASEASGYQDNIARLEEEIRHLKDEMARHLREYQDLLNVKMALDVEIATYRKLLEGEESRINLP
IQTFSALNFRETSPEQRGSEVHTKKTVMIKTIETRDGEVVSEATQQQHEVL

PREDICTED: desmin isoform 5 [Pan troglodytes] - chimp

NCBI Reference Sequence: XP_526039.2

>gi|114583482|ref|XP_526039.2| PREDICTED: desmin isoform 5 [Pan troglodytes]
MSQAYSSSQRVSSYRRTFGGAPGFPLGSPLSSPVFPRAGFGSKGSSSSVTSRVYQVSRTSGGAGGLGSLR
ASRLGTTRTPSSYGAGELLDFSLADAVNQEFLTTRTNEKVELQELNDRFANYIEKVRFLEQQNAALAAEV
NRLKGREPTRVAELYEEELRELRRQVEVLTNQRARVDVERDNLLDDLQRLKAKLQEEIQLKEEAENNLAA
FRADVDAATLARIDLERRIESLNEEIAFLKKVHEEEIRELQAQLQEQQVQVEMDMSKPDLTAALRDIRAQ
YETIAAKNISEAEEWYKSKVSDLTQAANKNNDALRQAKQEMMEYRHQIQSYTCEIDALKGTNDSLMRQMR
ELEDRFASEASGYQDNIARLEEEIRHLKDEMARHLREYQDLLNVKMALDVEIATYRKLLEGEESRINLPI
QTYSALNFRETSPEQRGSEVHTKKTVMIKTIETRDGEVVSEATQQQHEVL

PREDICTED: LOW QUALITY PROTEIN: desmin [Gallus gallus] - chicken

NCBI Reference Sequence: XP_004942895.1

>gi|513194584|ref|XP_004942895.1| PREDICTED: LOW QUALITY PROTEIN: desmin [Gallus gallus]
MSQSYSSSQRVSSYRRTFGGGTSPVFPRASFGSRGSGSSVTSRVYQVSRTSAVPTLSTFRTTRVTPLRTY
QSAYQGAGELLDFSLADAMNQEFLQTRTNEKVELQELNDRFANYIEKVRFLEQQNALMVAEVNRLRGKEP
TRVAEMYEEELRELRRQVDALTGQRARVEVERDNLLDDLQKLKQRLQEEIQLKEEAENNLAAFRADVDAA
TLARIDLERRIESLQEEIAFLKKVHEEEIRELQAQLQEQHIQVEMDISKPDLTAALRDIRAQYESIAAKN
IAEAEEWYKSKVSDLTQAANKNNDALRQAKQEMLEYRHQIQSYTCEIDALKGTNDSLMRQMREMEERFAX
EAGGYQDTIARLEEEIRHLKDEMARHLREYQDLLNVKMALDVEIATYRKLLEGEENRISIPMHQTFASAL
NFRETCHDQRGTRCPPWALSQAIGQPWPHPWPPRPAHGARLSEPARPAQLLPRQVTASR

desmin [Danio rerio] – zebra fish

NCBI Reference Sequence: NP_001070920.1

>gi|116875779|ref|NP_001070920.1| desmin [Danio rerio]
MSHSYASSSSSSYRRMFGGPGYMTPPMSRGVFGRASTGGSGSSRVSSRVYEVSKSSTSSPGFSGYRASS
YSMPNLSAGYTRSYGGMGETLDFSLADALNQEFLHTRTNEKAELQHLNDRFANYIEKVRMLEQQNQVLGV
EVERLRGREPTRIADLYEDEMRELRREVEVVTNHRSRVEVERDNLADDLQKLKLRLQEEIALREEAENNL
SAFRADVDAATLARLDLERRIETLQEEIAFLKKIHEEEIRELQAQMQETQVQIQMDMSKPDLTAALRDIR
AQYEGIAAKNIAEAEDWYKSKVSDLNQAVSKNNEALKQSKLETMEYRHQIQSYTCEIDSLKGSNESLMRQ
MRDMEDRHGREAGALQDTIARLEAEIANMKDEMARHLREYQDLLNVKMALDVEIATYRKLLEGEESRIVM
PMQSYSTISFRETSPEHQQRASEMHSKKTVLIKTIETRDGEVVSESTQHQQDLM

desmin [Xenopus laevis] – African clawed frog

GenPept Graphics

>gi|148224383|ref|NP_001080177.1| desmin [Xenopus laevis]
MSQSYSSNQRASSYRRTFGGGSPSFSTRSSFGSKGASSSSVSSRVYQVSRSTAAPSLSSFRATRVAPVRS
SYGADVLDFSLADAMNQEFLQTRTNEKVELQDLNDRFANYIEKVRYLEQQNQILVAEVNRLKGKEPTRVN
ELYEEEMRELRRQVDLVTNQRARVEVERDNLVDDLQKLKQRLQEEIQLKEDAENNLAAFRGDVDAATLAR
IDLERRIESLQEEIAFLKKIHEEEIRELQAQFQEQQLQVEIDVSKPDLTAALRDIRAQYENIAAKNVAEA
EEWYKSKVSDLNQAAKKNNDAMRQSKQEMMEYRHQIQSYTCEIDALKGTNDSLMRQMRDLEEKFSGEAAG
YQDTIGRLEEEIRNMKDEMARHLREYQDLLNVKMALDMEIATYRKLLEGEESRITLPIQTFSALSFRETS
PEQRASEVHTKKTVMIKTIETRDGEVLSEASQQHQEIL

desmin [Xenopus (Silurana) tropicalis] – western clawed frog
GenBank: CAJ82686.1
>gi|89272455|emb|CAJ82686.1| desmin [Xenopus (Silurana) tropicalis]
MSQSYSSNQRVSSYRRTFGGGSPSFNTRTSFGSKGASSSSVSSRVYQVSRSSAAPSLSSFRATRVAPVRS
SYGADVLDFSLADAMNQEFLQTRTNEKVELQELNDRFVNYIEKVRYLEQQNQILVAEVNKLKGKEPTRVN
ELYEEEMRELRRQVDLATNQRSRVEVERDNLVDDLQKLKLRLQEEIQLKEDAENNLAAFRADVDAATLAR
IDLERRIESLQEEIAFLKKIHEEEIRELQAQMQEQQLQVEIDVSKPDLTAALRDIRAQYENIAAKNVSEA
EEWYKSKVSDLNQAAQKNNDSLRQAKQEMMEYRHQIQSYTCEIDALKGTNDSLMRQMRDLEERFSGEAAG
YQDSIGRLEEEIRNMKDEMARHLREYQDLLNVKMALDVEIATYRKLLEGEESRITLPIQTYSALSFRETS
PEQRASEVHTKKTVMIKTIETRDGEVVSEASQQHQEIL

desmin [Canis lupus familiaris]
NCBI Reference Sequence: NP_001012394.1
>gi|59958381|ref|NP_001012394.1| desmin [Canis lupus familiaris]
MSQAYSSSQRVSSYRRTFGGAGGFPLGSPLGSPVFPRAGFGTKGSSSSVTSRVYQVSRTSGGAGGLGALR
AGRLGTGRAPSYSAGELLDFSLADAVNQEFLTTRTNEKVELQELNDRFANYIEKVRFLEQQNAALAAEVN
RLKGREPTRVAEIYEEELRELRRQVEVLTNQRARVDVERDNLLDDLQRLKAKLQEEIQLKEEAENNLAAF
RADVDAATLARIDLERRIESLNEEIAFLKKVHEEEIRELQAQLQEQQVQVEMDMSKPDLTAALRDIRAQY
ETIAAKNISEAEEWYKSKVSDLTQAANKNNDALRQAKQEMMEYRHQIQSYTCEIDALKGTNDSLMRQMRE
MEDRFASEASGYQDNIARLEEEIRHLKDEMARHLREYQDLLNVKMALDVEIATYRKLLEGEESRINLPIQ
TYSALNFRETSPEQRGSEVHTKKTVMIKTIETRDGEVVSEATQQQHEVL

desmin [Mesocricetus auratus] – Golden hamster
NCBI Reference Sequence: NP_001268541.1
>gi|528078510|ref|NP_001268541.1| desmin [Mesocricetus auratus]
MSQAYSSSQRVSSYRRTFGGAPSFSLGSPLSSPVFPRAGFGTKGSSSSVTSRVYQVSRTSGGAGGLGSLR
ASRLGSTRAPSYGAGELLDFSLADAVNQEFLATRTNEKVELQELNDRFANYIEKVRFLEQQNAALAAEVN
RLKGREPTRVAELYEEEMRELRRQVEVLTNQRARVDVERDNLIDDLQRLKAKLQEEIQLREEAENNLAAF
RADVDAATLARIDLERRIESLNEEIAFLKKVHEEEIRELQAQLQEQQVQVEMDMSKPDLTAALRDIRAQY
ETIAAKNISEAEEWYKSKVSDLTQAANKNNDALRQAKQEMMEYRHQIQSYTCEIDALKGTNDSLMRQMRE
LEDRFASEASGYQDNIARLEEEIRHLKDEMARHLREYQDLLNVKMALDVEIATYRKLLEGEESRINLPIQ
TFSALNFRETSPEQRGSEVHTKKTVMIKTIETRDGEVVSEATQQQHEVL

desmin [Papioanubis] – Olive Baboon

NCBI Reference Sequence: NP_001162488.1

>gi|281182830|ref|NP_001162488.1| desmin [Papioanubis]

MSQAYSSSQRVSSYRRTFGGAPSFPLGSPLSSPVFPRAGFGSKGSSSSVTSRVYQVSRTSGGAGGLGSLR
ASRLGTTRAPSSYGAGELLDFSLADAVNQEFLTTRTNEKVELQELNDRFANYIEKVRFLEQQNAALAAEV
NRLKGREPTRVAELYEEELRELRRQVEVLTNQRARVDVERDNLLDDLQRLKAKLQEEIQLKEEAENNLAA
FRADVDAATLARIDLERRIESLNEEIAFLKKVHEEEIRELQAQLQEQQVQVEMDMSKPDLTAALRDIRAQ
YETIAAKNISEAEEWYKSKVSDLTQAANKNNDALRQAKQEMMEYRHQIQSYTCEIDALKGTNDSLMRQMR
ELEDRFASEASGYQDNIARLEEEIRHLKDEMARHLREYQDLLNVKMALDVEIATYRKLLEGEESRINLPI
QTFSALNFRETSPEQRGSEVHTKKTVMIKTIETRDGEVVSEATQQQHEVL

PREDICTED: desmin [Ovis aries] - Sheep

NCBI Reference Sequence: XP_004005001.1

>gi|426221609|ref|XP_004005001.1| PREDICTED: desmin [Ovis aries]

MSQAYSSSQRVSSYRRTFGGAPSFPLGSPLSSPVFPRAGFGTKGSSSSVTSRVYQVSRTSGGAGGLGALR
ASRLGSTRVPSSYGAGELLDFSLADAVNQEFLTTRTNEKVELQELNDRFANYIEKVRFLEQQNAALAAEV
NRLKGREPTRVAEIYEEELRELRRQVEVLTNQRARVDVERDNLLDDLQRLKAKLQEEIQLKEEAENNLAA
FRADVDAATLARIDLERRIESLNEEIAFLKKVHEEEIRELQAQLQEQQVQVEMDMSKPDLTAALRDIRAQ
YETIAAKNISEAEEWYKSKVSDLTQAANKNNDALRQAKQEMMEYRHQIQSYTCEIDALKGTNDSLMRQMR
ELEDRFASEASGYQDNIARLEEEIRHLKDEMARHLREYQDLLNVKMALDVEIATYRKLLEGEESRINLPI
QTFSALNFRETSPEQRGSEVHTKKTVMIKTIETRDGEVVSEATQQQHEVL

PREDICTED: desmin [Mustela putorius furo] - Ferret

NCBI Reference Sequence: XP_004762908.1

>gi|511886224|ref|XP_004762908.1| PREDICTED: desmin [Mustela putorius furo]

MSQAYSSSQRVSSYRRTFGGAGGFPLVSPLGSPVFPRAGFGTKGSSSSMMSRVYQVSRTSGGAGGLGALR
ASRLGSARAPSSYGAGELLDFSLADAVNQEFLTTRTNEKVELQELNDRFANYIEKVRFLEQQNAALAAEV
NRLKGREPTRVAEIYEEELRELRRQVEVLTNQRARVDVERDNLLDDLQRLKAKLQEEIQLREEAENNLAA
FRADVDAATLARIDLERRIESLNEEIVFLKKVHEEEIRELQAQLQEQQVQVEMDMSKPDLTAALRDIRAQ
YETIAAKNISEAEEWYKSKVSDLTQAANKNNDALRQAKQEMMEYRHQIQSYTCEIDALKGTNDSLMRQMR
EMEDRFASEASGYQDNIARLEEEIRHLKDEMARHLREYQDLLNVKMALDVEIATYRKLLEGEESRINLPI
QTFSALNFRETSPEQRGSEVHTKKTVMIKTIETRDGEVVSEATQQQHEVL

desmin [Scyliorhinus stellaris] – Dog fish

GenBank: CAC83054.1

>gi|17221342|emb|CAC83054.1| desmin [Scyliorhinus stellaris]

MSFATSASSSSYRRTFGGNLMGSPQMSRASYSGGRFGGGGSNLSSRLLISKSSALPSYSSYRSKVPSRGY
EVVDFNIADALNQEFVQTRTNEKAEMQHLNDRFASYIEKVRFLEQQNKVLVAEVNRLKGQEPGRVNDLYE
QEMRDLRRQVDALTNEKVRVEVERDNLADDLLKLKQRLQDEIQLKEEAENTLAAFRHDVDAATLARLDLE
RRIEALQEEIAFLKKVHEEELRELQAQLQEQHVQVEMDLSKPDLTAALRDIRAQYECISAKNVQEAEEWY
KSKVSDLNQNATKNNDALRQAKQEVMEYRHQVQSYTCEIDALKGSNESLMRQMREMEERFAGEAAGYTDS
IARLEEEIRHLKDEMARHLREYQDLLNVKMALDVEIATYRKLLEGEESRIVVPVQSFSTLSFRETSPEQG
HAAEVHTKKTVTIKTYETRDGEVISESSQHQQEIS

PREDICTED: desmin [Columba livia] - Pigeon

NCBI Reference Sequence: XP_005498663.1

>gi|543714756|ref|XP_005498663.1| PREDICTED: desmin [Columba livia]

MSQSYSSSQRVSSYRRTFGGSPVFGGSPVFPRASFGTKGSSGSSVTSRVYQVSRPSAVPSLSSFRTTRVT
PLRSYQSAYQGAGELLDFSLADAMNQEFLQTRTNEKVELQELNDRFANYIEKVRFLEQQNALMVAEVNRL
RGKEPTRVAEMYEEELRELRRQVDALTGQRARVEVERDNLLDDLQKLKQRLQEEIQLKEEAENNLAAFRA
DVDAATLARIDLERRIESLQEEIAFLKKVHEEEIRELQAQLQEQHIQVEMDISKPDLTAALRDIRAQYES
IAAKNIAEAEEWYKSKVSDLTQAANKNNDALRQAKQEMLEYRHQIQSYTCEIDALKGTNDSLMRQMREME
ERFAGEAGGYQDTIARLEEEIRHLKDEMARHLREYQDLLNVKMALDVEIATYRKLLEGEENRISIPMHQT
FASALNFRETSPDQRGSEVHTKKTVMIKTIETRDGEVVSEATQQQHEVL

