On genetic aspects of hot water epilepsy, a sensory epilepsy triggered by touch and temperature stimuli

A thesis submitted for the degree of

Doctor of Philosophy

by

Shalini Roy Choudhury

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

August 2018

Dedicated to

Ma

Baba

For being there always.

Declaration

I hereby declare that this thesis entitled **"On genetic aspects of hot water epilepsy, a sensory epilepsy triggered by touch and temperature stimuli"** is an authentic record of research work carried out by me under the guidance of Prof. Anuranjan Anand at the Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, India. This work has not been submitted elsewhere for the award of any other degree.

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

Shalini Roy Choudhury

Place: Bangalore

Date:

Certificate

This is to certify that the work described in this thesis entitled **"On genetic aspects of hot water epilepsy, a sensory epilepsy triggered by touch and temperature stimuli"** is the result of the investigations carried out by Ms. Shalini Roy Choudhury in the Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, India, 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.

Prof. Anuranjan Anand
Molecular Biology and Genetics Unit
Jawaharlal Nehru Centre for Advanced Scientific Research
Bangalore 560 064

Place: Bangalore
Date:

Acknowledgements

I wish to thank a lot of people who have contributed in various ways to help me grow professionally and personally during these previous years at the JNCASR. I take this opportunity to thank some of them without whom this work would not have been possible.

I express my sincere gratitude towards my supervisor, Prof. Anuranjan Anand, for giving me an opportunity to explore the field of hot water epilepsy; to work on different, interesting and challenging projects. I thank him for his invaluable guidance, support, patience, discussions, encouragement, and positive criticisms all through; especially when things did not go as planned. I also thank him for providing the ideal scientific provisions and an organized and disciplined laboratory environment to explore and develop my potentials. Prof. Anand has been instrumental in shaping my interest in the field of human genetic disorders.

I extend my gratitude towards the faculty members of the Molecular Biology and Genetics Unit (MBGU): Prof. MRS Rao and Prof. Tapas Kundu for their valuable suggestions, Prof. Udaykumar Ranga, Prof. Kaustuv Sanyal, Prof. Hemalatha Balaram, Prof. Maneesha Inamdar, Prof. Namita Surolia and Dr. Ravi Manjithaya, for their constructive comments, discussions, and encouragement during departmental seminars, and at different instances of interactions I had chance to have with them. I thank Dr. James Chelliah for his helpful guidance, motivation and encouragement. I thank our clinical collaborators from NIMHANS, Bangalore: Dr. P. Satishchandra and Dr. Sanjib Sinha, for help in ascertaining the epilepsy families. I wish to thank all the families and subjects who participated in this study.

I extend my thanks to all the past and present members of the lab: Nishtha, Rahul, Mariyam, Manpreet, Praveen, Kalpita, Sourav, Vikas, Deepti, Minu, Varsha, Chandrashekhar, Mohan, Rammurthy, Shveta, Pooja, Shrilaxmi, Sambhavi, Girija and Moumita. I especially thank my seniors Nishtha, Praveen and Kalpita, for creating a stimulating lab environment, their constant support, scientific discussions, and interest in my work and for inspiring me with positivity. I thank Shveta for always being a lovely company. I am thankful to Sourav for his help with the genomics-related queries and for being a great friend and person who has influenced me in many ways.

I acknowledge support from the confocal, sequencing, and central facilities at MBGU, computer lab, library, academic and administrative sections. Many thanks to Mr. Sunil, Ms. Ramya, Ms. Nandhini, Mr. Raju, and Ms. Usha for their help and support. I specially thank Ms. Anitha, Ms. Greeshma, Ms. Suma, Dr. Archana, Ms. Uma, Ms. Tara, and Dr. Elizabeth for their interest in my general well-being. I thank all the staff at JNCASR who made life in the campus safe and comfortable.

I acknowledge financial support provided by CSIR and JNCASR.

I extend my thanks towards some people who had helped me during my work: Sutanuka, Anjali, Arnab, Pallabi, members of the Vascular Biology, HIV-AIDS and Autophagy labs. Some lovely people I met in and outside the campus: Simi, Neha, Stephanie, Prabhu, Debanjan, Abhik, Krishnendu, Shalini, Vandana, and Shiksha. My childhood friend Anindita, from whom I try to learn a thing or two about being meticulous. My love for food, movies, travel, books, art and culture always provided a warm nook during trying times. To Nishit, I cannot thank him enough for his unwavering support, for being a patient listener, and for the efforts to make me see the light when the days were not bright.

Little is possible without a loving family and parents. My grandmother who amazed me with her memory and progressive thoughts- I thank her for inspiring me in ways unconceivable. I thank Bonnie mashi for her love and care, and bauro mama for influencing me with his work and his kind deeds towards family and beyond. I wish to thank many people in my family who have influenced me in various ways through their ideas, achievements, actions, love and kindness. Thanks to my sister and brother-in-law who I look up to for their sincerity at work and their vigour in life. Finally, none of this would have been possible without the endless love, faith, and motivation my parents have endowed me with. Time and again they inspire me with their knowledge, zeal and positivity; and a way of life and being that I admire and try to imbibe.