# Appendix III

## A 4.1 Rare gene variants identified at the 5q12-q14 locus (GLH5 family)

| Gene name | Genomic position (Grch38) | Allele | Consequence | cDNA change | Amino acid change | rs ID |
|---|---|---|---|---|---|---|
| *PDE4D* | 60414870 | C/T | intron | c.-90+73072G>A | - | rs1417982737 |
| *PDE4D* | 60514741 | C/T | intron | c.-90+7310G>A | - | rs551579921 |
| *DEPDC1B* | 60620943 | G/A | intron | c.899-15087C>T | - | rs1219158005 |
| *DEPDC1B* | 60651967 | -/TTCCTCGTACATTGTGGATATTAGTC | intron | c.315-4435_315-4434insGACTAATATCCACAATGTACGAGGAA | - | rs1301922761 |
| *DEPDC1B* | 60698029 | T/- | intron | c.48+2017del | - | rs200688717 |
| *ELOVL7* | 60763645 | C/T | intron | c.499+582G>A | - | rs560441905 |
| *ELOVL7* | 60763645 | C/T | downstream | - | - | rs560441905 |
| *ELOVL7* | 60781147 | T/A | intron | c.64+6187A>T | - | rs780962916 |
| *ELOVL7* | 60817854 | A/G | intron | c.-35+26306T>C | - | rs769320359 |
| *ELOVL7* | 60819479 | T/G | intron | c.-35+24681A>C | - | rs1020084770 |
| *NDUFAF2* | 60950641 | -/TT | intron | c.127+5274_127+5275dup | - | rs1309492733 |
| *NDUFAF2* | 61070627 | -/CACT | intron | c.128-2497_128-2496insACTC | - | rs1244534996 |
| *NDUFAF2* | 61081266 | T/C | intron | c.217+8052T>C | - | - |
| *NDUFAF2* | 61105555 | -/A | intron | c.258+6533dup | - | rs903009606 |
| *NDUFAF2* | 61106478 | G/A | intron | c.258+7446G>A | - | rs550373310 |
| *NDUFAF2* | 61111106 | G/T | intron | c.258+12074G>T | - | rs567494900 |
| *NDUFAF2* | 61122949 | C/A | intron | c.258+23917C>A | - | rs549354711 |
| *NDUFAF2* | 61143970 | TTGTGTGTGTG/"TTG | intron | - | - | - |
| *ZSWIM6* | 61358636 | A/G | intron | c.676+25688A>G | - | rs532353624 |
| *ZSWIM6* | 61364932 | A/G | intron | c.676+31984A>G | - | - |
| *ZSWIM6* | 61405681 | G/A | intron | c.677-67000G>A | - | rs370544488 |
| *ZSWIM6* | 61429842 | T/C | intron | c.677-42839T>C | - | rs191357319 |
| *ZSWIM6* | 61431328 | G/C | intron | c.677-41353G>C | - | rs542792589 |
| *ZSWIM6* | 61450372 | A/G | intron | c.677-22309A>G | - | rs370909643 |
| *ZSWIM6* | 61493707 | T/G | intron | c.1183-553T>G | - | - |
| *KIF2A* | 62330690 | A/G | intron | c.-17-16440A>G | - | - |
| *KIF2A* | 62362031 | -/A | intron | c.947-405dup | - | rs1029470528 |
| *KIF2A* | 62376126 | C/A | intron | c.1717-1535C>A | - | - |
| *DIMT1* | 62396139 | -/TTTTTTTTTTTTTTTTTTT | intron | c.447-1533_447-1532insAAAAAAAAAAAAAAAAAAAA | - | - |
| *IPO11* | 62435857 | C/T | intron | c.-6-1417C>T | - | rs546680169 |
| *IPO11* | 62467980 | C/G | intron | c.649+717C>G | - | rs566722617 |
| *IPO11* | 62483049 | T/A | intron | c.829-52T>A | - | rs566434659 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *IPO11* | 62484873 | C/T | intron | c.1175-546C>T | - | rs534167278 |
| *IPO11* | 62559598 | -/T | intron | c.2461-1529dup | - | rs893248438 |
| *RNF180* | 64166125 | C/T | intron | c.-1+172C>T | - | - |
| *RNF180* | 64166125 | C/T | upstream | - | - | - |
| *RNF180* | 64174961 | TT/- | intron | c.-1+9031_-1+9032del | - | - |
| *RNF180* | 64192939 | -/TATATATATATATAT | intron | c.1-7869_1-7868insTATATATATATATAT | - | rs1554028955 |
| *RNF180* | 64201324 | T/C | intron | c.135+382T>C | - | rs138388925 |
| *RNF180* | 64224074 | GTGTGTGTGTGT/- | intron | c.1227+6697_1227+6708del | - | rs10694125 |
| *RNF180* | 64228915 | -/TTTT | intron | c.1227+11540_1227+11543dup | - | rs373212886 |
| *RNF180* | 64239690 | G/A | intron | c.1227+22294G>A | - | - |
| *RNF180* | 64255025 | A/G | intron | c.1227+37629A>G | - | rs559505101 |
| *RNF180* | 64260113 | A/G | intron | c.1227+42717A>G | - | rs754572276 |
| *RNF180* | 64282635 | A/C | intron | c.1228-42551A>C | - | - |
| *SHISAL2B* | 64713513 | G/A | intron | c.350-4376G>A | - | rs796225858 |
| *SREK1IP1* | 64743207 | A/G | intron | c.62-2007T>C | - | - |
| *SREK1IP1* | 64752144 | -/TTTTTTT | intron | c.61+2170_61+2171insAAAAAAA | - | - |
| *CWC27* | 64786878 | A/T | intron | c.599+251A>T | - | rs147048868 |
| *CWC27* | 64834827 | G/A | intron | c.938+30441G>A | - | rs145213940 |
| *CWC27* | 64839938 | G/C | intron | c.938+35552G>C | - | rs191914678 |
| *CWC27* | 64840211 | C/A | intron | c.938+35825C>A | - | - |
| *CWC27* | 64850225 | -/AAAAAAAAAAAA | intron | c.939-35209_939-35208insAAAAAAAAAAA | - | rs61174118 |
| *CWC27* | 64891767 | A/G | intron | c.1042+6221A>G | - | rs1214406432 |
| *CWC27* | 64991454 | G/A | intron | c.1256+14216G>A | - | - |
| *ADAMTS6* | 65179448 | C/A | intron | c.2911-6440G>T | - | rs10940017 |
| *ADAMTS6* | 65217862 | A/G | intron | c.2273-2375T>C | - | rs372153259 |
| *ADAMTS6* | 65230880 | -/AT | intron | c.1934-4663_1934-4662dup | - | - |
| *ADAMTS6* | 65236145 | T/C | intron | c.1933+5959A>G | - | rs373248800 |
| *ADAMTS6* | 65280708 | TTTTA/- | intron | c.1513-7261_1513-7257del | - | rs373879665 |
| *ADAMTS6* | 65284955 | T/C | intron | c.1512+6374A>G | - | rs375128376 |
| *ADAMTS6* | 65318753 | G/A | intron | c.1223+10625C>T | - | - |
| *ADAMTS6* | 65337050 | G/C | intron | c.1074-2965C>G | - | rs369308235 |
| *ADAMTS6* | 65361479 | C/T | intron | c.1074-27394G>A | - | rs369842353 |
| *ADAMTS6* | 65442633 | -/A | intron | c.1073+8841dup | - | rs942019307 |
| *CENPK* | 65524680 | -/A | intron | c.598-3153dup | - | rs753826622 |
| *CENPK* | 65547222 | G/A | intron | c.241+4342C>T | - | rs371367880 |
| *CENPK* | 65554836 | A/T | synonymous | c.72T>A | p.Leu24%3D | - |
| *CENPK* | 65554836 | A/T | 5' UTR | c.-19T>A | - | - |
| *CENPK* | 65554836 | A/T | intron | c.49+6627T>A | - | - |
| *CENPK* | 65554836 | A/T | upstream | - | - | - |

| | | | | | | |
|---|---|---|---|---|---|---|
| *TRIM23* | 65603314 | G/A | intron | c.1179+1597C>T | - | rs373915607 |
| *TRAPPC13* | 65634016 | T/C | intron | c.47-1285T>C | - | rs1449725778 |
| *TRAPPC13* | 65635742 | T/C | intron | c.116-202T>C | - | rs548523343 |
| *TRAPPC13* | 65641125 | G/T | intron | c.300+3345G>T | - | rs536727768 |
| *TRAPPC13* | 65646860 | G/A | intron | c.301-195G>A | - | rs536535119 |
| *TRAPPC13* | 65650274 | G/A | intron | c.429-536G>A | - | - |
| *TRAPPC13* | 65666110 | A/G | downstream | - | - | - |
| *SGTB* | 65666110 | A/G | 3' UTR | c.*4136T>C | - | - |
| *TRAPPC13* | 65666110 | A/G | 3' UTR | c.*1499A>G | - | - |
| *SGTB* | 65678213 | C/T | intron | c.681+2281G>A | - | rs375985377 |
| *SGTB* | 65698614 | G/T | intron | c.374+5665C>A | - | rs368877824 |
| *NLN* | 65739432 | G/A | intron | c.41+17018G>A | - | rs373309140 |
| *NLN* | 65775621 | C/T | intron | c.451-1806C>T | - | rs557865360 |
| *NLN* | 65802436 | G/A | intron | c.1528-7079G>A | - | rs375302124 |
| *NLN* | 65806639 | A/G | intron | c.1528-2876A>G | - | rs375958689 |
| *NLN* | 65832015 | AA/- | intron | c.777+19640_777+19641del | - | rs35924843 |
| *NLN* | 65832026 | A/T | intron | c.777+19635A>T | - | rs986153678 |
| *NLN* | 65835124 | T/C | intron | c.777+22733T>C | - | - |
| *ERBIN* | 65959188 | -/T | intron | c.-57-29437dup | - | rs929688196 |
| *ERBIN* | 65966715 | C/A | intron | c.-57-21920C>A | - | rs1345572844 |
| *ERBIN* | 65972431 | C/T | intron | c.-57-16204C>T | - | rs553806449 |
| *ERBIN* | 66007548 | AA/- | intron | c.308-4490_308-4489del | - | rs34972803 |
| *ERBIN* | 66010830 | G/T | intron | c.308-1219G>T | - | rs549522123 |
| *ERBIN* | 66022212 | T/C | intron | c.597+827T>C | - | rs371091766 |
| *ERBIN* | 66031860 | G/T | intron | c.1206+3517G>T | - | rs189365089 |
| *ERBIN* | 66040147 | C/T | intron | c.1306+1665C>T | - | rs557018075 |
| *ERBIN* | 66041936 | A/T | intron | c.1307-1141A>T | - | - |
| *ERBIN* | 66052089 | C/G | intron | c.2087+1123C>G | - | rs370332842 |
| *ERBIN* | 66063562 | G/A | intron | c.3634-8607G>A | - | rs1008481856 |
| *SREK1* | 66140881 | C/T | intron | c.-312+870C>T | - | - |
| *SREK1* | 66148494 | T/C | intron | c.161+3957T>C | - | rs376686015 |
| *SREK1* | 66164240 | G/A | intron | c.886+318G>A | - | rs372614474 |
| *SREK1* | 66184040 | -/A | downstream | - | - | rs559216238 |
| *MAST4* | 66643693 | -/T | intron | c.363+46684dup | - | rs1330225216 |
| *MAST4* | 66671861 | C/T | intron | c.363+74843C>T | - | rs527420482 |
| *MAST4* | 66724523 | C/T | intron | c.364-35186C>T | - | - |
| *MAST4* | 66729129 | T/C | intron | c.364-30580T>C | - | - |
| *MAST4* | 66789992 | TTTT/- | intron | c.642+1221_642+1224del | - | rs57743987 |
| *MAST4* | 66793437 | T/G | intron | c.642+4643T>G | - | - |
| *MAST4* | 66907159 | AGCG/- | intron | c.674+7176_674+7179del | - | - |
| *MAST4* | 66991743 | C/T | intron | c.675-62661C>T | - | - |
| *MAST4* | 67026246 | C/G | intron | c.92+21161C>G | - | rs767507219 |
| *MAST4* | 67085731 | G/A | intron | c.182-4431G>A | - | rs183409029 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *CD180* | 67182182 | C/T | 3' UTR | c.*675G>A | - | - |
| *CD180* | 67182182 | C/T | downstream | - | - | - |
| *PIK3R1* | 68214949 | AA/- | upstream | - | - | rs1316900119 |
| *PIK3R1* | 68217667 | CT/- | intron | c.-387+1717_-387+1718del | - | - |
| *PIK3R1* | 68217669 | -/GCGCGCGG | intron | c.-387+1721_-387+1722insCGCGCGGG | - | - |
| *PIK3R1* | 68227325 | TTATAGG/- | intron | c.334+310_334+316del | - | - |
| *PIK3R1* | 68227327 | G/- | intron | c.334+319del | - | - |
| *PIK3R1* | 68227328 | G/C | intron | c.334+319G>C | - | - |
| *PIK3R1* | 68227617 | A/G | intron | c.334+608A>G | - | rs530731247 |
| *PIK3R1* | 68237893 | G/A | intron | c.334+10884G>A | - | - |
| *PIK3R1* | 68246463 | C/T | intron | c.40+6526C>T | - | rs1383695692 |
| *PIK3R1* | 68260815 | A/G | intron | c.41-12575A>G | - | rs546497626 |
| *PIK3R1* | 68294742 | A/T | intron | c.668+64A>T | - | rs535033913 |
| *PIK3R1* | 68294742 | A/T | downstream | - | - | rs535033913 |
| *PIK3R1* | 68297199 | T/C | intron | c.1086-213T>C | - | rs1166065523 |
| *SLC30A5* | 69111446 | A/C | intron | c.448-1694A>C | - | rs560440158 |
| *SLC30A5* | 69120664 | T/C | intron | c.1570-1030T>C | - | rs371242858 |
| *SLC30A5* | 69131122 | G/C | downstream | - | - | rs749487065 |
| *CENPH* | 69190662 | -/TTAA | intron | c.134+895_134+898dup | - | - |
| *MRPS27* | 72242429 | A/- | intron | c.282-4301del | - | rs750987632 |
| *MRPS27* | 72264541 | A/G | intron | c.282-26413T>C | - | - |
| *MRPS27* | 72287042 | A/C | intron | c.281+8489T>G | - | rs1055365435 |
| *PTCD2* | 72360798 | C/T | 3' UTR | c.*2371C>T | - | - |
| *PTCD2* | 72362653 | A/G | 3' UTR | c.*4226A>G | - | rs531904274 |
| *ZNF366* | 72450872 | G/A | intron | c.1525-3455C>T | - | - |
| *TNPO1* | 72876255 | G/A | intron | c.801+518G>A | - | rs542656367 |
| *FCHO2* | 72983922 | T/C | intron | c.126-5505T>C | - | rs555520675 |
| *FCHO2* | 73032615 | GAT/- | intron | c.797-2037_797-2035del | - | rs546964572 |
| *TMEM171* | 73123849 | G/C | missense | c.476G>C | p.Gly159Ala | COSV55129751 |
| *UTP15* | 73580757 | -/T | 3' UTR | c.*675dup | - | rs1034231085 |
| *UTP15* | 73580757 | -/T | downstream | - | - | rs1034231085 |
| *ARHGEF28* | 73676239 | TG/- | intron | c.-11-8602_-11-8601del | - | rs202060147 |
| *ARHGEF28* | 73805673 | T/C | intron | c.1024+10282T>C | - | - |
| *ARHGEF28* | 73810450 | T/C | intron | c.1024+15059T>C | - | - |
| *ARHGEF28* | 73822480 | A/T | intron | c.1025-9858A>T | - | - |
| *ARHGEF28* | 73841946 | TTATC/- | intron | c.1427+1186_1427+1190del | - | - |
| *ARHGEF28* | 73891339 | C/T | intron | c.3388-713C>T | - | rs1561489660 |
| *ENC1* | 74634415 | AAT/- | intron | c.*32+269_*32+271del | - | rs552080426 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *HEXB* | 74661013 | GA/- | intron | c.-377+20472_-377+20473del | - | rs141154481 |
| *HEXB* | 74661234 | A/G | intron | c.-377+20676A>G | - | rs79454547 |
| *HEXB* | 74682361 | -/AAAAC | intron | c.-376-6952_-376-6948dup | - | - |
| *HEXB* | 74700105 | C/T | intron | c.669+2999C>T | - | rs559089327 |
| *HEXB* | 74708621 | AAAA/- | intron | c.771+3301_771+3304del | - | - |
| *NSA2* | 74766304 | A/T | upstream | - | - | rs567010282 |
| *GFM2* | 74766304 | A/T | intron | c.-25+634T>A | - | rs567010282 |
| *NSA2* | 74773802 | C/T | intron | c.523-2802C>T | - | rs554698672 |
| *FAM169A* | 74838580 | A/C | intron | c.318+385T>G | - | - |
| *FAM169A* | 74838580 | A/C | downstream | - | - | - |
| *GCNT4* | 75034971 | C/A | intron | c.-1-4933G>T | - | rs1255318909 |
| *GCNT4* | 75048995 | C/T | intron | c.-142-958G>A | - | rs552931788 |
| *ANKRD31* | 75093209 | G/A | intron | c.5161-1808C>T | - | rs139159707 |
| *ANKRD31* | 75097085 | G/C | intron | c.5161-5684C>G | - | rs1043701326 |
| *ANKRD31* | 75154293 | T/C | missense | c.1760A>G | p.Asp587Gly | rs768240858 |
| *ANKRD31* | 75160751 | C/T | intron | c.1708-6406G>A | - | rs1018560474 |
| *ANKRD31* | 75210515 | A/G | intron | c.326+313T>C | - | rs575769762 |
| *ANKRD31* | 75217332 | C/A | intron | c.288+4917G>T | - | rs186514757 |
| *HMGCR* | 75337270 | G/- | 5' UTR | c.-51del | - | rs1412704333 |
| *HMGCR* | 75337270 | G/- | upstream | - | - | rs1412704333 |
| *HMGCR* | 75337270 | G/- | intron | c.-24+656del | - | rs1412704333 |
| *CERT1* | 75382099 | T/C | intron | c.1489-22A>G | - | rs554480700 |
| *COL4A3BP* | 75382099 | T/C | intron | c.1873-22A>G | - | rs554480700 |
| *CERT1* | 75400704 | TTATATATAA/- | intron | c.1018-407_1018-398del | - | - |
| *CERT1* | 75400704 | TTATATATAA/- | upstream | - | - | - |
| *COL4A3BP* | 75400704 | TTATATATAA/- | intron | c.1402-407_1402-398del | - | - |
| *CERT1* | 75415766 | -/T | intron | c.837+1109dup | - | rs1235931030 |
| *COL4A3BP* | 75415766 | -/T | intron | c.1221+1109dup | - | rs1235931030 |
| *POLK* | 75532458 | -/T | intron | c.-13-14540dup | - | rs745417117 |
| *POLK* | 75576972 | -/A | intron | c.694+50dup | - | rs748170450 |
| *POLK* | 75588764 | A/G | intron | c.1260-1580A>G | - | rs920822377 |
| *POLK* | 75591417 | A/G | intron | c.1356+977A>G | - | - |
| *POLK* | 75591417 | A/G | downstream | - | - | - |
| *POLK* | 75606432 | T/C | downstream | - | - | - |
| *ANKDD1B* | 75642787 | TG/- | intron | c.798+6904_798+6905del | - | - |
| *ANKDD1B* | 75643440 | T/C | intron | c.798+7558T>C | - | rs1282424 |
| *POC5* | 75693241 | T/C | intron | c.691-741A>G | - | rs1458542557 |
| *POC5* | 75706569 | -/T | intron | c.224-783dup | - | rs796946990 |
| *POC5* | 75717964 | A/C | upstream | - | - | rs190691992 |
| *SV2C* | 75876130 | T/C | intron | c.-102+27642T>C | - | rs1413582775 |
| *SV2C* | 75881815 | C/T | intron | c.-102+33327C>T | - | rs1246271886 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *SV2C* | 75882089 | C/- | intron | c.-102+33601del | - | rs1284048891 |
| *SV2C* | 75882176 | G/T | intron | c.-102+33688G>T | - | rs937118128 |
| *SV2C* | 75882177 | A/T | intron | c.-102+33689A>T | - | rs1243043186 |
| *SV2C* | 75882202 | G/A | intron | c.-102+33714G>A | - | rs1457993182 |
| *SV2C* | 75947436 | C/T | intron | c.-102+98948C>T | - | rs1306944684 |
| *SV2C* | 76035226 | -/T | intron | c.-101-96413dup | - | rs1181946600 |
| *SV2C* | 76077198 | T/C | intron | c.-101-54452T>C | - | - |
| *SV2C* | 76113889 | C/T | intron | c.-101-17761C>T | - | rs149527206 |
| *SV2C* | 76170148 | T/G | intron | c.581-24771T>G | - | rs546317209 |
| *SV2C* | 76257320 | -/GT | intron | c.914-27832_914-27831dup | - | - |
| *SV2C* | 76312266 | G/T | intron | c.2000+10721G>T | - | rs370142721 |
| *IQGAP2* | 76484806 | TC/- | intron | c.146+23151_146+23152del | - | rs1297304528 |
| *IQGAP2* | 76568833 | A/G | intron | c.304-1747A>G | - | rs553611224 |
| *IQGAP2* | 76572836 | A/G | intron | c.381+2179A>G | - | rs1038855554 |
| *IQGAP2* | 76610041 | G/T | intron | c.1358-979G>T | - | rs1001249761 |
| *IQGAP2* | 76618250 | G/C | intron | c.1521+7067G>C | - | rs762944161 |
| *F2RL2* | 76618250 | G/C | missense | c.457C>G | p.His153Asp | rs762944161 |
| *IQGAP2* | 76650847 | C/T | intron | c.2095-1903C>T | - | rs531481245 |
| *ZBED3* | 77083842 | G/A | intron | c.-153+3269C>T | - | rs768652733 |
| *AC022414.1* | 77147136 | T/G | intron | c.-34+28615T>G | - | - |
| *PDE8B* | 77315048 | A/G | intron | c.399+2995A>G | - | rs764026369 |
| *AC022414.1* | 77315048 | A/G | intron | c.27+2995A>G | - | rs764026369 |
| *PDE8B* | 77392057 | A/G | intron | c.1168-8191A>G | - | - |
| *AC022414.1* | 77392057 | A/G | intron | c.796-8191A>G | - | - |
| *WDR41* | 77478444 | G/A | intron | c.167+11013C>T | - | rs779081277 |
| *WDR41* | 77493075 | A/G | upstream | - | - | - |
| *WDR41* | 77493075 | A/G | intron | c.43-3503T>C | - | - |
| *WDR41* | 77549083 | G/A | intron | c.43-59511C>T | - | rs549986895 |
| *WDR41* | 77570601 | T/A | intron | c.42+49878A>T | - | rs1180719903 |
| *WDR41* | 77602243 | G/A | intron | c.42+18236C>T | - | rs371634531 |
| *WDR41* | 77611284 | C/A | intron | c.42+9195G>T | - | rs1334574621 |
| *OTP* | 77635398 | A/C | intron | c.447+1423T>G | - | - |
| *TBCA* | 77790697 | A/G | intron | c.-20+77937T>C | - | rs76581280 |
| *TBCA* | 77799832 | G/- | intron | c.-20+68802del | - | rs543181854 |
| *TBCA* | 77841824 | T/A | intron | c.-20+26810A>T | - | - |
| *TBCA* | 77845100 | T/G | intron | c.-20+23534A>C | - | - |
| *TBCA* | 77850916 | G/A | intron | c.-20+17718C>T | - | - |
| *AP3B1* | 78122835 | G/T | intron | c.1968+5195C>A | - | - |
| *AP3B1* | 78122842 | A/C | intron | c.1968+5188T>G | - | - |
| *AP3B1* | 78164789 | A/G | intron | c.1230+821T>C | - | rs576629360 |
| *AP3B1* | 78275483 | -/T | intron | c.129-7889dup | - | rs935681124 |
| *LHFPL2* | 78533227 | T/C | intron | c.-185-22829A>G | - | rs573970161 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *LHFPL2* | 78563401 | G/C | intron | c.-186+1412C>G | - | rs539227969 |
| *LHFPL2* | 78600652 | C/T | intron | c.-245+31612G>A | - | rs1382660712 |
| *LHFPL2* | 78647829 | -/A | intron | c.-350+669dup | - | rs1009848200 |
| *LHFPL2* | 78647829 | -/A | upstream | - | - | rs1009848200 |
| *LHFPL2* | 78704961 | TGTGATAG ATGTG/- | intron | c.-350+13123_-350+13135del | - | rs1221883063 |
| *LHFPL2* | 78713154 | G/A | intron | c.-350+4930C>T | - | rs1410113712 |
| *LHFPL2* | 78713313 | G/T | intron | c.-350+4771C>A | - | rs1174823208 |
| *LHFPL2* | 78713314 | A/T | intron | c.-350+4770T>A | - | rs1376656994 |
| *LHFPL2* | 78713340 | T/C | intron | c.-350+4744A>G | - | rs1351828225 |
| *LHFPL2* | 78713347 | C/T | intron | c.-350+4737G>A | - | rs1300181766 |
| *ARSB* | 78847235 | TCC/- | intron | c.1143-7809_1143-7807del | - | - |
| *ARSB* | 78860954 | C/A | intron | c.1143-21528G>T | - | rs535028539 |
| *ARSB* | 78871220 | C/T | intron | c.1142+14364G>A | - | rs1296152883 |
| *ARSB* | 78930777 | A/G | intron | c.898+24518T>C | - | - |
| *ARSB* | 78965599 | C/G | intron | c.500-993G>C | - | - |
| *DMGDH* | 79054023 | C/T | intron | c.540+161G>A | - | - |
| *JMY* | 79258917 | C/T | intron | c.1033-18993C>T | - | - |
| *JMY* | 79275174 | T/C | intron | c.1033-2736T>C | - | rs79342964 |
| *JMY* | 79282501 | A/C | intron | c.1206+4418A>C | - | - |
| *JMY* | 79293624 | A/G | intron | c.1527+2325A>G | - | rs181427991 |
| *HOMER1* | 79375001 | C/T | downstream | - | - | rs533348338 |
| *HOMER1* | 79375001 | C/T | 3' UTR | c.*1008G>A | - | rs533348338 |
| *HOMER1* | 79427370 | -/T | intron | c.294+23619dup | - | rs550570484 |
| *HOMER1* | 79441898 | C/T | intron | c.294+9092G>A | - | - |
| *HOMER1* | 79483392 | T/C | intron | c.6-26374A>G | - | - |
| *TENT2* | 79658765 | T/C | intron | c.1071+1764T>C | - | - |
| *PAPD4* | 79658765 | T/C | intron | c.1071+1764T>C | - | - |
| *TENT2* | 79670386 | G/A | intron | c.1208+1358G>A | - | rs532297293 |
| *PAPD4* | 79670386 | G/A | intron | c.1208+1358G>A | - | rs532297293 |
| *CMYA5* | 79713292 | GCTGCA/- | intron | c.150-15628_150-15623del | - | - |
| *CMYA5* | 79713296 | C/G | intron | c.150-15619C>G | - | - |
| *CMYA5* | 79718398 | AT/- | intron | c.150-10514_150-10513del | - | rs541182855 |
| *CMYA5* | 79730596 | C/A | missense | c.1831C>A | p.Gln611Lys | rs373188852 |
| *CMYA5* | 79748528 | CCTATCTAT /"CCTAT | intron | - | - | - |
| *THBS4* | 79997406 | TA/- | intron | c.-229-922_-229-921del | - | rs1028195063 |
| *THBS4* | 80005791 | G/T | intron | c.-186+7416G>T | - | rs539947461 |
| *THBS4* | 80064971 | A/- | intron | c.1126-438del | - | - |
| *THBS4* | 80083825 | G/A | downstream | - | - | - |
| *SERINC5* | 80169971 | G/T | intron | c.552-425C>A | - | rs924450804 |
| *SERINC5* | 80183313 | G/A | intron | c.196-5249C>T | - | rs1025938379 |
| *SERINC5* | 80191480 | A/- | intron | c.195+11406del | - | rs1167390197 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *SERINC5* | 80195362 | T/G | intron | c.195+7524A>C | - | - |
| *SERINC5* | 80232321 | A/G | intron | c.27+23575T>C | - | rs550944856 |
| *SERINC5* | 80232322 | A/T | intron | c.27+23574T>A | - | rs576955815 |
| *SERINC5* | 80244068 | G/T | intron | c.27+11828C>A | - | rs556483749 |
| *ZFYVE16* | 80448413 | G/T | intron | c.3103+9G>T | - | rs543112994 |
| *ZFYVE16* | 80475488 | TCAGTTC/- | intron | c.4461+652_4461+658del | - | - |
| *FAM151B* | 80534677 | G/A | intron | c.672-6996G>A | - | rs1212217745 |
| *ANKRD34B* | 80570376 | C/T | upstream | - | - | rs554996227 |
| *DHFR* | 80626857 | TTT/- | 3' UTR | c.*2230_*2232del | - | - |
| *DHFR* | 80628209 | -/A | 3' UTR | c.*877dup | - | rs547127745 |
| *DHFR* | 80628209 | -/A | downstream | - | - | rs547127745 |
| *DHFR* | 80646634 | T/C | intron | c.242+2755A>G | - | rs547440317 |
| *MSH3* | 80708055 | A/G | intron | c.1341-17398A>G | - | rs530827133 |
| *MSH3* | 80709237 | G/A | intron | c.1341-16216G>A | - | rs553843898 |
| *MSH3* | 80725217 | AAA/- | intron | c.1341-222_1341-220del | - | rs33919693 |
| *MSH3* | 80812211 | C/T | intron | c.2656-1373C>T | - | rs547059748 |
| *MSH3* | 80814143 | AAAAAAC/- | intron | c.2813+401_2813+407del | - | rs1352527938 |
| *RASGRF2* | 80960476 | G/T | 5' UTR | c.-263G>T | - | - |
| *RASGRF2* | 80960476 | G/T | upstream | - | - | - |
| *RASGRF2* | 80992266 | A/G | intron | c.288+31240A>G | - | rs1433070682 |
| *RASGRF2* | 80994355 | C/G | intron | c.288+33329C>G | - | rs537383454 |
| *RASGRF2* | 80994360 | T/C | intron | c.288+33334T>C | - | rs1303632399 |
| *RASGRF2* | 81049206 | G/- | intron | c.395+6223del | - | rs1336139340 |
| *ZCCHC9* | 81310161 | TAAAA/"T | intron | - | - | - |
| *ACOT12* | 81310161 | TAAAA/"T | 3' UTR | - | - | - |
| *ACOT12* | 81339256 | G/A | intron | c.1129-3355C>T | - | rs530539279 |
| *ACOT12* | 81354764 | T/C | intron | c.496+5139A>G | - | rs555172952 |
| *SSBP2* | 81592876 | C/A | intron | c.282+22597G>T | - | rs570146211 |
| *SSBP2* | 81643999 | G/T | intron | c.135+6268C>A | - | - |
| *SSBP2* | 81677829 | G/T | intron | c.63-27490C>A | - | - |
| *SSBP2* | 81712362 | -/GATTTCTGGCTAAGGATGGTAGGTTGAACTTACATTTAC | intron | c.62+38618_62+38619insGTAAATGTAAGTTCAACCTACCATCCTTAGCCAGAAATC | - | rs141328482 |
| *ATG10* | 82005316 | A/C | intron | c.108+17638A>C | - | rs528284827 |
| *ATG10* | 82071052 | C/T | intron | c.216+12450C>T | - | rs538130951 |
| *ATG10* | 82077025 | C/T | intron | c.216+18423C>T | - | rs557106821 |
| *ATG10* | 82106067 | C/A | intron | c.216+47465C>A | - | - |
| *ATG10* | 82127813 | C/T | intron | c.217-36586C>T | - | rs192417385 |
| *ATG10* | 82141850 | -/A | intron | c.217-22537dup | - | rs949240420 |
| *ATG10* | 82220411 | C/T | intron | c.454-32151C>T | - | rs568446204 |
| *RPS23* | 82273305 | -/A | downstream | - | - | - |

| | | | | | | |
|---|---|---|---|---|---|---|
| *ATP6AP1L* | 82312089 | C/T | intron | c.132-522C>T | - | rs562662347 |
| *ATP6AP1L* | 82312089 | C/T | 3' UTR | c.*225C>T | - | rs562662347 |
| *TMEM167A* | 83056822 | G/A | 3' UTR | c.*262C>T | - | rs569130199 |
| *TMEM167A* | 83068259 | A/C | intron | c.4-3142T>G | - | rs559264486 |
| *XRCC4* | 83209091 | GTGTGTGT/- | intron | c.745+4194_745+4201del | - | rs35019637 |
| *XRCC4* | 83267480 | A/T | intron | c.893+8803A>T | - | - |
| *XRCC4* | 83300270 | G/A | intron | c.893+41593G>A | - | rs558960212 |
| *VCAN* | 83479244 | A/C | intron | c.-6-4269A>C | - | - |
| *VCAN* | 83479245 | A/C | intron | c.-6-4268A>C | - | - |
| *VCAN* | 83479246 | A/C | intron | c.-6-4267A>C | - | - |
| *VCAN* | 83509942 | A/G | intron | c.749-2161A>G | - | rs544245229 |
| *VCAN* | 83553242 | A/G | intron | c.9494-122A>G | - | rs544957308 |
| *VCAN* | 83572389 | A/G | intron | c.9736-27A>G | - | rs541600152 |
| *HAPLN1* | 83651679 | G/C | intron | c.472+774C>G | - | rs113959569 |
| *HAPLN1* | 83651679 | G/C | downstream | - | - | rs113959569 |
| *HAPLN1* | 83661579 | G/T | intron | c.101-8755C>A | - | rs570810208 |
| *HAPLN1* | 83719505 | C/A | intron | c.-27+1284G>T | - | - |
| *EDIL3* | 83947867 | G/A | intron | c.1294-4299C>T | - | rs184311438 |
| *EDIL3* | 83965484 | T/C | intron | c.1138-2124A>G | - | rs560824443 |
| *EDIL3* | 84013300 | T/G | intron | c.1137+47000A>C | - | rs571134136 |
| *EDIL3* | 84047772 | C/G | intron | c.1137+12528G>C | - | rs1044847411 |
| *EDIL3* | 84149607 | A/T | intron | c.356-12253T>A | - | rs1159018142 |
| *EDIL3* | 84238693 | T/C | intron | c.197-8809A>G | - | - |
| *EDIL3* | 84293780 | T/- | intron | c.68-39568del | - | rs1044967360 |
| *EDIL3* | 84341836 | C/A | intron | c.67+42472G>T | - | - |