Table of Contents

Abbreviations

A	adenine
C	cytosine
G	guanine
T	thymine
θ	recombination fraction
°C	degree Celsius
cm	centimetre
µg	microgram
µl	microlitre
µM	micromolar
Mb	megabases
mg	milligrams
mM	millimolar
M	molar
ml	millilitre
mm	millimetre
ng	nanogram
nm	nanometre
pmol	picomole
BCA	bicinchoninic acid
bp	base pairs
BSA	bovine serum albumin
BWA	burrows wheeler aligner
C-terminal	carboxy-terminal
cDNA	complementary deoxyribose nucleic acid
CO_2	carbon dioxide
CPS	complex partial seizures
DAPI	4', 6-diamidino-2-phenylindole
DMSO	dimethyl sulfoxide
DNA	deoxyribose nucleic acid
dNTP	deoxyribonucleotide triphosphate
EDTA	ethylene diamine tetra-acetate
ER	endoplasmic reticulum
EEG	electroencephalogram
FBS	Fetal bovine serum
g	unit of acceleration
g	grams
GDP	guanosine diphosphate
GFP	green fluorescent protein
GTCS	Generalized tonic clonic seizure
HCl	hydrochloric acid
HWE	hot water epilepsy
HPA	Human Protein Atlas
ILAE	International League Against Epilepsy
kb	kilobases
kDa	kiloDaltons

LOD	logarithm of odds
MAF	minor allele frequency
MW	molecular weight
mRNA	messenger ribonucleic acid
NMD	nonsense mediated decay
N-terminal	amino-terminal
P-bodies	processing bodies
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PFA	paraformaldehyde
pH	power of hydrogen
PTCs	premature translation-termination codons
PVDF	polyvinylidene difluoride
RNA	ribonucleic acid
rpm	revolutions per minute
SD	standard deviation
SDS	sodium dodecyl sulphate
SIFT	Sorting Intolerant From Tolerant
SNP	single-nucleotide polymorphism
SNV	single-nucleotide variation
TAE	tris-acetate-EDTA
Taq	*Thermus aquaticus*
U	units
UTR	untranslated region
WT	wild-type
Z_{max}	maximum LOD score

Chapter 1

Introduction

1.1. Epilepsy and seizures

Epilepsy is a relatively common neurological condition defined by recurrent, usually unprovoked seizures resulting from episodic cortical hyperexcitability. The International League Against Epilepsy (ILAE) defines it as a disorder of the brain characterized by a predisposition to generate epileptic seizures and by its neurobiologic, cognitive, psychological, and social consequences (Fisher et al 2005). The most recent operational clinical definition suggests epilepsy as a brain disease complying any of the following conditions: (i) at least two unprovoked (or reflex) seizures occurring more than 24 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). Epilepsy has worldwide lifetime cumulative incidence of about 3% (Hildebrand et al 2013). All epilepsies exhibit seizures. An epileptic seizure is a transient occurrence of signs and/or symptoms caused by synchronous or abnormal excessive neuronal activity in the brain that is usually intermittent, self-limiting and occurs with or without loss of consciousness. Generalized seizures involve all of the brain or large portions of it, are asymmetric in origin, and in most cases, spread rapidly among cortical and subcortical structures manifesting bilateral symptoms. These seizures likely involve disturbances in existing thalamocortical networks (Kramer and Cash 2012). The diagnosis of generalized epilepsy is made on clinical basis, determined by typical interictal EEG discharges. Focal epilepsies comprise unifocal and multifocal seizures, and those that involve single hemisphere (Scheffer et al 2017). Although well distinguished, these events often occur as a spectrum of overlapping or related mechanisms. Epilepsy is considered as a disorder of cortical network organization (Kramer and Cash 2012).

1.1.1. Classification of seizures

The first classification of seizures proposed in 1989, has been revised twice over time to provide a more practical classification for the disorder (Fisher et al 2017). The basic classification by ILAE 2017 categorizes seizures beginning with its first identification for

being focal or generalized. When difficult to identify the onset, the cases are considered as seizures of unknown onset. And, there is a category for unclassified seizures, due to insufficient information or inability to place them under any specific category. Diagnosis of seizure type is followed by the next step of diagnosis of the epilepsy type.

1.1.2. Classification of epilepsy

Epilepsy is a heterogeneous clinical entity with high variation in its etiological factors and clinical manifestation. John Hughlings Jackson's work in the mid-19th century has significantly influenced our modern day understanding of epilepsy. He had laid the foundation for a scientific approach to the study of the disorder by describing definite physiology and semiology of different types of epilepsy seizures (Magiorkinis 2014). Subsequently, several classification systems have been developed for epilepsy. The current International Classification of Epilepsies and Epileptic Syndromes (ICE) is based on the revised ILAE classification of epilepsies (Scheffer et al 2017). The clinically more relevant, renewed classification of epilepsy presents itself at three levels: diagnosis of seizure type, followed by diagnosis of epilepsy type, and diagnosis of an epilepsy syndrome, if any. The diagnosis of seizure type is based on the ILAE 2017 classification of seizures. The diagnosis of epilepsy type assumes the patient is already diagnosed for epilepsy (according to ILAE 2014 definition of epilepsy). The epilepsy type includes generalized epilepsy, focal epilepsy, a new category of combined generalized and focal epilepsy, and an unknown category where epilepsy is diagnosed but the type cannot be determined. Generalized and focal epilepsies are diagnosed clinically, supported with EEG findings showing typical epileptiform discharges. The combined generalized and focal epilepsy category that includes Dravet syndrome and Lennox-Gastaut syndrome is diagnosed employing clinical and ictal EEG observations. Epileptiform discharges are not essential for diagnosis of epilepsy. Diagnosis of an epilepsy syndrome is based on a group of criteria including seizure types, EEG and imaging features that appear to occur together. It can have a variety of comorbidities which mostly serve as a guide to management.

1.1.3. Genetic basis of epilepsy

Human epilepsy is a heterogeneous condition in which genetic factors contribute to the etiology of at least 35% of all epilepsies (McNamara and Puranam 1998). Despite diverse causes that underlie epilepsy, it is considered to be a highly genetic and in many cases, a

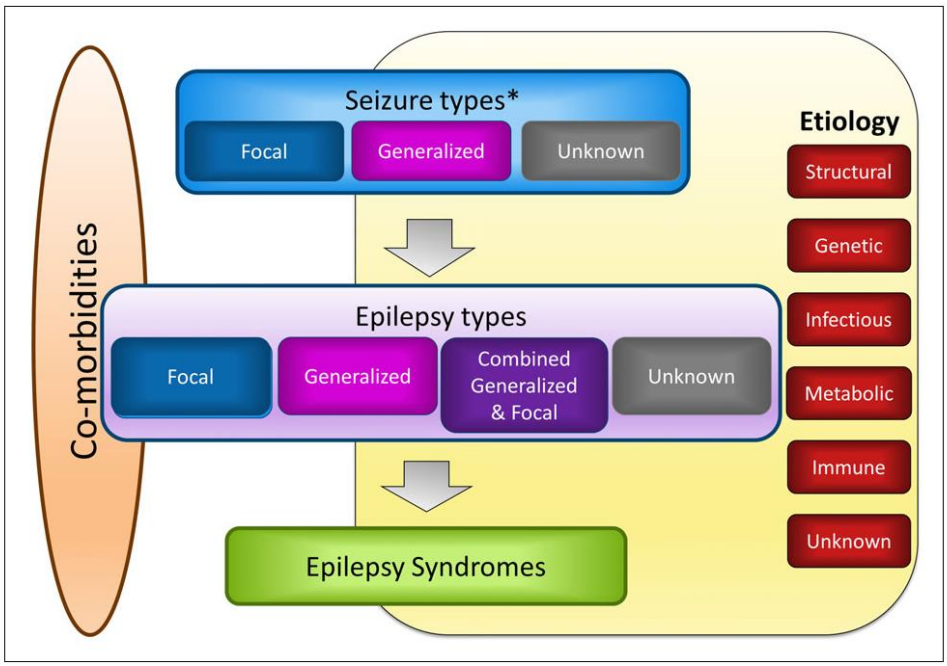


Figure 1.1: ILAE 2017 framework for classification of epilepsies (Scheffer et al 2017)

heritable condition (Poduri and Lowenstein 2011). Evidence from a variety of investigations including twin studies, familial empirical risk calculations, family aggregation studies, studies involving single-gene epilepsy disorders, and animal models has helped to establish the fact that there is substantial genetic basis to the pathophysiology of epilepsy (McNamara and Puranam 1998). The genetic basis of epilepsies was first identified in monogenic familial cases and in disorders where seizures occur as comorbidity (Thomas and Berkovic 2014). Before the introduction of large-scale sequencing methods, epilepsy genetic studies had been majorly based on classical genetic approach of studying Mendelian inheritance in large pedigrees, by linkage mapping of critical loci, and sequencing relevant genes to identify pathogenic mutations (Poduri and Lowenstein 2011). Mode of genetic inheritance in familial epilepsies has been mostly autosomal dominant. Genetic influences are usually stronger in generalized, compared to focal epilepsies. The variable penetrance and expressivity seen in epilepsy suggest that besides a primary mutation, there might be influences of other genetic and non-genetic factors in phenotypic manifestation.

Major recent advances in the field have been the identification of copy number variations (Olson 2014) and *de novo* mutations (Allen 2013). More than 40 genes have been identified in human families and mouse models. Many of these genes encode voltage-gated ion channels, entitling certain classes of epilepsy as channelopathies. In addition, identification

of a few non-ion channel genes has provided evidence for the basis of somewhat unanticipated pathways and mechanisms underlying seizure activity (Robinson and Gardiner 2004). These include neuronal maturation and migration factors (*EFHC1, PCDH19, LIS1*), neurotransmitter release machinery (*STXBP1, SYN1*), neurotransmitter receptors (*GABRA1, CHRNA7*), transcription factors (*MECP2, ARX*) (Poduri and Lowenstein 2011), and RNA binding proteins (*NOVA2*). By identifying the specific genetic mutation in an epilepsy patient, the diagnosis, prognosis, treatment, management and risk assessment in family members can be effectively conducted and integrated (Garofalo et al 2012). Identification of epilepsy-causing genes has opened a wider area to delve into finer and more intricate questions directed towards the understanding of brain development, neuronal networks, and epileptogenesis.