## A 4.2 Primer pairs for the PCR- and Sanger-based sequencing experiments

| ***GLH5-WGS variant primers*** | |
|---|---|
| **Name** | **Sequence (5'-3')** |
| TRAPPC13-3'UTR-F | gctattctattttggtaagttg |
| TRAPPC13-3'UTR-R | agatcctttctaagatggcatt |
| CD180-3'UTR-F | cactacccacccccactaca |
| CD180-3'UTR-R | cctgagaaccaaaccacgat |
| PTDC2-3'UTRa-F | tgataggcatttcattcttt |
| PTDC2-3'UTRa-R | aaaaacaatctacaaaagcaca |
| PTDC2-3'UTRb-F | ttcccagccttcagaactgt |

| ***GLH5-WGS variant primers*** | |
|---|---|
| **Name** | **Sequence (5'-3')** |
| PTDC2-3'UTRb-R | tttcacaaccaagtccattctt |
| HOMER1-3'UTR-F | caactaacgaaggcagctgag |
| HOMER1-3'UTR-R | ttcactgagaagagcctattga |
| RASGRF2-5'UTR-F | ttgccctttaaatcggtctc |
| RASGRF2-5'UTR-R | gttgtagcgcacgctcttct |
| TMEM167A-3'UTR-F | ctcttttatctccgaacattga |
| TMEM167A-3'UTR-R | ccagaggaaaagcaagcact |

| ***TMEM171 gene sequencing primers*** | |
|---|---|
| **Name** | **Sequence (5'-3')** |
| TMEM171-Exon-2a-F | gcctcattgtggtgtgatttc |
| TMEM171-Exon-2a-R | cagttggagccacatccag |
| TMEM171-Exon-2b-F | tttgcccagtgccttatc |
| TMEM171-Exon-2b-R | ccagacgagaaccaagagc |
| TMEM171-Exon-3-F | gcttccatttattacccacctg |

| ***TMEM171 cDNA primers*** | |
|---|---|
| **Name** | **Sequence (5'-3')** |
| TMEM171-Insert-1F | ggacagacacgtcagcaaac |
| TMEM171-Insert-1R | cggctctctccacagatgaa |
| TMEM171-Insert-2F | gcccagtgccttatctttgg |
| TMEM171-Insert-2R | tagttccaggactctcagcg |
| TMEM171-Insert-3F | cccctccaccaccttacttt |

| | |
|---|---|
| TMEM171-Exon-3-R | tcacacttgaatgcctcctg |
| TMEM171-Exon-4-F | gttttcatcccctccccttc |
| TMEM171-Exon-4-R | gaagcactaatggctggaatc |

| | |
|---|---|
| TMEM171-Insert-3R | gtggggaaggctcagaagat |

## A 4.3 Primer pairs used for cloning and site-directed mutagenesis reactions (SDM)

| *TMEM171 cloning primers* | |
|---|---|
| **Name** | **Sequence (5'-3')** |
| **pcDNA3.1-TMEM171** | |
| TMEM171-clone-F | gctggctagcgccaccatgtctcctgcagctgctgct |
| TMEM171-clone-R | tagactcgagttacggtggggaaggctcaga |
| **3X-FlAG-TMEM171** | |
| Tmem171-3x-flag-NotI-FP | gctggcggccgccatgtctcctgcagctgc |

| *TMEM171 interactor cloning primers* | |
|---|---|
| **Name** | **Sequence (5'-3')** |
| **EGFPN1 – SNAP25** | |
| SNAP25-EGFPN1-XhoI-FP | gctgctcgaggccaccatggccgaagacgcagac |
| SNAP25-EGFPN1-SacII-RP | tctagaccgcggaccacttcccagcatc |
| **3X-FlAG-TMEM171** | |
| Tmem171-3x-flag-BamHI-RP | tctagaggatccttacggtggggaaggctcag |

| *TMEM171 SDM primers* | |
|---|---|
| **Name** | **Sequence (5'-3')** |
| TMEM171-Q90R-F | gctccgtgcagggctgcggagaggtcagcagatgg |
| TMEM171-Q90R-R | ccatctgctgacctctccgcagccctgcacggagc |
| TMEM171-R156W-F | ctggcgactcagagccctggatgtgtgggttcctt |
| TMEM171-R156W-R | aaggaacccacacatccagggctctgagtcgccag |
| TMEM171-G159A-F | agagccccggatgtgtgcgttcctttctctgcaga |
| TMEM171-G159A-R | tctgcagagaaaggaacgcacacatccggggctct |
| TMEM171-C67T-SDM-F | cacgtcagcaaactcatcttctgcttctttgtcttcg |
| TMEM171-C67T-SDM-R | cgaagacaaagaagcagaagatgagtttgctgacgtg |

| *TMEM171 SDM primers* | |
|---|---|
| **Name** | **Sequence (5'-3')** |
| TMEM171-46X-SDM-F | atctttgggttccaggcataacaatataagcccctcccag |
| TMEM171-46X-SDM-R | ctgggaggggcttatattgttatgcctggaacccaaagat |
| TMEM171-81X-SDM-F | cgctcccgggcgtaacttcagctcc |
| TMEM171-81X-SDM-R | ggagctgaagttacgcccgggagcg |
| TMEM171-138X-SDM-F | ctggatgtggctcctaatgggcgcaggaacc |
| TMEM171-138X-SDM-R | ggttcctgcgcccattaggagccacatccag |
| TMEM171-184X-SDM-F | cgtggttgcccatgttaagtaaagaaacacgctgaatgctg |
| TMEM171-184X-SDM-R | cagcattcagcgtgtttctttacttaacatgggcaaccacg |

## A 4.4 *TMEM171* rare pathogenic variants identified in the Epi25 consortium study

| Chr position | cDNA position | Amino acid position | Consequence | Allele count Case | Allele number Case | MAF gnomAD v2 |
|---|---|---|---|---|---|---|
| 72419280 | c.80T>C | p.Val27Ala | missense | 1 | 18340 | 0.0004 |
| 72419319 | c.119C>T | p.Ser40Phe | missense | 2 | 18340 | 0.00005 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 72419379 | c.179A>G | p.Lys60Arg | missense | 3 | 18340 | 0.00007 |
| 72419385 | c.185C>A | p.Ala62Glu | missense | 1 | 18340 | _ |
| 72419433 | c.233G>A | p.Arg78His | missense | 1 | 18340 | 0.00006 |
| 72419474 | c.274G>A | p.Gly92Ser | missense | 1 | 18340 | 0.000007 |
| 72419523 | c.323A>G | p.Gln108Arg | missense | 1 | 18340 | _ |
| 72419528 | c.328G>T | p.Ala110Ser | missense | 1 | 18340 | _ |
| 72419567 | c.367G>A | p.Gly123Ser | missense | 1 | 18340 | 0.0001 |
| 72419574 | c.374T>C | p.Leu125Pro | missense | 1 | 18340 | _ |
| 72419612 | c.414delC | p.Asn139ThrfsTer7 | loss of function | 1 | 18340 | 0.00001 |
| 72419660 | c.460G>C | p.Glu154Gln | missense | 1 | 18340 | _ |
| 72419709 | c.509T>C | p.Ile170Thr | missense | 2 | 18340 | 0.00005 |
| 72419744 | c.544C>T | p.His182Tyr | missense | 1 | 18338 | 0.000003 |
| 72419763 | c.563C>T | p.Thr188Met | missense | 2 | 18338 | 0.00004 |
| 72424209 | c.641-8T>C | _ | splice region | 1 | 18338 | _ |
| 72424210 | c.641-5dupC | – | splice region | 1 | 18336 | 0.00002 |
| 72424211 | c.641-6C>T | _ | splice region | 1 | 18338 | 0.000007 |
| 72424239 | c.668delC | p.Pro223LeufsTer123 | loss of function | 4 | 18338 | 0.0001 |
| 72424243 | c.667C>T | p.Pro223Ser | missense | 2 | 18340 | 0.000003 |
| 72424325 | c.755delC | p.Pro252LeufsTer94 | loss of function | 1 | 18336 | 0.0005 |
| 72427421 | c.839C>A | p.Thr280Asn | missense | 1 | 18320 | 0.00005 |
| 72427477 | c.895G>T | p.Glu299Ter | loss of function | 2 | 18332 | 0.00003 |

## A 4.5 TMEM171protein sequence (NCBI, FASTA format)

transmembrane protein 171 isoform 1 [Homo sapiens]
NCBI Reference Sequence: NP_775761.4

```
>gi|239735596|ref|NP_775761.4| transmembrane protein 171 isoform 1 [Homo
sapiens]
MSPAAAAEPDGDQQDRHVSKLIFCFFVFGAVLLCVGVLLSIFGFQACQYKPLPDCPMVLKVAGPACAVVG
LGAVILARSRAQLQLRAGLQRGQQMDPDRAFICGESRQFAQCLIFGFLFLTSGMLISVLGIWVPGCGSNW
AQEPLNETDTGDSEPRMCGFLSLQIMGPLIVLVGLCFFVVAHVKKRNTLNAGQDASEREEGQIQIMEPVQ
VTVGDSVIIFPPPPPPYFPESSASAVAESPGTNSLLPNENPPSYYSIFNYGRTPTSEGAASERDCESIYT
ISGTNSSSEASHTPHLPSELPPRYEEKENAAATFLPLSSEPSPP
```

transmembrane protein 171 [Mus musculus]
NCBI Reference Sequence: NP_001020777.1

```
>gi|71037373|ref|NP_001020777.1| transmembrane protein 171 [Mus musculus]
MSSVGTAEPDGDQRDRHVSKLIFFLFVFGAALLCVGVLLSIFGYQACQYKPLSHCSIVLKIAGPSCAVVG
LGAVILARSRARLHLRERQRQGLQDPDQSFFCGESRQFAQCLIFGFLFLTSGMLISILGIWVPGCDSDWA
QEPLNETNTGEGEPQICGFLSLQIMGPLVVLVGLCFFVVAHVKKKNNLSSSRDTSEVEGGHAHSTEPVHI
```

TVGDSVIIFPPPPPPYFPESSAAAPSPGANSLHQIENPPSYSSLFNYGTPTPENQGAASEREQELIYTIS
GQGSSSERSYTGHLPLDLPPRYEEKETAPATPLGAPSDASPP

transmembrane protein 171 [Rattus norvegicus]
NCBI Reference Sequence: NP_001013925.1
>gi|62079273|ref|NP_001013925.1| transmembrane protein 171 [Rattus norvegicus]
MSSVGTAEPDGDQRDRQVSKLIFFLFVFGAVLLCVGVLISIFGYQACQYKPLSHCSMVLKIAGPSCAVMG
LGTVILARSRARLQLRERQRQGHQDPDQSFFCGESRQFAQCLIFGFLFLTSGMLISILGIWVPGCGSDWA
QEPLNETNSGEGEPQICSFLSLQIMGPLIVLVGLCFFVVAHVKKKSNLSSSRDTSEIEGGHTHSTEPVHI
TVGDSVIIFPPPPPPYFAESSAAAPSPGANSLHRIENPPSYSSLFNLSRTPTPENQGAASERDREVIYTI
SGPGSSSESSHTGHLPLDLPPRYEEKETAPATPLGAPSESSPP

transmembrane protein 171 [Bos taurus]
NCBI Reference Sequence: NP_001014951.2
>gi|75832108|ref|NP_001014951.2| transmembrane protein 171 [Bos taurus]
MSPAAAAEPDGVRGDRHVSKLIFFLFVIGAILLCVGVLLSIFGFQACQYKTFPDCSMMLKIAGPACAVVG
LGAVILARSRARLQQTEERLRGNNQADSDRPFLCGESRQFVQCLIFGFLFLTSGMLISVLGIWVPGCGSD
WMQESLNETDTADSEPQICGFLSLQILGPLIVLVGLCFFVVAHVKKRSNLNGDQDASESEERQTQSLEPI
QVTVGDAVIIFPPPPPPYFPESSASAAARSPGTDGLLPDESPPSYYSIFQSGSPTPEGQGAASERDCELI
YTISGTASSSETSHTLHLLSELPPRYEEKETATTTSLSPSSEPSLP

PREDICTED: transmembrane protein 171 [Pan troglodytes]
NCBI Reference Sequence: XP_001152753.1
>gi|114599719|ref|XP_001152753.1| PREDICTED: transmembrane protein 171 [Pan troglodytes]
MSPAAAAEPDGDQQDRHVSKLIFCFFVFGAVLLCVGVLLSIFGFQACQYKPLPDCPMVLKVAGPACAVVG
LGAVILARSRAQLQLRAGLQRGQQMDPDRAFICGESRQFAQCLIFGFLFLTSGMLISVLGIWVPGCGSNW
AQEPLNETDTGDSEPRMCGFLSLQIMGPLIVLVGLCFFVVAHVKKRNTLNAGQDASEREEGQIQIMEPVQ
VTVGDSVIIFPPPPPPYFPESSASAVAESPGTNSLLPNENPPSYYSIFNYGTPTSEGAASERDCESIYTI
SGMNSSSEASHTPHLPSELPPRYEEKENAAATFLPLSSEPSPP

PREDICTED: transmembrane protein 171 [Macaca mulatta]
NCBI Reference Sequence: XP_001099752.2
>gi|297294522|ref|XP_001099752.2| PREDICTED: transmembrane protein 171 [Macaca mulatta]
MRRRTPRPASAHSRAVSGDAGPSLRAAEILEGRKHIPHCLQEGALLDMSPAAAAEPDGDQQDRHVSKLIF
CFFVFGAVLLCVGVLLSIFGFQACQYKPLPDCPIVLKVAGPACAVVGLGAVILARSRAQLQLRAGLQRGR
QMDPDRAFICGESRQFAQCLIFGFLFLTSGMLISVLGIWVPGCGSDWAREPLNETDTGDSKPRMCGFLSL
QIMGPLIVLVGLCFFVIAHIKKRNTLNIGQDASEREEGQIQSMEAVQVTVGDSVMIFPPPPPPYFPESSA
SAVTESPGTNSLLPNESPPSYYSIFNHGTPASEGVASERDCESIYTISRTNSSSEVSHTPHLPSELPPRY
EEKENAAATFLPLSSEPSPL

transmembrane protein 171 [Sus scrofa]
NCBI Reference Sequence: NP_001231252.1
>gi|346716214|ref|NP_001231252.1| transmembrane protein 171 [Sus scrofa]

MSPVAAAESDGVHRDRHVSKLIFFLFVMGAILLCVGVLLSIFGFQACQYETLPDCSMVLKIAGPACAVVG
LGAVILARSRARLQQSEGRLRGNQGDSDRAFLCGESRQFVQCLIFGFLFLTSGMLISVLGIWVPGCGSDW
AQEPLNETDTADSESQICGFLSLQILGPLIVLVGLCFFVVAHVKKRNNLNGGQDASESEERQPQHMEPVQ
VTVGDAVIIFPPPPPPYFPETSAAAASQGPGADDLLPNESPPSYYSIFHYGTPPPEGQGVTSERDCDSIY
TISGTAPTFETSHTPSLSSELPPRYEEKETAATTSLSPSFEPSPP

PREDICTED: transmembrane protein 171 [Gorilla gorilla gorilla]
NCBI Reference Sequence: XP_004058731.1
>gi|426384348|ref|XP_004058731.1| PREDICTED: transmembrane protein 171
[Gorilla gorilla gorilla]
MSPAAAAEPDGDQQDGHVSKLIFCFFVFGAVLLCVGVLLSIFGFQACQYKPLPDCPMVLKVAGPACAVVG
LGAVILARSRAQLQLRAGLQRGQQMDPDQAFICGESRQFAQCLIFGFLFLTSGMLISVLGIWVPGCGSNW
VQEPLNETDTGDSEPRMCGFLSLQIMGPLIVLVGLCFFVVAHVKKRNTLNAGQDASEREEGQIQIMEPVQ
VTVGDSVIIFPPPPPPYFPESSASVVAESPGTNSLLPNENPPSYYSIFNYGRTPTSEGAASERDCESIYT
ISGTNSSSEASHTPHLPSELPPRYEEKENAAATFLPLSSEPSPP

PREDICTED: transmembrane protein 171 [Ovis aries]
NCBI Reference Sequence: XP_004016942.1
>gi|426246319|ref|XP_004016942.1| PREDICTED: transmembrane protein 171 [Ovis
aries]
MSPAAAAEPDGVQGDRHVSKLIFFLFVIGAILLCVGVLLSIFGFQACQYETFPDCSMVLKIAGPACAVVG
LGAVILARSRARLQQTEERLRGNNQRDSDRPFLCGESRQFVQCLIFGFLFLTSGMLISVLGIWVPGCSSN
WMQEPLNETDTADSEPQICGFLSLQILGPLIVLVGLCFFVVAHVKKRNDLNGAQDASESEERQTQSLEPI
QVTVGDAVIIFPPPPPPYFPESSASAAARSPGTDGLLPDESPPSYYSIFHSGSQTPEGQGAASERDCESI
YTISGTASSSEPSHTLHLLSELPPRYEEKETAATTSLSPSSEPSPP

PREDICTED: transmembrane protein 171 [Felis catus]
NCBI Reference Sequence: XP_003981120.1
>gi|410948804|ref|XP_003981120.1| PREDICTED: transmembrane protein 171
[Felis catus]
MSPVAAAEPDGVQPDRSVSKVIFFLFVFGAILLCVGVLLSIFGFQTCQYKTFPDCSMVLKVAGPACAAVG
LGAVILARSRARLQLREGHLRGSQADPDRAFLCGESRQFAQCLIFGFLFLTSGMLISILGIWVPGCGSDW
AQEPLNETDTADVEPQICGFLSLQIMGPLIVLMGLCFFVVAHVKKRNNLNVGQEASEAEDRQIQNTEPVQ
VTVGDAVIIFPPPPPPYFPESSASAVTRSPGAHGLLPNENPPSYYSIFNYGRTPTPGGQGVGSERDCESI
YTISGSPSSAGISYTPHLSSELPPRYEEKEIAATTSLSPSSEPSRP