1.1.4. Animal models of epilepsy

Identification of several gene mutations in various human epilepsies has provided information for its genetic basis. However, studying the functional shortcomings of these genes in the human disease scenario remains a challenging task. Animal models of epilepsy attempt to unravel several of the underlying neuroanatomical, biochemical and genetic factors that lead to their cause (Grone and Baraban 2015). As early as 1869, John Hughlings Jackson valued the study of comparative physiology of spontaneous seizures in dogs. Soon after, work on rabbits, guineapigs, cats and dogs indicated that electrical stimulation of cortex generated seizures resembling those of human epilepsy (Ferrier 1873, Grone and Baraban 2015). Cats were used as models for research comprising anti-convulsant drugs during 1930s (Putnam and Merritt 1937, Grone and Baraban 2015). Earlier, dogs, non-human primates and cats were widely used models for epilepsy. Relatively less commonly used species also provide unusual insight into epilepsy mechanism. *Tritonia diomedea* (a nudibranch mollusc) has been important in studying brain injury in epilepsy (Sakurai et al 2014). *Xenopus* tadpoles have been used to study the biochemical process of resistance to subsequent seizures (Bell et al 2011). Turtles allow the study of resistance to hyperexcitability and cell death (Pamenter et al 2012). Chimeric brains made in birds with embryological implants help studying epileptic foci (Margoliash et al 2010, Teillet et al 1991). In recent times, rodent models of epilepsy have become appreciably the most common animal models. Mice models for type I lissencephaly, tuberous sclerosis complex and Dravet syndrome have helped to identify underlying circuit aberrations in epileptic conditions. *Drosophila melanogaster* and *Caenorhabditis elegans* are simple models that

have been helpful in understanding the basic mechanism of neurotransmission in vertebrates. *Danio rerio* (zebrafish) serves as another elegant system for modeling acute seizures and genetic epilepsies. Zebrafish models for *Ube3a* in Angelman's syndrome (Hortopan et al 2010), *Scn1a* in Dravet syndrome (Baraban et al 2013) and *Ocrl1* in Lowe's syndrome (Ramirez et al 2012) have provided for effective confirmation of epilepsy with insights into the disorder. Simple invertebrates to higher order vertebrates serve as models for epilepsy. Choosing a right animal model for its relevance to the question in hand can provide significant understanding of the mechanisms underlying epilepsy.

1.2. Reflex epilepsy

Reflex or sensory epilepsy comprises a class of about 1% of all human epilepsies (Puranam and McNamara 2001). It can be defined as *sensory precipitation of seizures* (proposed by Pennfield, 1941) where seizures are consistently precipitated by distinctly identifiable and controllable triggers, that may be exogenous (environmental) or endogenous (mental activities) in nature (Engel and Pedley 2008). Reflex epilepsies are identified by the precipitant stimulus and the clinical-EEG response (Panayiotopoulos 1996). Many relatively simple stimuli that are normal to unaffected individuals can evoke seizure activity in the brain of reflex epilepsy patients. These include visual, auditory, somatosensory stimuli, or complex ones like reading, writing, doing arithmetic, solving puzzles or even thinking about specific topics. While seizures observed in reflex epilepsy are not different from the spontaneous seizures observed in non-reflex epilepsy, they do require a specific stimulus for presentation. The reflex epilepsies are interesting and intriguing, and probably represent extremes of a continuum, causing rapid transition of the brain from normal to a seizing condition when provoked by a stimulus.

1.2.1. Classification of reflex epilepsy

The ILAE defines reflex seizures (RS) as those that are evoked by a specific afferent stimulus or by activity of the patient (Blume et al 2001), and reflex epilepsy syndromes as those where all epileptic seizures are precipitated by sensory stimuli (Engel 2001, Illingworth and Ring 2013). The proposed ILAE classification in 2001 determined the types of reflex epilepsies as idiopathic photosensitive occipital lobe epilepsy (IPOLE) and other visual sensitive epilepsies, primary reading epilepsy, startle, language-induced, eating, musicogenic and thinking induced epilepsy. It also enlists the precipitating triggers as visual

stimuli (5% of reflex epilepsies), thinking, eating, reading, praxis, music, somatosensory, proprioceptive, hot water and startle (Kasteleijn-Nolst Trenité 2012).

1.2.2. Stimuli for reflex seizures

A single sensory stimulus in the form of an unexpected sound or surprise can induce seizures in startle epilepsy. These seizures have been reported in case of Down syndrome, Lennox-Gastaut syndrome and dysplasias. On the other hand, most often, repeated stimuli is a key to seizure generation where inducing factors can be in the form of flickering lights, loud noise, bathing in hot water, playing video games, eating, and rubbing of skin. The stimuli for reflex induction of seizures are highly variable in content and can have different latencies in susceptible individuals. Sudden noise or tactile stimulus causes rare reflex epilepsy in infants characterized by myoclonus with polyspike-waves (Ricci et al 1995). With increasing age of onset, the complexity of stimuli increases from simple sensory to complex cognitive-emotional (Koepp et al 2016).

1.2.3. Response to stimuli

According to ILAE 2001, reflex seizures can be encountered in both generalized and focal epilepsy syndromes. Clinically, reflex seizures are of three types: pure reflex seizures; those that are associated with spontaneous seizures, focal, and generalized epilepsy; and isolated seizures that can occur without a diagnosis of epilepsy (Xue and Ritaccio 2006). *Generalized reflex seizures*, such as myoclonic jerks (MJs) and generalized tonic-clonic seizures (GTCS) can occur independently or be a part of epileptic syndromes. MJs are the most common type of seizures in the limbs, trunk or jaw muscles in reading epilepsy, or eyelids in eyelid myoclonia with absences. Absence seizures constitute mostly of specific stimuli like photic, pattern, fixation-off, proprioceptive, cognitive, emotional or linguistic (Duncan and Panayiotopoulos 1995). Absences are more common in self-induction.

Focal reflex seizures are mostly seen in focal or lobular epilepsy like visual seizures of photosensitive type in occipital lobe, or complex focal temporal lobe seizures of musicogenic epilepsy. Seizures may start from a focus but may spread to a broader area later, like in the cases of photic stimulation where spikes are first marked at occipital lobe which spread to other cortical areas or become generalized as in the case of photoparoxysmal responses of idiopathic generalized epilepsy (IGE). Conversely, reading may elicit responses strictly confined to the brain regions sub-serving reading, such as alexia associated with focal ictal EEG paroxysms. A lot of variability is observed among individuals towards

response to same stimulus. Age-related changes in the semiology and topography of physiological networks involved in triggering reflex seizures at different stages of brain development suggest that reflex seizures and epilepsies are closely linked to the synaptic pruning in adolescence that makes the brain functionally more segregated (Koepp et al 2016). Hence, with brain maturation, the seizure semiology too can change.

1.2.4. Self-induction

Self-induction of seizures by patients is regarded as the most uncooperative behaviour in the treatment of epilepsy (Binnie 1988). Auto induction is an interesting phenomenon in some reflex epilepsy cases wherein individuals invoke pleasure by repeated exposure to the seizure inducing stimuli. A large number of reports on self-induction are due to photosensitive epilepsy (PSE) exhibited by a quarter of patients, both children as well as adults (Kasteleijn-Nolst Trenité 1989). Patients often get 'magnetically' attracted to light, a condition called sunflower epilepsy (Ames and Saffer 1983) and voluntarily 'sunflower maneuver' their fingers in front of the eyes to induce seizures. As reported by many patients, self-induction is supposed to give a feeling of pleasure and relieve stress (Binnie 1988). About 30% of hot water epilepsy (HWE) cases in India exhibit self-induction behavior (Satishchandra 2003). This behavior is often observed in children with Dravet syndrome exhibiting learning disability but in able individuals, self-induction can be often controlled. 'Rub' epilepsy is another example under this category where patients evoke seizures deliberately by repeated rubbing an area of the skin (Kanemoto et al 2001). Self-induction has been seen predominantly in PSE, with an occurrence of 24-30% in some selected series (Darby et al 1980, Kasteleijn-Nolst Trenité 1989, Binnie 1988). A comparative Turkish study between HWE and PSE showed high frequency of self-induction in the HWE group (Bebek et al 2006). Self-induced seizures are fiercely resistant to antiepileptic treatment. Dopamine antagonist trials in some patients have suggested lessening this behaviour, by making self-induced seizures non-rewarding (Binnie 1988).

1.2.5. Seizure suppression through stimuli

Predisposition to seizures occurs as an outcome of genetic determinants. Experiencing a seizure on an already genetically predisposed setting may lead to further aggravation of epilepsy. However, in certain paradigms, a stimulus can both precipitate and abort a seizure. The type of response to the stimulus depends largely on the state of cortical activation at the instant of input. An input to neurons in the resting condition or at the intermediate state of

activation may precipitate a seizure. Seizure interruption may occur when epileptic discharge has already begun or a critical mass of neurons responding to a strong stimulus block the spread of seizure activity (Wolf 2005). Response to intermittent photic stimulation (IPS) has been shown to inhibit seizures in 30% of focal epilepsy and 6% of generalized epilepsy patients through EEG studies. But a rebound has been noted after withdrawal of IPS (Stevens 1962). Suppression of seizure is observed in photoparoxysmal responses during IPS when the patient is presented with pictures, listens to clicking sounds or performs calculations. Exercising too can provoke or inhibit seizures.