PREDICTED: transmembrane protein 171 [Saimiri boliviensisboliviensis]
NCBI Reference Sequence: XP_003925806.1
>gi|403267366|ref|XP_003925806.1| PREDICTED: transmembrane protein 171
[Saimiri boliviensisboliviensis]
MSPAAAAEPDGDQQDRHVSKLIFCFFVFGAVLLCVGVLLSIFGFQACQYKPLPDCSMVLKVAGPACAVIG
LGAVILARSRAQLQLHTGLQRGHQMDPDRAFMCGESRQFAQCLIFGFLFLTSGMLISVLGIWVPGCGSDW
AQESLNETDTADTEPRICGFLSLQIMGPLIVLVGLCFFVVAHIKKRNTNAGQDASESEEGQIQSTEPVRV
TVGDSVIIFPPPPPPYFPESSASAVTESPGTNSQLPNENPPSYYSVFNYGTPTSESQGAASERDCASIYT
ISGTNSSEVSRTPHLPSELPPRYEEKENAAATFLPPSSEPFPP

PREDICTED: transmembrane protein 171 [Otolemurgarnettii]
NCBI Reference Sequence: XP_003785951.1
>gi|395825464|ref|XP_003785951.1| PREDICTED: transmembrane protein 171 [Otolemurgarnettii]
MSPTAAAEPDGEQQDRLVSKLIFFFFVFGAILLCVGVLLSIFGFQACQYESLPHCSMVLKVAGPACAMVG
LGAVILARSRARLRFHQGRRQGHQADPDQAFICGESRQFAQCLVFGFLFLTSGMLISVLGVWVPGCGSGW
EQEPLNDTDSADEEPQICGFLSLQIMGPLIVLVGLCFFVVAHVKKRNELNADQDAPDREERQTQSTEPVQ
VTVGDSVIVFPPPPPPYFSESSAAAVTRSPGANGLLPNENPPPYYSIFNYGTPTPESQGTVSERNRESVY
TISGTSPSFEISHTPHLSSESPPRYEEKDNTAGTPLSPSSVPSPP

PREDICTED: transmembrane protein 171 [Papioanubis]
NCBI Reference Sequence: XP_003899842.1
>gi|402871812|ref|XP_003899842.1| PREDICTED: transmembrane protein 171 [Papioanubis]
MRRRTPRPASAHSSAVSGDAGPSLRAAEILEGRKHIPHCLQEGALLDMSPEAAAEPDGDQQDRHVSKLIF
CFFVFGAVLLCVGVLLSIFGFQACQYKPLPDCPMVLKVAGPACAVVGLGAVILARSRAQLQLRAGLQRGR
QMDPDRAFICGESRQFAQCLIFGFLFLTSGMLISVLGIWVPGCGSDWAREPLNETDTGDSKPRMCGFLSL
QIMGPLIVLVGLCFFVIAHIKKRNPLNAGQDASEREEGQIQSMEAVQVTVGDSVIIFPPPPPPYFPESSA
SAVTESPGTNSLLPNESPPSYYSIFNYGTPASEGVASERDCESIYTISRTNSSSEVSHTPHLPSELPPRY
EEKENAAATFLPLSSEPSPL

PREDICTED: transmembrane protein 171 [Pan paniscus]
NCBI Reference Sequence: XP_003810530.1
>gi|397478388|ref|XP_003810530.1| PREDICTED: transmembrane protein 171 [Pan paniscus]
MSPAAAAEPDGDQQDRHVSKLIFCFFVFGAVLLCVGVLLSIFGFQACQYKPLPDCPMVLKVAGPACAVVG
LGAVILARSRAQLQLRAGLQRGQQMDPDRAFICGESRQFAQCLIFGFLFLTSGMLISVLGIWVPGCGSNW
AQEPLNETDTGDSEPRMCGFLSLQIMGPLIVLVGLCFFVVAHVKKRNTLNAGQDASEREEGQIQIMEPVQ
VTVGDSVIIFPPPPPPYFPESSASAVTESPGTNSLLPNENPPSYYSIFNYGTPTSEGAASERDCESIYTI
SGMNSSSEASHTPHLPSELPPRYEEKENAAATFLPLSSEPSPP

PREDICTED: transmembrane protein 171 [Sarcophilus harrisii]
NCBI Reference Sequence: XP_003759514.1
>gi|395510503|ref|XP_003759514.1| PREDICTED: transmembrane protein 171 [Sarcophilus harrisii]
MSPATAPEPGGDVSGRRISKIIFFLFVFGVILLCAGILLSIFGFQTCHYETFLDCSAVLKIVGPSCAVIG
LGAIILARSRAKLQLRERQLQGNQIDPDSFFLCGESRQFAQFLIFGFLFLTSGMLISVLGIWVPGCNSGW
PQPPFNDTDSSSAEPQACGFLSLQIMGPLIVLIGLCFFVAAHIKKRHNLNSSQDSSENEERRSQLNEPVQ
VTVGDAVIIFPPPPPPYFSDVSTPTVVTRNSSSNSLPLSENPPSYYSIFNYGTQIPEVQGAAFYRDTESI
YTISGNNLSSETSPASHISSELPPRYEEKETVNVDHPTSSTSSSSSSQPSPP

PREDICTED: transmembrane protein 171 [Nomascusleucogenys]
NCBI Reference Sequence: XP_003266102.1
>gi|332233819|ref|XP_003266102.1| PREDICTED: transmembrane protein 171 [Nomascusleucogenys]

MSPAAAAEPDGDQQDRHVGKLIFCFFVFGAVLLCVGVLLSIFGFQACQYKPLPDCPLVLKVAGPACAVVG
LGAVILARSRAQLQLRAGLQRGQQMDPDRAFICGESRQFAQCLIFGFLFLTSGMLISILGIWVPGCGSNW
AQEPLNETDTGDSEPRMCGFLSLQIMGPLIVLVGLCFFVVAHVKKRNTLNAGQDASEREEGQIQIMEPVQ
VTVGDSVIIFPPPPPPYFPESSASAVAESPGTNSLLPNENPPSYYSIFNYGTPTSEGAASERDCESIYTI
SRMNSSSEVSHTPHLPSELPPRYEEKENAAATFLPLSSEPSPP

PREDICTED: transmembrane protein 171 [Callithrix jacchus]
NCBI Reference Sequence: XP_002744920.1
>gi|296194364|ref|XP_002744920.1| PREDICTED: transmembrane protein 171
[Callithrix jacchus]
MSPAAAADPDGDQQDRHISKLIFCFFVFGAVLLCLGLLLSIFGFQACQYKSLPDCSMVLKVPPRLQHVLK
CGAACAVIGLGAVILARSRAQLQLRTGLQRGHQMDSDRAFICGESRQFAQCLIFGFLFLTSGMFISVLGI
WVPGCGSDWAQEPLNETDTADMEPRICGFLSLQIMGPLIVLVGLCFFVVAHIKKRNTLNAGQNASESEEG
QIQSTEPVQITVGDSVIIFPPPPPPYFPDSSASAVTESPGTNSQLPNENPPSYYSIFNYGTPTSENQGAA
SERDCESIYTISRTNSSSEVSHIPHLPSELPPRYEEKENAAATFLPPSSEPFPP

# References

Alfradique I, Vasconcelos MM. 2007. Juvenile myoclonic epilepsy. Arq Neuropsiquiatr 65:1266–1271.

Allen AS, Bellows ST, Berkovic SF, Bridgers J, Burgess R, Cavalleri G, Chung S-K, Cossette P, Delanty N, Dlugos D, Epstein MP, Freyer C, et al. 2017. Ultra-rare genetic variation in common epilepsies: a case-control sequencing study. The Lancet Neurology 16:135–143.

Almeida A, Bolaños JP, Moreno S. 2005. Cdh1/Hct1-APC Is Essential for the Survival of Postmitotic Neurons. J Neurosci 25:8115–8121.

Amenduni M, De Filippis R, Cheung AYL, Disciglio V, Epistolato MC, Ariani F, Mari F, Mencarelli MA, Hayek Y, Renieri A, Ellis J, Meloni I. 2011. iPS cells to model CDKL5-related disorders. Eur J Hum Genet 19:1246–1255.

Antonucci F, Corradini I, Fossati G, Tomasoni R, Menna E, Matteoli M. 2016. SNAP-25, a Known Presynaptic Protein with Emerging Postsynaptic Functions. Front Synaptic Neurosci 8:.

Arango-Lievano M, Boussadia B, De Terdonck LDT, Gault C, Fontanaud P, Lafont C, Mollard P, Marchi N, Jeanneteau F. 2018. Topographic Reorganization of Cerebrovascular Mural Cells under Seizure Conditions. Cell Rep 23:1045–1059.

Arsov T, Mullen SA, Damiano JA, Lawrence KM, Huh LL, Nolan M, Young H, Thouin A, Dahl H-HM, Berkovic SF, Crompton DE, Sadleir LG, et al. 2012. Early onset absence epilepsy: 1 in 10 cases is caused by GLUT1 deficiency. Epilepsia 53:e204-207.

Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, Donnelly P, Eichler EE, Flicek P, Gabriel SB, et al. 2015. A global reference for human genetic variation. Nature 526:68–74.

Bai D, Alonso ME, Medina MT, Bailey JN, Morita R, Cordova S, Rasmussen A, Ramos-Peek J, Ochoa A, Jara A, Donnadieu FR, Cadena G, et al. 2002. Juvenile myoclonic epilepsy: linkage to chromosome 6p12 in Mexico families. Am J Med Genet 113:268–274.

Bailey JN, Nijs L de, Bai D, Suzuki T, Miyamoto H, Tanaka M, Patterson C, Lin Y-C, Medina MT, Alonso ME, Serratosa JM, Durón RM, et al. 2018. Variant Intestinal-Cell Kinase in Juvenile Myoclonic Epilepsy. N Engl J Med 378:1018–1028.

Bailey JN, Patterson C, Nijs L de, Durón RM, Nguyen V-H, Tanaka M, Medina MT, Jara-Prado A, Martínez-Juárez IE, Ochoa A, Molina Y, Suzuki T, et al. 2017. EFHC1 variants in juvenile myoclonic epilepsy: reanalysis according to NHGRI and ACMG guidelines for assigning disease causality. Genet Med 19:144–156.

Bär H, Kostareva A, Sjöberg G, Sejersen T, Katus HA, Herrmann H. 2006. Forced expression of desmin and desmin mutants in cultured cells: impact of myopathic missense mutations in the central coiled-coil domain on network formation. Exp Cell Res 312:1554–1565.

Bär H, Strelkov SV, Sjöberg G, Aebi U, Herrmann H. 2004a. The biology of desmin filaments: how do mutations affect their structure, assembly, and organisation? J Struct Biol 148:137–152.

Bär H, Strelkov SV, Sjöberg G, Aebi U, Herrmann H. 2004b. The biology of desmin filaments: how do mutations affect their structure, assembly, and organisation? Journal of Structural Biology 148:137–152.

Baykan B, Madia F, Bebek N, Gianotti S, Güney AI, Cine N, Bianchi A, Gökyiğit A, Zara F. 2004. Autosomal recessive idiopathic epilepsy in an inbred family from Turkey: identification of a putative locus on chromosome 9q32-33. Epilepsia 45:479–487.

Bell RD, Winkler EA, Sagare AP, Singh I, LaRue B, Deane R, Zlokovic BV. 2010. Pericytes Control Key Neurovascular Functions and Neuronal Phenotype in the Adult Brain and during Brain Aging. Neuron 68:409–427.

Bozzi Y, Casarosa S, Caleo M. 2012. Epilepsy as a Neurodevelopmental Disorder. Front Psychiatry 3:.

Canevini MP, Mai R, Di Marco C, Bertin C, Minotti L, Pontrelli V, Saltarelli A, Canger R. 1992. Juvenile myoclonic epilepsy of Janz: clinical observations in 60 patients. Seizure 1:291–298.

Cao F, Liu JJ, Zhou S, Cortez MA, Snead OC, Han J, Jia Z. 2020. Neuroligin 2 regulates absence seizures and behavioral arrests through GABAergic transmission within the thalamocortical circuitry. Nat Commun 11:3744.

Capetanaki Y, Bloch RJ, Kouloumenta A, Mavroidis M, Psarras S. 2007. Muscle intermediate filaments and their links to membranes and membranous organelles. Experimental Cell Research 313:2063–2076.

Catterall WA, Few AP. 2008. Calcium Channel Regulation and Presynaptic Plasticity. Neuron 59:882–901.

Cavalleri GL, Walley NM, Soranzo N, Mulley J, Doherty CP, Kapoor A, Depondt C, Lynch JM, Scheffer IE, Heils A, Gehrmann A, Kinirons P, et al. 2007. A multicenter study of BRD2 as a risk factor for juvenile myoclonic epilepsy. Epilepsia 48:706–712.

Chen W-J, Lin Y, Xiong Z-Q, Wei W, Ni W, Tan G-H, Guo S-L, He J, Chen Y-F, Zhang Q-J, Li H-F, Lin Y, et al. 2011. Exome sequencing identifies truncating mutations in PRRT2 that cause paroxysmal kinesigenic dyskinesia. Nat Genet 43:1252–1255.

Cheng J, Korte N, Nortley R, Sethi H, Tang Y, Attwell D. 2018. Targeting pericytes for therapeutic approaches to neurological disorders. Acta Neuropathol 136:507–523.

Chinwalla AT, Cook LL, Delehaunty KD, Fewell GA, Fulton LA, Fulton RS, Graves TA, Hillier LW, Mardis ER, McPherson JD, Miner TL, Nash WE, et al. 2002. Initial sequencing and comparative analysis of the mouse genome. Nature 420:520–562.

Clemen CS, Herrmann H, Strelkov SV, Schröder R. 2013. Desminopathies: pathology and mechanisms. Acta Neuropathol 125:47–75.

Cossette P, Liu L, Brisebois K, Dong H, Lortie A, Vanasse M, Saint-Hilaire J-M, Carmant L, Verner A, Lu W-Y, Wang YT, Rouleau GA. 2002. Mutation of GABRA1 in an autosomal dominant form of juvenile myoclonic epilepsy. Nat Genet 31:184–189.

Delgado-Escueta AV. 2007. Advances in Genetics of Juvenile Myoclonic Epilepsies. Epilepsy Curr 7:61–67.

Delgado-Escueta AV, Koeleman BPC, Bailey JN, Medina MT, Durón RM. 2013. The quest for juvenile myoclonic epilepsy genes. Epilepsy Behav 28 Suppl 1:S52-57.

Delplanque J, Devos D, Huin V, Genet A, Sand O, Moreau C, Goizet C, Charles P, Anheim M, Monin ML, Buée L, Destée A, et al. 2014. TMEM240 mutations cause spinocerebellar ataxia 21 with mental retardation and severe cognitive impairment. Brain 137:2657–2663.

Devinsky O, Vezzani A, O'Brien TJ, Jette N, Scheffer IE, Curtis M de, Perucca P. 2018. Epilepsy. Nat Rev Dis Primers 4:1–24.

Di Cunto F, Imarisio S, Hirsch E, Broccoli V, Bulfone A, Migheli A, Atzori C, Turco E, Triolo R, Dotto GP, Silengo L, Altruda F. 2000. Defective Neurogenesis in Citron Kinase Knockout Mice by Altered Cytokinesis and Massive Apoptosis. Neuron 28:115–127.

Dib C, Fauré S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, et al. 1996. A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 380:152–154.

Dibbens LM, Feng H-J, Richards MC, Harkin LA, Hodgson BL, Scott D, Jenkins M, Petrou S, Sutherland GR, Scheffer IE, Berkovic SF, Macdonald RL, et al. 2004. GABRD encoding a protein for extra- or peri-synaptic GABAA receptors is a susceptibility locus for generalized epilepsies. Hum Mol Genet 13:1315–1319.

Dibbens LM, Mullen S, Helbig I, Mefford HC, Bayly MA, Bellows S, Leu C, Trucks H, Obermeier T, Wittig M, Franke A, Caglayan H, et al. 2009. Familial and sporadic 15q13.3 microdeletions in idiopathic generalized epilepsy: precedent for disorders with complex inheritance. Hum Mol Genet 18:3626–3631.

Dimassi S, Labalme A, Lesca G, Rudolf G, Bruneau N, Hirsch E, Arzimanoglou A, Motte J, Martin A de S, Boutry-Kryza N, Cloarec R, Benitto A, et al. 2014. A subset of genomic alterations detected in rolandic epilepsies contains candidate or known epilepsy genes including GRIN2A and PRRT2. Epilepsia 55:370–378.

Durner M, Janz D, Zingsem J, Greenberg DA. 1992. Possible association of juvenile myoclonic epilepsy with HLA-DRw6. Epilepsia 33:814–816.

Durner M, Keddache MA, Tomasini L, Shinnar S, Resor SR, Cohen J, Harden C, Moshe SL, Rosenbaum D, Kang H, Ballaban-Gil K, Hertz S, et al. 2001. Genome scan of idiopathic generalized epilepsy: evidence for major susceptibility gene and modifying genes influencing the seizure type. Ann Neurol 49:328–335.

Duy PQ, David WB, Kahle KT. 2019. Identification of KCC2 Mutations in Human Epilepsy Suggests Strategies for Therapeutic Transporter Modulation. Frontiers in Cellular Neuroscience 13:515.

Eggenschwiler R, Moslem M, Fráguas MS, Galla M, Papp O, Naujock M, Fonfara I, Gensch I, Wähner A, Beh-Pajooh A, Mussolino C, Tauscher M, et al. 2016. Improved bi-allelic modification of a transcriptionally silent locus in patient-derived iPSC by Cas9 nickase. Sci Rep 6:38198.

Elmslie FV, Rees M, Williamson MP, Kerr M, Kjeldsen MJ, Pang KA, Sundqvist A, Friis ML, Chadwick D, Richens A, Covanis A, Santos M, et al. 1997. Genetic mapping of a major

susceptibility locus for juvenile myoclonic epilepsy on chromosome 15q. Hum Mol Genet 6:1329–1334.

Epi4K: Gene discovery in 4,000 genomes. 2012. Epilepsia 53:1457–1467.

Epi25 Collaborative. Electronic address: s.berkovic@unimelb.edu.au, Epi25 Collaborative. 2019. Ultra-Rare Genetic Variation in the Epilepsies: A Whole-Exome Sequencing Study of 17,606 Individuals. Am J Hum Genet 105:267–282.

Escayg A, De Waard M, Lee DD, Bichet D, Wolf P, Mayer T, Johnston J, Baloh R, Sander T, Meisler MH. 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 66:1531–1539.

Escayg A, Heils A, MacDonald BT, Haug K, Sander T, Meisler MH. 2001. A Novel SCN1A Mutation Associated with Generalized Epilepsy with Febrile Seizures Plus—and Prevalence of Variants in Patients with Epilepsy. The American Journal of Human Genetics 68:866–873.

Etain B, Dumaine A, Mathieu F, Chevalier F, Henry C, Kahn J-P, Deshommes J, Bellivier F, Leboyer M, Jamain S. 2010. A SNAP25 promoter variant is associated with early-onset bipolar disorder and a high expression level in brain. Mol Psychiatry 15:748–755.

Ewels P, Magnusson M, Lundin S, Käller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32:3047–3048.

Fang G, Yu H, Kirschner MW. 1998. Direct binding of CDC20 protein family members activates the anaphase-promoting complex in mitosis and G1. Mol Cell 2:163–171.

Feng Y-CA, Howrigan DP, Abbott LE, Tashman K, Cerrato F, Singh T, Heyne H, Byrnes A, Churchhouse C, Watts N, Solomonson M, Lal D, et al. 2019. Ultra-Rare Genetic Variation in the Epilepsies: A Whole-Exome Sequencing Study of 17,606 Individuals. The American Journal of Human Genetics 105:267–282.

Fisher RS, Acevedo C, Arzimanoglou A, Bogacz A, Cross JH, Elger CE, Engel J, Forsgren L, French JA, Glynn M, Hesdorffer DC, Lee BI, et al. 2014. ILAE official report: a practical clinical definition of epilepsy. Epilepsia 55:475–482.

Fitzgerald TW, Gerety SS, Jones WD, Kogelenberg M van, King DA, McRae J, Morley KI, Parthiban V, Al-Turki S, Ambridge K, Barrett DM, Bayzetinova T, et al. 2015. Large-scale discovery of novel genetic causes of developmental disorders. Nature 519:223–228.

Fossati G, Morini R, Corradini I, Antonucci F, Trepte P, Edry E, Sharma V, Papale A, Pozzi D, Defilippi P, Meier JC, Brambilla R, et al. 2015. Reduced SNAP-25 increases PSD-95 mobility and impairs spine morphogenesis. Cell Death Differ 22:1425–1436.

Foyaca-Sibat H. 2011. Novel Aspects on Epilepsy.

Fruscione F, Valente P, Sterlini B, Romei A, Baldassari S, Fadda M, Prestigio C, Giansante G, Sartorelli J, Rossi P, Rubio A, Gambardella A, et al. 2018. PRRT2 controls neuronal excitability by negatively modulating Na+ channel 1.2/1.6 activity. Brain 141:1000–1016.

Garner JP. 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 Journal 55:438–456.

Ghiani CA, Starcevic M, Rodriguez-Fernandez IA, Nazarian R, Cheli VT, Chan LN, Malvar JS, Vellis J de, Sabatti C, Dell'Angelica EC. 2010. The dysbindin-containing complex (BLOC-1) in brain: developmental regulation, interaction with SNARE proteins and role in neurite outgrowth. Molecular Psychiatry 15:204–215.

Goetz SC, Anderson KV. 2010. The Primary Cilium: A Signaling Center During Vertebrate Development. Nat Rev Genet 11:331–344.

Green MR, Sambrook J. 2017. Isolation of High-Molecular-Weight DNA Using Organic Solvents. Cold Spring Harb Protoc 2017:pdb.prot093450.

Greenberg DA, Cayanis E, Strug L, Marathe S, Durner M, Pal DK, Alvin GB, Klotz I, Dicker E, Shinnar S, Bromfield EB, Resor S, et al. 2005. Malic enzyme 2 may underlie susceptibility to adolescent-onset idiopathic generalized epilepsy. Am J Hum Genet 76:139–146.

Gross RA. 1992. A brief history of epilepsy and its therapy in the western hemisphere. Epilepsy Research 12:65–74.

Han S, An Z, Luo X, Zhang L, Zhong X, Du W, Yi Q, Shi Y. 2018. Association between CMYA5 gene polymorphisms and risk of schizophrenia in Uygur population and a meta-analysis. Early Interv Psychiatry 12:15–21.

Harrison V, Connell L, Hayesmoore J, McParland J, Pike MG, Blair E. 2011. Compound heterozygous deletion of NRXN1 causing severe developmental delay with early onset epilepsy in two sisters. Am J Med Genet A 155A:2826–2831.

Hawi Z, Matthews N, Wagner J, Wallace RH, Butler TJ, Vance A, Kent L, Gill M, Bellgrove MA. 2013. DNA Variation in the SNAP25 Gene Confers Risk to ADHD and Is Associated with Reduced Expression in Prefrontal Cortex. PLoS One 8:.

Heinzen EL, Depondt C, Cavalleri GL, Ruzzo EK, Walley NM, Need AC, Ge D, He M, Cirulli ET, Zhao Q, Cronin KD, Gumbs CE, et al. 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 91:293–302.

Helbig I, Mefford HC, Sharp AJ, Guipponi M, Fichera M, Franke A, Muhle H, Kovel C de, Baker C, Spiczak S von, Kron KL, Steinich I, et al. 2009. 15q13.3 microdeletions increase risk of idiopathic generalized epilepsy. Nat Genet 41:160–162.

Hempelmann A, Taylor KP, Heils A, Lorenz S, Prud'homme J-F, Nabbout R, Dulac O, Rudolf G, Zara F, Bianchi A, Robinson R, Gardiner RM, et al. 2006. Exploration of the genetic architecture of idiopathic generalized epilepsies. Epilepsia 47:1682–1690.

Heron SE, Grinton BE, Kivity S, Afawi Z, Zuberi SM, Hughes JN, Pridmore C, Hodgson BL, Iona X, Sadleir LG, Pelekanos J, Herlenius E, et al. 2012. PRRT2 mutations cause benign familial infantile epilepsy and infantile convulsions with choreoathetosis syndrome. Am J Hum Genet 90:152–160.

Heron SE, Khosravani H, Varela D, Bladen C, Williams TC, Newman MR, Scheffer IE, Berkovic SF, Mulley JC, Zamponi GW. 2007. Extended spectrum of idiopathic generalized epilepsies associated with CACNA1H functional variants. Ann Neurol 62:560–568.

Heron SE, Phillips HA, Mulley JC, Mazarib A, Neufeld MY, Berkovic SF, Scheffer IE. 2004. Genetic variation of CACNA1H in idiopathic generalized epilepsy. Ann Neurol 55:595–596.

Higurashi N, Uchida T, Lossin C, Misumi Y, Okada Y, Akamatsu W, Imaizumi Y, Zhang B, Nabeshima K, Mori MX, Katsurabayashi S, Shirasaka Y, et al. 2013. A human Dravet syndrome model from patient induced pluripotent stem cells. Molecular Brain 6:19.

Hinz L, Hoekstra SD, Watanabe K, Posthuma D, Heine VM. 2019. Generation of Isogenic Controls for In Vitro Disease Modelling of X-Chromosomal Disorders. Stem Cell Rev and Rep 15:276–285.

Hodge RD, Bakken TE, Miller JA, Smith KA, Barkan ER, Graybuck LT, Close JL, Long B, Johansen N, Penn O, Yao Z, Eggermont J, et al. 2019. Conserved cell types with divergent features in human versus mouse cortex. Nature 573:61–68.

Homa M, Loyens A, Eddarkaoui S, Faivre E, Deramecourt V, Maurage C-A, Buée L, Huin V, Sablonnière B. 2020. The TMEM240 Protein, Mutated in SCA21, Is Expressed in Purkinje Cells and Synaptic Terminals. Cerebellum 19:358–369.

Hong YJ, Do JT. 2019. Neural Lineage Differentiation From Pluripotent Stem Cells to Mimic Human Brain Tissues. Frontiers in Bioengineering and Biotechnology 7:400.

Hsiung A, Naya FJ, Chen X, Shiang R. 2019. A schizophrenia associated CMYA5 allele displays differential binding with desmin. J Psychiatr Res 111:8–15.

Huang L, Szymanska K, Jensen VL, Janecke AR, Innes AM, Davis EE, Frosk P, Li C, Willer JR, Chodirker BN, Greenberg CR, McLeod DR, et al. 2011. TMEM237 is mutated in individuals with a Joubert syndrome related disorder and expands the role of the TMEM family at the ciliary transition zone. Am J Hum Genet 89:713–730.

Hughes S, Chan-Ling T. 2004. Characterization of Smooth Muscle Cell and Pericyte Differentiation in the Rat Retina In Vivo. Invest Ophthalmol Vis Sci 45:2795–2806.

Ilardi JM, Mochida S, Sheng ZH. 1999. Snapin: a SNARE-associated protein implicated in synaptic transmission. Nat Neurosci 2:119–124.

International League Against Epilepsy Consortium on Complex Epilepsies. 2018. Genome-wide mega-analysis identifies 16 loci and highlights diverse biological mechanisms in the common epilepsies. Nat Commun 9:5269.

Izsak J, Seth H, Andersson M, Vizlin-Hodzic D, Theiss S, Hanse E, Ågren H, Funa K, Illes S. 2019. Robust Generation of Person-Specific, Synchronously Active Neuronal Networks Using Purely Isogenic Human iPSC-3D Neural Aggregate Cultures. Front Neurosci 0:

Jallon P, Latour P. 2005a. Epidemiology of Idiopathic Generalized Epilepsies. Epilepsia 46:10–14.

Jallon P, Latour P. 2005b. Epidemiology of idiopathic generalized epilepsies. Epilepsia 46 Suppl 9:10–14.

Jang C-Y, Coppinger JA, Seki A, Yates JR, Fang G. 2009. Plk1 and Aurora A regulate the depolymerase activity and the cellular localization of Kif2a. J Cell Sci 122:1334–1341.

Janz D, Christian W. 1957. Impulsiv-Petit mal. Deutsche Zeitschrift f Nervenheilkunde 176:346–386.

Jayalakshmi SS, Srinivasa Rao B, Sailaja S. 2010. Focal clinical and electroencephalographic features in patients with juvenile myoclonic epilepsy. Acta Neurol Scand 122:115–123.

Kahle KT, Merner ND, Friedel P, Silayeva L, Liang B, Khanna A, Shang Y, Lachance-Touchette P, Bourassa C, Levert A, Dion PA, Walcott B, et al. 2014. Genetically encoded impairment of neuronal KCC2 cotransporter function in human idiopathic generalized epilepsy. EMBO Rep 15:766–774.

Kapoor A, Ratnapriya R, Kuruttukulam G, Anand A. 2007. A novel genetic locus for juvenile myoclonic epilepsy at chromosome 5q12-q14. Hum Genet 121:655–662.

Kapoor A, Satishchandra P, Ratnapriya R, Reddy R, Kadandale J, Shankar SK, Anand A. 2008. An idiopathic epilepsy syndrome linked to 3q13.3-q21 and missense mutations in the extracellular calcium sensing receptor gene. Ann Neurol 64:158–167.

Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, et al. 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. bioRxiv 531210.

Kearney JA. 2014. Epi4K Phase I: Gene Discovery in Epileptic Encephalopathies by Exome Sequencing: Epileptic Encephalopathies Exome Sequencing. Epilepsy Curr 14:208–210.

Kibbe WA. 2007. OligoCalc: an online oligonucleotide properties calculator. Nucleic Acids Res 35:W43-46.

Kim AH, Puram SV, Bilimoria PM, Ikeuchi Y, Keough S, Wong M, Rowitch D, Bonni A. 2009. A centrosomal Cdc20-APC pathway controls dendrite morphogenesis in postmitotic neurons. Cell 136:322–336.

Kinirons P, Verlaan DJ, Dubé M-P, Poirier J, Deacon C, Lortie A, Clément J-F, Desbiens R, Carmant L, Cieuta-Walti C, Shevell M, Rouleau GA, et al. 2008. A novel locus for idiopathic generalized epilepsy in French-Canadian families maps to 10p11. Am J Med Genet A 146A:578–584.

Kogut I, McCarthy SM, Pavlova M, Astling DP, Chen X, Jakimenko A, Jones KL, Getahun A, Cambier JC, Pasmooij AMG, Jonkman MF, Roop DR, et al. 2018. High-efficiency RNA-based reprogramming of human primary fibroblasts. Nat Commun 9:745.

Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, et al. 2002. A high-resolution recombination map of the human genome. Nat Genet 31:241–247.

Konishi Y, Stegmüller J, Matsuda T, Bonni S, Bonni A. 2004. Cdh1-APC controls axonal growth and patterning in the mammalian brain. Science 303:1026–1030.

Kovel CGF de, Pinto D, Haan GJ de, Kasteleijn-Nolst Trenité DG, Lindhout D, Koeleman BPC. 2007. Association analysis of BRD2 (RING3) and epilepsy in a Dutch population. Epilepsia 48:2191–2192.

Kovel CGF de, Trucks H, Helbig I, Mefford HC, Baker C, Leu C, Kluck C, Muhle H, Spiczak S von, Ostertag P, Obermeier T, Kleefuss-Lie AA, et al. 2010. Recurrent microdeletions at 15q11.2 and 16p13.11 predispose to idiopathic generalized epilepsies. Brain 133:23–32.

Krepischi ACV, Knijnenburg J, Bertola DR, Kim CA, Pearson PL, Bijlsma E, Szuhai K, Kok F, Vianna-Morgante AM, Rosenberg C. 2010. Two distinct regions in 2q24.2-q24.3 associated with idiopathic epilepsy. Epilepsia 51:2457–2460.

Kuijlaars J, Oyelami T, Diels A, Rohrbacher J, Versweyveld S, Meneghello G, Tuefferd M, Verstraelen P, Detrez JR, Verschuuren M, De Vos WH, Meert T, et al. 2016. Sustained synchronized neuronal network activity in a human astrocyte co-culture system. Sci Rep 6:36529.

Kyttälä A, Moraghebi R, Valensisi C, Kettunen J, Andrus C, Pasumarthy KK, Nakanishi M, Nishimura K, Ohtaka M, Weltner J, Van Handel B, Parkkonen O, et al. 2016. Genetic Variability Overrides the Impact of Parental Cell Type and Determines iPSC Differentiation Potential. Stem Cell Reports 6:200–212.

Lachance-Touchette P, Brown P, Meloche C, Kinirons P, Lapointe L, Lacasse H, Lortie A, Carmant L, Bedford F, Bowie D, Cossette P. 2011. Novel α1 and γ2 GABAA receptor subunit mutations in families with idiopathic generalized epilepsy. Eur J Neurosci 34:237–249.

Laemmli UK. 1970. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. Nature 227:680–685.

Lathrop GM, Lalouel JM, Julier C, Ott J. 1984. Strategies for multilocus linkage analysis in humans. Proc Natl Acad Sci USA 81:3443–3446.

Lau CG, Takayasu Y, Rodenas-Ruano A, Paternain AV, Lerma J, Bennett MVL, Zukin RS. 2010. SNAP-25 Is a Target of Protein Kinase C Phosphorylation Critical to NMDA Receptor Trafficking. J Neurosci 30:242–254.