1.3. Overview of types of reflex epilepsies

1.3.1. Photosensitive epilepsy

Photosensitive reflex epilepsy or photoparoxysmal response (PPR) is the most common reflex epilepsy type (Doose et al 2002) and is characterized by seizure generation due to visual cues. Classically, an abnormal response to light meant photosensitivity, but with advent of EEG and newer technology, an abnormal response to flicker stimulation during EEG recording is generally called photosensitivity (Zifkin and Trenité 2000). PPR is induced by intermittent light frequency of 15-18 Hz (Hughes 2008) and discharges are majorly seen in the occipital lobe of the cortex (Wolf and Goosses 1986). Photosensitivity is reported in about 10% of patients with epilepsy compared with less than 0.5% in otherwise healthy individuals (Gregory et al 1993). It is twice as frequent in females as in males (Wolf and Goosses 1986). Children experience higher rate of photosensitivity, about 7.6% during 5-15 years of age (reviewed in Doose and Waltz 1993). The precipitating factors in PPR are flickering lights, such as sun shining through tree branches, stroboscopic disco lights, and coloured flashes or backgrounds on television or produced by video games on computer screens (Koepp et al 2016). There is evidence for genetic transmission of PPR, found in clinical studies of relatives of PPR-positive epilepsy and non-epilepsy subjects with an autosomal dominant mode of inheritance with age-dependent penetrance for photosensitivity. Existence of PPR independently and as a part of several autosomal recessive diseases makes it genetically heterogeneous (Stephani et al 2004). Linkage studies have identified loci conferring susceptibility to PPR at 5q35.3 (de Kovel et al 2010), 6p21.1 (Tauer et al 2005), 7q32 (Pinto et al 2005), 8q21.1 (de Kovel et al 2010), 13q31.3 (Tauer et al 2005) and 16p13.3 (Pinto et al 2005). Recently, *BRD2* has emerged as an associated candidate in photosensitive epilepsy (Yavuz et al 2012). No mutation, but many rare SNPs

have been found in genetic epilepsy patients exhibiting PPR in Turkey (Yavuz et al 2012). Evidences also point towards *CHD2* as novel gene for photosensitive epilepsy, with higher prevalence of unique variants among patients as compared to the control cohort (Taylor et al 2013). The seizures are typically generalized tonic-clonic (GTCS), sometimes seen with absences or partial motor seizures. The clinical response of the patients can range from subtle eyelid myoclonus to a GTCS convulsion.

PPR is not an exclusive epilepsy syndrome. The propensity of seizures induced by light is a genetically determined trait that may be asymptomatic throughout life or manifest with epileptic seizures. Genetic studies of PPR without epilepsy or other disorders has mostly been conducted in families with one PPR-positive epileptic member and indicates autosomal dominant inheritance. PPR with epilepsy is mostly seen in idiopathic generalized and partial, and in symptomatic epilepsies. It is highest in myoclonic epilepsy in infancy and childhood, myoclonic astatic epilepsy (MAE), childhood absence epilepsies (CAE), epilepsy with myoclonic absences (EMA), facial myoclonia with absences (FMA), self-induced photosensitive epilepsy, juvenile absence epilepsy (JAE), juvenile myoclonic epilepsy (JME), eyelid myoclonia with absences, and epilepsy with generalized tonic–clonic seizure (GTCS) on awakening. In these epilepsies, rates of photosensitive patients reach values up to 30% (Wolf and Goosses 1986). Children with neonatal seizures develop PPR in 44% cases (Doose et al 2002).

Precipitation of seizures in photosensitive patients depends on the activation of a critical neuronal mass in the occipital cortex (Guerrini et al 1995). Photosensitive individuals have an intrinsic hyperexcitability underlying the visual cortex, which can predispose to a large scale neuronal synchronization. In some cases like eyelid myoclonia or absences, frontal lobe origin is noted. Earlier, the prevailing view on human photosensitivity was that it is a generalized epilepsy form with the involvement of thalamocortical reticular system only. But now, the occipital origin of PPR is well accepted with evidence on primary role of visual cortex in pattern-induced and photic stimulation (Fisher et al 2005, Panayiotopoulos 2005). PPR induced by television and video games has highlighted the role of coloured stimuli, especially those in the deep red band, 680-700 nm (Harding 1998). This identified two distinct photosensitivity mechanisms: a quality-of-light that depends on changes in luminance, and a wavelength-dependent mechanism that might synergistically contribute to evoke a PPR. Mechanistically, it is unclear whether it is a defect in the excitatory or

inhibitory pathway or in both. Evidences suggest a heightened excitatory phenomenon rather than a failed inhibition. A dopamine receptor agonist, apomorphine is shown to block PPR in IGE patients but opiate antagonists like naloxone fail to do so suggesting the failure of dopaminergic system (Mervaala et al 1990). Interestingly, *Papio papio* baboons, the most prominent model for PPR, when induced by intermittent photic stimuli show EEG discharges with clonic and myoclonic jerks with motor cortex origin, in an otherwise healthy animal (Killam 1969, Silva-Barrat and Menini 1990). The large majority of patients with PPR do not need anticonvulsant therapy, but, when needed, the drug of choice is valproate. Stimulus avoidance and modification can be an effective treatment in some patients and can sometimes be combined with antiepileptic drug treatment (Verrotti et al 2004). Self-induced seizures are best handled with psychiatric or behavioural interventions.