Layouni S, Buresi C, Thomas P, Malafosse A, Dogui M. 2010. BRD2 and TAP-1 genes and juvenile myoclonic epilepsy. Neurol Sci 31:53–56.

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 40:117–119.

Leal-Campanario R, Alarcon-Martinez L, Rieiro H, Martinez-Conde S, Alarcon-Martinez T, Zhao X, LaMee J, Popp PJO, Calhoun ME, Arribas JI, Schlegel AA, Stasi LLD, et al. 2017. Abnormal Capillary Vasodynamics Contribute to Ictal Neurodegeneration in Epilepsy. Sci Rep 7:43276.

Lee H-Y, Huang Y, Bruneau N, Roll P, Roberson EDO, Hermann M, Quinn E, Maas J, Edwards R, Ashizawa T, Baykan B, Bhatia K, et al. 2012a. Mutations in the novel protein PRRT2 cause paroxysmal kinesigenic dyskinesia with infantile convulsions. Cell Rep 1:2–12.

Lee Y-C, Lee M-J, Yu H-Y, Chen C, Hsu C-H, Lin K-P, Liao K-K, Chang M-H, Liao Y-C, Soong B-W. 2012b. PRRT2 mutations in paroxysmal kinesigenic dyskinesia with infantile convulsions in a Taiwanese cohort. PLoS One 7:e38543.

Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, et al. 2016. Analysis of protein-coding genetic variation in 60,706 humans. Nature 536:285–291.

Lemke JR, Riesch E, Scheurenbrand T, Schubach M, Wilhelm C, Steiner I, Hansen J, Courage C, Gallati S, Bürki S, Strozzi S, Simonetti BG, et al. 2012. Targeted next generation sequencing as a diagnostic tool in epileptic disorders. Epilepsia 53:1387–1398.

Lenzen KP, Heils A, Lorenz S, Hempelmann A, Höfels S, Lohoff FW, Schmitz B, Sander T. 2005a. Supportive evidence for an allelic association of the human KCNJ10 potassium channel gene with idiopathic generalized epilepsy. Epilepsy Res 63:113–118.

Lenzen KP, Heils A, Lorenz S, Hempelmann A, Sander T. 2005b. Association analysis of malic enzyme 2 gene polymorphisms with idiopathic generalized epilepsy. Epilepsia 46:1637–1641.

Leu C, Kovel CGF de, Zara F, Striano P, Pezzella M, Robbiano A, Bianchi A, Bisulli F, Coppola A, Giallonardo AT, Beccaria F, Trenité DK-N, et al. 2012. Genome-wide linkage meta-analysis identifies susceptibility loci at 2q34 and 13q31.3 for genetic generalized epilepsies. Epilepsia 53:308–318.

Li H, Durbin R. 2009a. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760.

Li H, Durbin R. 2009b. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

Li M, Niu F, Zhu X, Wu X, Shen N, Peng X, Liu Y. 2015. PRRT2 Mutant Leads to Dysfunction of Glutamate Signaling. International Journal of Molecular Sciences 16:9134–9151.

Li X, Poschmann S, Chen Q, Fazeli W, Oundjian NJ, Snoeijen-Schouwenaars FM, Fricke O, Kamsteeg E-J, Willemsen M, Wang QK. 2018. De novo BK channel variant causes epilepsy by affecting voltage gating but not Ca2+ sensitivity. Eur J Hum Genet 26:220–229.

Liu AW, Delgado-Escueta AV, Gee MN, Serratosa JM, Zhang QW, Alonso ME, Medina MT, Cordova S, Zhao HZ, Spellman JM, Donnadieu FR, Peek JR, et al. 1996. Juvenile myoclonic epilepsy in chromosome 6p12-p11: locus heterogeneity and recombinations. Am J Med Genet 63:438–446.

Liu AW, Delgado-Escueta AV, Serratosa JM, Alonso ME, Medina MT, Gee MN, Cordova S, Zhao HZ, Spellman JM, Peek JR. 1995. Juvenile myoclonic epilepsy locus in chromosome 6p21.2-p11: linkage to convulsions and electroencephalography trait. Am J Hum Genet 57:368–381.

Liu S-P, Hsu Y-H, Huang C-Y, Ho M-C, Cheng Y-C, Wen C-H, Lu H-E, Tsai C-H, Shyu W-C, Hsieh PCH. 2018. Generation of novel induced pluripotent stem cell (iPSC) line from a 16-year-old sialidosis patient with NEU-1 gene mutation. Stem Cell Research 28:39–43.

Liu X, Jian X, Boerwinkle E. 2011. dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. Hum Mutat 32:894–899.

Loucks CM, Park K, Walker DS, McEwan AH, Timbers TA, Ardiel EL, Grundy LJ, Li C, Johnson J-L, Kennedy J, Blacque OE, Schafer W, et al. 2019. EFHC1, implicated in juvenile myoclonic epilepsy, functions at the cilium and synapse to modulate dopamine signaling. eLife 8:e37271.

Lowther C, Costain G, Stavropoulos DJ, Melvin R, Silversides CK, Andrade DM, So J, Faghfoury H, Lionel AC, Marshall CR, Scherer SW, Bassett AS. 2015. Delineating the 15q13.3 microdeletion phenotype: a case series and comprehensive review of the literature. Genet Med 17:149–157.

Lu H, Anujan P, Zhou F, Zhang Y, Chong YL, Bingle CD, Roy S. 2019. *Mcidas* mutant mice reveal a two-step process for the specification and differentiation of multiciliated cells in mammals. Development 146:dev172643.

Ma S, Blair MA, Abou-Khalil B, Lagrange AH, Gurnett CA, Hedera P. 2006. Mutations in the GABRA1 and EFHC1 genes are rare in familial juvenile myoclonic epilepsy. Epilepsy Res 71:129–134.

Madaule P, Eda M, Watanabe N, Fujisawa K, Matsuoka T, Bito H, Ishizaki T, Narumiya S. 1998. Role of citron kinase as a target of the small GTPase Rho in cytokinesis. Nature 394:491–494.

Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res 47:W636–W641.

Majolo F, Marinowic DR, Palmini ALF, DaCosta JC, Machado DC. 2019. Migration and Synaptic Aspects of Neurons Derived from Human Induced Pluripotent Stem Cells from Patients with Focal Cortical Dysplasia II. Neuroscience 408:81–90.

Manivannan SN, Roovers J, Smal N, Myers CT, Turkdogan D, Roelens F, Kanca O, Chung H-L, Scholz T, Hermann K, Bierhals T, Caglayan SH, et al. 2021. De novo FZR1 loss-of-function variants cause developmental and epileptic encephalopathies including Myoclonic Atonic Epilepsy. medRxiv 2021.06.12.21256778.

Marcet B, Chevalier B, Coraux C, Kodjabachian L, Barbry P. 2011. MicroRNA-based silencing of Delta/Notch signaling promotes multiple cilia formation. Cell Cycle 10:2858–2864.

Mas C, Taske N, Deutsch S, Guipponi M, Thomas P, Covanis A, Friis M, Kjeldsen MJ, Pizzolato GP, Villemure J-G, Buresi C, Rees M, et al. 2004. Association of the connexin36 gene with juvenile myoclonic epilepsy. Journal of Medical Genetics 41:e93–e93.

Matsuura T, Yamagata T, Burgess DL, Rasmussen A, Grewal RP, Watase K, Khajavi M, McCall AE, Davis CF, Zu L, Achari M, Pulst SM, et al. 2000. Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. Nature Genetics 26:191–194.

Mattson RH. 2003. Overview: Idiopathic Generalized Epilepsies. Epilepsia 44:2–6.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303.

McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P, Cunningham F. 2016. The Ensembl Variant Effect Predictor. Genome Biol 17:.

McNeely KC, Dwyer ND. 2020. Cytokinesis and postabscission midbody remnants are regulated during mammalian brain development. PNAS 117:9584–9593.

Medina MT, Suzuki T, Alonso ME, Durón RM, Martínez-Juárez IE, Bailey JN, Bai D, Inoue Y, Yoshimura I, Kaneko S, Montoya MC, Ochoa A, et al. 2008. Novel mutations in Myoclonin1/EFHC1 in sporadic and familial juvenile myoclonic epilepsy. Neurology 70:2137–2144.

Mefford HC. 2014. CNVs in Epilepsy. Curr Genet Med Rep 2:162–167.

Mefford HC, Muhle H, Ostertag P, Spiczak S von, Buysse K, Baker C, Franke A, Malafosse A, Genton P, Thomas P, Gurnett CA, Schreiber S, et al. 2010. Genome-Wide Copy Number Variation in Epilepsy: Novel Susceptibility Loci in Idiopathic Generalized and Focal Epilepsies. PLOS Genetics 6:e1000962.

Meyer K, Kirchner M, Uyar B, Cheng J-Y, Russo G, Hernandez-Miranda LR, Szymborska A, Zauber H, Rudolph I-M, Willnow TE, Akalin A, Haucke V, et al. 2018. Mutations in Disordered Regions Can Cause Disease by Creating Dileucine Motifs. Cell 175:239-253.e17.

Miller DT, Shen Y, Weiss LA, Korn J, Anselm I, Bridgemohan C, Cox GF, Dickinson H, Gentile J, Harris DJ, Hegde V, Hundley R, et al. 2009. Microdeletion/duplication at 15q13.2q13.3 among individuals with features of autism and other neuropsychiatric disorders. J Med Genet 46:242–248.

Moen T, Brodtkorb E, Michler RP, Holst A. 1995. Juvenile myoclonic epilepsy and human leukocyte antigens. Seizure 4:119–122.

Moreira DP, Griesi-Oliveira K, Bossolani-Martins AL, Lourenço NCV, Takahashi VNO, Rocha KM da, Moreira ES, Vadasz E, Meira JGC, Bertola D, O'Halloran E, Magalhães TR, et al. 2014. Investigation of 15q11-q13, 16p11.2 and 22q13 CNVs in autism spectrum disorder Brazilian individuals with and without epilepsy. PLoS One 9:e107705.

Moschetta S, Fiore LA, Fuentes D, Gois J, Valente KD. 2011. Personality traits in patients with juvenile myoclonic epilepsy. Epilepsy Behav 21:473–477.

Muhle H, Spiczak S von, Gaus V, Kara S, Helbig I, Hampe J, Franke A, Weber Y, Lerche H, Kleefuss-Lie AA, Elger CE, Schreiber S, et al. 2010. Role of GRM4 in idiopathic generalized epilepsies analysed by genetic association and sequence analysis. Epilepsy Res 89:319–326.

Murakami H, Tamura N, Enomoto Y, Shimasaki K, Kurosawa K, Hanada K. 2020. Intellectual disability-associated gain-of-function mutations in CERT1 that encodes the ceramide transport protein CERT. PLOS ONE 15:e0243980.

Myers KA, Johnstone DL, Dyment DA. 2019. Epilepsy genetics: Current knowledge, applications, and future directions. Clinical Genetics 95:95–111.

Nadadhur AG, Alsaqati M, Gasparotto L, Cornelissen-Steijger P, Hugte E van, Dooves S, Harwood AJ, Heine VM. 2019. Neuron-Glia Interactions Increase Neuronal Phenotypes in Tuberous Sclerosis Complex Patient iPSC-Derived Models. Stem Cell Reports 12:42–56.

Naseer MI, Faheem M, Chaudhary AG, Kumosani TA, Al-Quaiti MM, Jan MM, Saleh Jamal H, Al-Qahtani MH. 2015. Genome wide analysis of novel copy number variations duplications/deletions of different epileptic patients in Saudi Arabia. BMC Genomics 16 Suppl 1:S10.

Neubauer BA, Waldegger S, Heinzinger J, Hahn A, Kurlemann G, Fiedler B, Eberhard F, Muhle H, Stephani U, Garkisch S, Eeg-Olofsson O, Müller U, et al. 2008. KCNQ2 and KCNQ3 mutations contribute to different idiopathic epilepsy syndromes. Neurology 71:177–183.

Nicholas CR, Chen J, Tang Y, Southwell DG, Chalmers N, Vogt D, Arnold CM, Chen Y-JJ, Stanley EG, Elefanty AG, Sasai Y, Alvarez-Buylla A, et al. 2013. Functional Maturation of hPSC-Derived Forebrain Interneurons Requires an Extended Timeline and Mimics Human Neural Development. Cell Stem Cell 12:573–586.

Nijs L de, Léon C, Nguyen L, LoTurco JJ, Delgado-Escueta AV, Grisar T, Lakaye B. 2009. EFHC1 interacts with microtubules to regulate cell division and cortical development. Nature Neuroscience 12:1266–1274.

Nijs L de, Wolkoff N, Coumans B, Delgado-Escueta AV, Grisar T, Lakaye B. 2012. Mutations of EFHC1, linked to juvenile myoclonic epilepsy, disrupt radial and tangential migrations during brain development. Hum Mol Genet 21:5106–5117.

Nijs L de, Wolkoff N, Grisar T, Lakaye B. 2013. Juvenile myoclonic epilepsy as a possible neurodevelopmental disease: Role of EFHC1 or Myoclonin1. Epilepsy Behav 28:S58–S60.

Nikolakopoulou AM, Montagne A, Kisler K, Dai Z, Wang Y, Huuskonen MT, Sagare AP, Lazic D, Sweeney MD, Kong P, Wang M, Owens NC, et al. 2019. Pericyte loss leads to circulatory failure and pleiotrophin depletion causing neuron loss. Nat Neurosci 22:1089–1098.

Noctor SC, Flint AC, Weissman TA, Dammerman RS, Kriegstein AR. 2001. Neurons derived from radial glial cells establish radial units in neocortex. Nature 409:714–720.

Obeid T, Rab MO el, Daif AK, Panayiotopoulos CP, Halim K, Bahakim H, Bamgboye E. 1994. Is HLA-DRW13 (W6) associated with juvenile myoclonic epilepsy in Arab patients? Epilepsia 35:319–321.

Pal DK, Evgrafov OV, Tabares P, Zhang F, Durner M, Greenberg DA. 2003. BRD2 (RING3) is a probable major susceptibility gene for common juvenile myoclonic epilepsy. Am J Hum Genet 73:261–270.

Pan P-Y, Tian J-H, Sheng Z-H. 2009. Snapin facilitates the synchronization of synaptic vesicle fusion. Neuron 61:412–424.

Parihar R, Mishra R, Singh SK, Jayalakshmi S, Mehndiratta MM, Ganesh S. 2014. Association of the GRM4 gene variants with juvenile myoclonic epilepsy in an Indian population. J Genet 93:193–197.

Pfleger CM, Kirschner MW. 2000. The KEN box: an APC recognition signal distinct from the D box targeted by Cdh1. Genes Dev 14:655–665.

Poirier K, Lebrun N, Broix L, Tian G, Saillour Y, Boscheron C, Parrini E, Valence S, Pierre BS, Oger M, Lacombe D, Geneviève D, et al. 2013. Mutations in TUBG1 , DYNC1H1 , KIF5C and KIF2A cause malformations of cortical development and microcephaly. Nature Genetics 45:639–647.

Pólvora-Brandão D, Joaquim M, Godinho I, Aprile D, Álvaro AR, Onofre I, Raposo AC, Pereira de Almeida L, Duarte ST, Rocha ST da. 2018. Loss of hierarchical imprinting regulation at the Prader-Willi/Angelman syndrome locus in human iPSCs. Hum Mol Genet 27:3999–4011.

Pozzi D, Condliffe S, Bozzi Y, Chikhladze M, Grumelli C, Proux-Gillardeaux V, Takahashi M, Franceschetti S, Verderio C, Matteoli M. 2008. Activity-dependent phosphorylation of Ser187 is required for SNAP-25-negative modulation of neuronal voltage-gated calcium channels. PNAS 105:323–328.

Prè D, Nestor MW, Sproul AA, Jacob S, Koppensteiner P, Chinchalongporn V, Zimmer M, Yamamoto A, Noggle SA, Arancio O. 2014. A Time Course Analysis of the Electrophysiological Properties of Neurons Differentiated from Human Induced Pluripotent Stem Cells (iPSCs). PLOS ONE 9:e103418.

Puranam RS, Jain S, Kleindienst AM, Saxena S, Kim M-K, Kelly Changizi B, Padma MV, Andrews I, Elston RC, Tiwari HK, McNamara JO. 2005. A locus for generalized tonic-clonic seizure susceptibility maps to chromosome 10q25-q26. Ann Neurol 58:449–458.

Qin R, Cao S, Lyu T, Qi C, Zhang W, Wang Y. 2017. CDYL Deficiency Disrupts Neuronal Migration and Increases Susceptibility to Epilepsy. Cell Reports 18:380–390.

Raju PK, Satishchandra P, Nayak S, Iyer V, Sinha S, Anand A. 2017. Microtubule-associated defects caused by EFHC1 mutations in juvenile myoclonic epilepsy. Hum Mutat 38:816–826.

Ratnapriya R, Vijai J, Kadandale JS, Iyer RS, Radhakrishnan K, Anand A. 2010. A locus for juvenile myoclonic epilepsy maps to 2q33-q36. Hum Genet 128:123–130.

Revinski DR, Zaragosi L-E, Boutin C, Ruiz-Garcia S, Deprez M, Thomé V, Rosnet O, Gay A-S, Mercey O, Paquet A, Pons N, Ponzio G, et al. 2018. CDC20B is required for deuterosome-mediated centriole production in multiciliated cells. Nat Commun 9:4668.

Riban V, Bouilleret V, Pham-Lê BT, 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 112:101–111.

Rodríguez C, Sánchez-Morán I, Álvarez S, Tirado P, Fernández-Mayoralas DM, Calleja-Pérez B, Almeida Á, Fernández-Jaén A. 2019. A novel human Cdh1 mutation impairs anaphase promoting complex/cyclosome activity resulting in microcephaly, psychomotor retardation, and epilepsy. Journal of Neurochemistry 151:103–115.

Rohena L, Neidich J, Truitt Cho M, Gonzalez KD, Tang S, Devinsky O, Chung WK. 2013. Mutation in SNAP25 as a novel genetic cause of epilepsy and intellectual disability. Rare Dis 1:e26314.

Rossetto MG, Zanarella E, Orso G, Scorzeto M, Megighian A, Kumar V, Delgado-Escueta AV, 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 20:4248–4257.

Rozycka A, Steinborn B, Trzeciak WH. 2009. The 1674+11C>T polymorphism of CHRNA4 is associated with juvenile myoclonic epilepsy. Seizure 18:601–603.

Rudolf G, Lesca G, Mehrjouy MM, Labalme A, Salmi M, Bache I, Bruneau N, Pendziwiat M, Fluss J, Bellescize J de, Scholly J, Møller RS, et al. 2016. Loss of function of the retinoid-related nuclear receptor (RORB) gene and epilepsy. Eur J Hum Genet 24:1761–1770.

Sakai D, Dixon J, Dixon MJ, Trainor PA. 2012. Mammalian Neurogenesis Requires Treacle-Plk1 for Precise Control of Spindle Orientation, Mitotic Progression, and Maintenance of Neural Progenitor Cells. PLOS Genetics 8:e1002566.

Sander T, Bockenkamp B, Hildmann T, Blasczyk R, Kretz R, Wienker TF, Volz A, Schmitz B, Beck-Mannagetta G, Riess O, Epplen JT, Janz D, et al. 1997. Refined mapping of the epilepsy susceptibility locus EJM1 on chromosome 6. Neurology 49:842–847.

Sander T, Schulz H, Saar K, Gennaro E, Riggio MC, Bianchi A, Zara F, Luna D, Bulteau C, Kaminska A, Ville D, Cieuta C, et al. 2000. Genome search for susceptibility loci of common idiopathic generalised epilepsies. Hum Mol Genet 9:1465–1472.

Santos BP dos, Marinho CRM, Marques TEBS, Angelo LKG, Malta MV da S, Duzzioni M, Castro OW de, Leite JP, Barbosa FT, Gitaí DLG. 2017. Genetic susceptibility in Juvenile Myoclonic Epilepsy: Systematic review of genetic association studies. PLoS One 12:e0179629.

Sawamoto K, Wichterle H, Gonzalez-Perez O, Cholfin JA, Yamada M, Spassky N, Murcia NS, Garcia-Verdugo JM, Marin O, Rubenstein JLR, Tessier-Lavigne M, Okano H, et al. 2006a. New neurons follow the flow of cerebrospinal fluid in the adult brain. Science 311:629–632.

Sawamoto K, Wichterle H, Gonzalez-Perez O, Cholfin JA, Yamada M, Spassky N, Murcia NS, Garcia-Verdugo JM, Marin O, Rubenstein JLR, Tessier-Lavigne M, Okano H, et al. 2006b. New neurons follow the flow of cerebrospinal fluid in the adult brain. Science 311:629–632.

Schapira M, Tyers M, Torrent M, Arrowsmith CH. 2017. WD40 repeat domain proteins: a novel target class? Nature Reviews Drug Discovery 16:773–786.

Scheffer IE, Berkovic S, Capovilla G, Connolly MB, French J, Guilhoto L, Hirsch E, Jain S, Mathern GW, Moshé SL, Nordli DR, Perucca E, et al. 2017a. ILAE Classification of the Epilepsies Position Paper of the ILAE Commission for Classification and Terminology. Epilepsia 58:512–521.

Scheffer IE, Berkovic S, Capovilla G, Connolly MB, French J, Guilhoto L, Hirsch E, Jain S, Mathern GW, Moshé SL, Nordli DR, Perucca E, et al. 2017b. ILAE classification of the epilepsies: Position paper of the ILAE Commission for Classification and Terminology. Epilepsia 58:512–521.

Scheffer IE, Mefford HC. 2014. Epilepsy: Beyond the single nucleotide variant in epilepsy genetics. Nat Rev Neurol 10:490–491.

Selak S, Paternain AV, Aller IM, Picó E, Rivera R, Lerma J. 2009. A Role for SNAP25 in Internalization of Kainate Receptors and Synaptic Plasticity. Neuron 63:357–371.

Sharp AJ, Mefford HC, Li K, Baker C, Skinner C, Stevenson RE, Schroer RJ, Novara F, De Gregori M, Ciccone R, Broomer A, Casuga I, et al. 2008. A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. Nat Genet 40:322–328.

Shihab HA, Rogers MF, Gough J, Mort M, Cooper DN, Day INM, Gaunt TR, Campbell C. 2015. An integrative approach to predicting the functional effects of non-coding and coding sequence variation. Bioinformatics 31:1536–1543.

Shinawi M, Liu P, Kang S-HL, Shen J, Belmont JW, Scott DA, Probst FJ, Craigen WJ, Graham BH, Pursley A, Clark G, Lee J, et al. 2010. Recurrent reciprocal 16p11.2 rearrangements associated

with global developmental delay, behavioural problems, dysmorphism, epilepsy, and abnormal head size. J Med Genet 47:332–341.

Sigl R, Wandke C, Rauch V, Kirk J, Hunt T, Geley S. 2009. Loss of the mammalian APC/C activator FZR1 shortens G1 and lengthens S phase but has little effect on exit from mitosis. Journal of Cell Science 122:4208–4217.

Stefansson H, Rujescu D, Cichon S, Pietiläinen OPH, Ingason A, Steinberg S, Fossdal R, Sigurdsson E, Sigmundsson T, Buizer-Voskamp JE, Hansen T, Jakobsen KD, et al. 2008. Large recurrent microdeletions associated with schizophrenia. Nature 455:232–236.

Steffens M, Leu C, Ruppert A-K, Zara F, Striano P, Robbiano A, Capovilla G, Tinuper P, Gambardella A, Bianchi A, La Neve A, Crichiutti G, et al. 2012. Genome-wide association analysis of genetic generalized epilepsies implicates susceptibility loci at 1q43, 2p16.1, 2q22.3 and 17q21.32. Hum Mol Genet 21:5359–5372.

Stogmann E, Lichtner P, Baumgartner C, Bonelli S, Assem-Hilger E, Leutmezer F, Schmied M, Hotzy C, Strom TM, Meitinger T, Zimprich F, Zimprich A. 2006. Idiopathic generalized epilepsy phenotypes associated with different EFHC1 mutations. Neurology 67:2029–2031.

Strand AD, Aragaki AK, Baquet ZC, Hodges A, Cunningham P, Holmans P, Jones KR, Jones L, Kooperberg C, Olson JM. 2007. Conservation of Regional Gene Expression in Mouse and Human Brain. PLOS Genetics 3:e59.

Strehlow V, Swinkels MEM, Thomas RH, Rapps N, Syrbe S, Dorn T, Lemke JR. 2016. Generalized Epilepsy and Myoclonic Seizures in 22q11.2 Deletion Syndrome. Mol Syndromol 7:239–246.

Striano P, Weber YG, Toliat MR, Schubert J, Leu C, Chaimana R, Baulac S, Guerrero R, LeGuern E, Lehesjoki A-E, Polvi A, Robbiano A, et al. 2012. GLUT1 mutations are a rare cause of familial idiopathic generalized epilepsy. Neurology 78:557–562.

Sunkel CE, Glover DM. 1988. polo, a mitotic mutant of Drosophila displaying abnormal spindle poles. Journal of Cell Science 89:25–38.

Suzuki T, Delgado-Escueta AV, Aguan K, Alonso ME, Shi J, Hara Y, Nishida M, Numata T, Medina MT, Takeuchi T, Morita R, Bai D, et al. 2004. Mutations in EFHC1 cause juvenile myoclonic epilepsy. Nat Genet 36:842–849.

Suzuki T, Miyamoto H, Nakahari T, Inoue I, Suemoto T, Jiang B, Hirota Y, Itohara S, Saido TC, Tsumoto T, Sawamoto K, Hensch TK, et al. 2009. Efhc1 deficiency causes spontaneous myoclonus and increased seizure susceptibility. Hum Mol Genet 18:1099–1109.

Szafranski P, Schaaf CP, Person RE, Gibson IB, Xia Z, Mahadevan S, Wiszniewska J, Bacino CA, Lalani S, Potocki L, Kang S-H, Patel A, et al. 2010. Structures and Molecular Mechanisms for Common 15q13.3 Microduplications Involving CHRNA7: Benign or Pathological? Hum Mutat 31:840–850.

Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S. 2007. Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. Cell 131:861–872.

Tan GW, Kondo T, Murakami N, Imamura K, Enami T, Tsukita K, Shibukawa R, Funayama M, Matsumoto R, Ikeda A, Takahashi R, Inoue H. 2017. Induced pluripotent stem cells derived

from an autosomal dominant lateral temporal epilepsy (ADLTE) patient carrying S473L mutation in leucine-rich glioma inactivated 1 (LGI1). Stem Cell Research 24:12–15.

Tan NCK, Mulley JC, Berkovic SF. 2004. Genetic Association Studies in Epilepsy: "The Truth Is Out There." Epilepsia 45:1429–1442.

Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. 2015. Sambamba: fast processing of NGS alignment formats. Bioinformatics 31:2032–2034.

Terré B, Lewis M, Gil-Gómez G, Han Z, Lu H, Aguilera M, Prats N, Roy S, Zhao H, Stracker TH. 2019. Defects in efferent duct multiciliogenesis underlie male infertility in GEMC1-, MCIDAS- or CCNO-deficient mice. Development 146:.

Terré B, Piergiovanni G, Segura-Bayona S, Gil-Gómez G, Youssef SA, Attolini CS-O, Wilsch-Bräuninger M, Jung C, Rojas AM, Marjanović M, Knobel PA, Palenzuela L, et al. 2016. GEMC1 is a critical regulator of multiciliated cell differentiation. The EMBO Journal 35:942–960.

Thakran S, Guin D, Singh P, Singh P, Kukal S, Rawat C, Yadav S, Kushwaha SS, Srivastava AK, Hasija Y, Saso L, Ramachandran S, et al. 2020. Genetic Landscape of Common Epilepsies: Advancing towards Precision in Treatment. Int J Mol Sci 21:.

The Epileptology of Théodore Herpin (1799–1865) - Eadie - 2002 - Epilepsia - Wiley Online Library.

Thounaojam R, Langbang L, Itisham K, Sobhani R, Srivastava S, Ramanujam B, Verma R, Tripathi M, Aguan K. 2017. EFHC1 mutation in Indian juvenile myoclonic epilepsy patient. Epilepsia Open 2:84–89.

Tian J, Tian C, Ding Y, Li Z, Geng Q, Xiahou Z, Wang J, Hou W, Liao J, Dong M-Q, Xu X, Li J. 2015. Aurora B-dependent phosphorylation of Ataxin-10 promotes the interaction between Ataxin-10 and Plk1 in cytokinesis. Sci Rep 5:.

Tidball AM, Parent JM. 2016. Concise Review: Exciting Cells: Modeling Genetic Epilepsies with Patient-Derived Induced Pluripotent Stem Cells. STEM CELLS 34:27–33.

Tigges U, Welser-Alves JV, Boroujerdi A, Milner R. 2012. A novel and simple method for culturing pericytes from mouse brain. Microvasc Res 84:74–80.

Tompson S, Young T. 2017. Assaying the Effects of Splice Site Variants by Exon Trapping in a Mammalian Cell Line. BIO-PROTOCOL 7:.

Trujillo CA, Gao R, Negraes PD, Gu J, Buchanan J, Preissl S, Wang A, Wu W, Haddad GG, Chaim IA, Domissy A, Vandenberghe M, et al. 2019. Complex Oscillatory Waves Emerging from Cortical Organoids Model Early Human Brain Network Development. Cell Stem Cell 25:558-569.e7.

Turco EM, Vinci E, Altieri F, Ferrari D, Torres B, Goldoni M, Lamorte G, Tata AM, Mazzoccoli G, Postorivo D, Della Monica M, Bernardini L, et al. 2018. Copy number variations in healthy subjects. Case study: iPSC line CSSi005-A (3544) production from an individual with variation in 15q13.3 chromosome duplicating gene CHRNA7. Stem Cell Research 32:73–77.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3--new capabilities and interfaces. Nucleic Acids Res 40:e115.

Valente EM, Logan CV, Mougou-Zerelli S, Lee JH, Silhavy JL, Brancati F, Iannicelli M, Travaglini L, Romani S, Illi B, Adams M, Szymanska K, et al. 2010. Mutations in TMEM216 perturb ciliogenesis and cause Joubert, Meckel and related syndromes. Nat Genet 42:619–625.

Vijai J, Cherian PJ, Stlaja PN, Anand A, Radhakrishnan K. 2003a. Clinical characteristics of a South Indian cohort of juvenile myoclonic epilepsy probands. Seizure 12:490–496.

Vijai J, Kapoor A, Ravishankar HM, Cherian PJ, Girija AS, Rajendran B, Rangan G, Jayalakshmi S, Mohandas S, Radhakrishnan K, Anand A. 2003b. Genetic association analysis of KCNQ3 and juvenile myoclonic epilepsy in a South Indian population. Hum Genet 113:461–463.

Wang Q, He K, Li Z, Chen J, Li W, Wen Z, Shen J, Qiang Y, Ji J, Wang Y, Shi Y. 2014a. The CMYA5 gene confers risk for both schizophrenia and major depressive disorder in the Han Chinese population. The World Journal of Biological Psychiatry 15:553–560.

Wang W, Wu T, Kirschner MW. 2014b. The master cell cycle regulator APC-Cdc20 regulates ciliary length and disassembly of the primary cilium. Elife 3:e03083.

Weinstein J, Jacobsen FW, Hsu-Chen J, Wu T, Baum LG. 1994. A novel mammalian protein, p55CDC, present in dividing cells is associated with protein kinase activity and has homology to the Saccharomyces cerevisiae cell division cycle proteins Cdc20 and Cdc4. Mol Cell Biol 14:3350–3363.

Wiegand C, Banerjee I. 2019. Recent advances in the applications of iPSC technology. Current Opinion in Biotechnology 60:250–258.

Wu Y-T, Hsu Y-H, Huang C-Y, Ho M-C, Cheng Y-C, Wen C-H, Ko H-W, Lu H-E, Chen Y-C, Tsai C-L, Hsu Y-C, Wei Y-H, et al. 2018. Generation of an induced pluripotent stem cell (iPSC) line from a 40-year-old patient with the A8344G mutation of mitochondrial DNA and MERRF (myoclonic epilepsy with ragged red fibers) syndrome. Stem Cell Research 27:10–14.

Xu C, Min J. 2011. Structure and function of WD40 domain proteins. Protein Cell 2:202–214.

Yamamuro T, Kano K, Naito K. 2008. Functions of FZR1 and CDC20, Activators of the Anaphase-Promoting Complex, During Meiotic Maturation of Swine Oocytes. Biol Reprod 79:1202–1209.

Yang X, Feng M, Jiang X, Wu Z, Li Z, Aau M, Yu Q. 2009a. miR-449a and miR-449b are direct transcriptional targets of E2F1 and negatively regulate pRb–E2F1 activity through a feedback loop by targeting CDK6 and CDC25A. Genes Dev 23:2388–2393.

Yang Y, Kim AH, Yamada T, Wu B, Bilimoria PM, Ikeuchi Y, Iglesia N de la, Shen J, Bonni A. 2009b. A Cdc20-APC ubiquitin signaling pathway regulates presynaptic differentiation. Science 326:575–578.

Zafeiriou M-P, Bao G, Hudson J, Halder R, Blenkle A, Schreiber M-K, Fischer A, Schild D, Zimmermann W-H. 2020. Developmental GABA polarity switch and neuronal plasticity in Bioengineered Neuronal Organoids. Nat Commun 11:3791.

Zara F, Bianchi A, Avanzini G, Di Donato S, Castellotti B, Patel PI, Pandolfo M. 1995. Mapping of genes predisposing to idiopathic generalized epilepsy. Hum Mol Genet 4:1201–1207.

Zhou B, Zhu Y-B, Lin L, Cai Q, Sheng Z-H. 2010. Snapin deficiency is associated with developmental defects of the central nervous system. Bioscience Reports 31:151–158.