1.3.2. Audiogenic seizures

Acoustic stimulus induced seizures are called audiogenic seizures. Mice, genetically susceptible to sound-induced reflex seizures are extensively used to study the genetic factors contributing to sound-induced convulsions. Rare in humans, this class of epilepsy is mostly seen in mice and rats exhibiting rapid whirl and abnormal behavior due to high pitched sound. Audiogenic seizure susceptibility, many a times is a consequence of alterations in noise levels, acoustic deprivation and kanamycin (Bac et al 1998), alcohol withdrawal (Faingold and Riaz 1995) and progesterone levels (Hom et al 1993, Voiculescu et al 1994). In the mid-1940-s, audiogenic epilepsy was noted in DBA/2J strain (Hall 1947) and was also described in different laboratory mouse strains both inbred (Fuller and Smith 1953) and outbred. The first observation of mouse audiogenic feature was made in Pavlov laboratory in apetite-conditioning experiments in mice. The bell used during food presentation unexpectedly induced seizures in some mice (Krushinsky et al 1970). The albino mouse strain Frings (selected by Frings and Frings, 1953) provided the first genetic evidence for audiogenic susceptibility, with a locus monogenic audiogenic seizure susceptible (*mass1*) mapping to chromosome 13 (Skradski et al 2001). The gene codes for *very large G-protein-coupled receptor-1 (vlgr1)* with a naturally occurring nonsense mutation V2250X, rendering the mice susceptible to seizures (McMillan and White 2004). In humans, *mass1* localizes to chromosome 5q14 (Skradski et al 2001). *hmass1* intriguingly localizes to a region close to two critical loci, familial febrile convulsions (*FEB4*) and Usher syndrome type 2c (*USH2C*). Another gene known in Black Swiss mice (Misawa et al 2002) is *jams1* (juvenile audiogenic monogenic seizures) mapping to chromosome 10. Large portions of this chromosomal

region are syntenic to human chromosome 19p13.3 region, which is responsible for a familial form of juvenile febrile convulsions (Puttagunta et al 2000). Knockout of certain genes in mice result in audiogenic seizure phenotype: *fmr1* (Chen and Toth 2001, Pacey et al 2009), *interleukin6* (De Luca et al 2004), 5HT2C receptor (Brennan et al 1997) and *Lgi1* (Boillot and Baulac 2015, Chabrol et al 2010). Triple-knockout mice of PAR bZip proteins, *Dbp*, *Hlf* and *Tef* (Gachon et al 2004) and double-knockout mice of *Gd3s* (GD3 synthase) and *GalNAcT* (beta-1, 4-N-acetylgalactosaminyltransferase) (Kawai et al 2001) exhibit susceptibility to audiogenic seizures. Certain classes of acquired audiogenic susceptibility include lesions of *substantia nigra*, ischemia, and dietary manipulations such as magnesium deficiency (Bac et al 1998). Recently, extracellular signal-regulated kinases 1 and 2 (ERK1/2) have been anticipated as one of the important contributors to audiogenic seizures in seizure-prone rats (Glazova et al 2015). It is supported by experiments using ERK1/2 inhibitor SL327 which blocks seizures. Selenoproteins were also implicated in causing severe neurological dysfunction leading to audiogenic seizures (Byrns et al 2014). Not only anomalies in "sensory part" of acoustic impulses in the brain pathway results in audiogenic seizure development, but abnormal biochemical and physiological status of central auditory and other structures are important issues (Garcia-Cairasco 2002). Audiogenic seizure fit initiates in the brain stem structure, namely inferior colliculli (Poletaeva et al 2011). Audiogenic epileptic EEG has been demonstrated in medulla (Krushinsky et al 1963, Krushinsky et al 1970), midbrain (N'Gouemo and Faingold 1999, Doretto et al 1994), lateral geniculate bodies (Krushinsky 1963, Ribak et al 1994), but not in hippocampus and neocortex (Merrill et al 2005, Moraes et al 2005, Semiokhina et al 2006).

1.3.3. Other reflex epilepsies/stimuli

Photosensitive and audiogenic epilepsies are comparatively well studied than other types of reflex epilepsies. Reading epilepsy manifests with jaw myoclonus (Matthews and Wright 1967, Radhakrishnan et al 1995) or with focal seizures of alexia. It is mostly triggered by reading aloud. No loci or genes are reported for this epilepsy, although a Mendelian mode of transmission of EEG pattern has been observed in a multigeneration family (Yacubian and Wolf 2015). Epilepsy provoked by eating is a rare condition. It is usually focal in origin and occurs in the setting of symptomatic epilepsies (Italiano et al 2016). *MECP2* duplication has been reported in two patients with this disorder (de Palma et al 2012, Martinez et al 2011). Musicogenic seizures usually occur in patients with focal symptomatic or cryptogenic temporal lobe epilepsy. Here, seizures are triggered by a specific type of music

or song. Although spontaneous seizures are documented, it is a complex phenomenon where seizures have a temporal lobe origin with limbic system distribution. Praxis induction of seizures is another interesting subset of reflex epilepsies where seizures are induced by cognition-guided higher mental activities. This type of epilepsy is idiopathic and generalized in nature characterized by myoclonic, absence or tonic-clonic seizures. The triggering factors can range from decision-making, playing chess, cards, or other games, calculation, manipulation, drawing, and writing (Raciot et al 2016). Somatosensory stimuli can trigger seizures in the form of startle responses. Cutaneous stimulation by rubbing, touching or tapping, brushing teeth, and even external ear stimulation can act as trigger for somatosensory epilepsy. Seizure semiology can range from myoclonic jerks to somatosensorial or motor seizures. Seizures induced by movements, or proprioceptive stimuli are evoked by passive or active movement without startle. The seizures are usually brief tonic seizures or simple partial attacks induced by movement of a limb and usually occur in subjects with fixed cerebral lesions and motor deficits. Reflex seizures should be distinguished from provocative precipitants in a certain situation or by general internal precipitants (such as emotional stress, fatigue, high body temperature, lack of sleep, specific stages of normal sleep, and menstrual cycle) or external precipitants (such as alcohol consumption, hyperventilation, or specific foods).

Table 1.1: Genes and loci known for reflex epilepsies

Reflex seizure type	Genes and loci identified
Photosensitivity/ photoparoxysmal response (PPR)	*CHD2*, *SCNM1*, 6p21, 7q32, 13q31, 16p13
Musicogenic/audiogenic	*LGI1*, *SCN1A*
Eating	*MECP2*

1.4. Hot water epilepsy

Epilepsy precipitated by the stimulus of contact with hot water for example during bathing is known as hot water epilepsy (HWE). Seizures triggered by contact with hot water have been variously described as water immersion epilepsy (Mofenson et al 1965), bathing epilepsy (Shaw et al 1988, Lenoir et al 1989), and hot water epilepsy (Satishchandra 2003). HWE was first described in 1945, in a 10-year-old boy from New Zealand who had developed a staring expression, stiffened extremities, and loss of consciousness every time he was given a bath (Allen 1945). Following this, several isolated cases were reported from Australia

(Keipert 1969), United States of America (Stensman and Ursing I971), Canada (Szymonowicz and Meloff 1978), United Kingdom (Moran et al 1976), and Japan (Kurata 1979, Miyao et al 1982, Morimoto et al 1985). A large number of patients with HWE have been reported from India, especially from the southern parts of the country (Mani et al 1968, Satishchandra et al 1985, Velmurugendran 1985). The first case of HWE reported in India was an 11-year old boy displaying troubled behaviour in the background of impaired consciousness during hot water bathing (Mani et al 1968). Following this, a series of 42 cases were reported with similar behavioural and clinical pattern while bathing in hot water (Mani et al 1968). The phenotype seemed like a unique type of reflex epilepsy and was substantiated by a large number of cases reported subsequently: 60 patients in 1972 (Mani et al 1972) and 108 cases in 1974 (Mani et al 1974). Satishchandra and colleagues (1988) published the largest series of HWE reports comprising 279 cases accounting for about 3.6-3.9% of epilepsy patients in south India (Satishchandra et al 1988). Apart from its prevalence in India, a series of 21 HWE patients have been reported from Turkey where it accounts for 0.6% of all epilepsies (Bebek et al 2001). HWE has been officially categorized as a reflex epilepsy type under the ILAE classification (Engel 2001).

1.4.1. Clinical features and seizure phenomenology

HWE is most prevalent in those parts of the country where hot water bathing following copious application of warm oil on the head, is a cultural custom. The bathing tradition in southern India involves filling of water in the bucket and pouring over the head using a mug. The water temperature ranges between 40-50°C. HWE has also been reported in isolated cases of tub and shower bathing (Satishchandra 2003). In HWE, the frequency of seizures is usually dependent on the frequency of head bathing. At a later stage in the natural history, 5-10% of these patients experience seizures even during body bath when water in not poured over the head (Satishchandra 2003). A study on 45 HWE patients from Karnataka reported that one-third of the patients display 1:1 relationship of seizure episodes to hot water bathing (Meghana et al 2012). The seizure semiology in these patients is typically complex partial seizures (CPS; now termed focal seizures with impaired cognition) with or without secondary generalization. In Indian HWE cases, 67-80% patients exhibit CPS (Satishchandra et al 1988, Mani et al 1974) while in Turkey, 95% of the cases presented with CPS (Bebek et al 2001). One-third of all the cases have primary generalized tonic-clonic seizures (GTCS). The seizures typically last for thirty seconds to three minutes and could manifest either at the beginning or end of the bath. Although there is no specific age group

for its onset, HWE is mostly seen in children. The mean age of seizure onset is 13±11 (mean±SD) years among Indian patients (Satishchandra et al 1988). A study of 21 HWE cases in Turkey identified the mean age of onset at 12 years (ranging from 1.5 years to 27 years) where 57% of the patients experienced their initial seizures in the first decade of their lives (Bebek et al 2001). The onset of HWE is marked by a confused look, sense of fear, visual and auditory hallucinations, irrelevant speech and complex automatisms (Satishchandra 2003). A number of HWE patients show a phenomenon of self-induction of seizures. A sense of intense desire and pleasure provokes these patients to continuously pour hot water over their head to induce seizure. One-third of HWE patients in Turkey reported 'self-induced' seizure (Bebek et al 2001) whereas in India about 10% of HWE patients showed this tendency (Satishchandra et al 1988). A contrary phenomenon of self abortion of seizures was reported in 12.7% of HWE cases who deliberately distracted themselves during an aura to abort a seizure event (Savitha et al 2007).

Interestingly, spontaneous non-reflex seizures (NRS) develop in some of the HWE patients later in their lifetime. Mani et al (1974) and Subrahmanyam et al (1972) reported 16-38% of such cases where NRS was observed after two years of HWE. Satishchandra et al (1988) reported development of NRS in 25% of patients. In a more recent study, 47.9% cases showed spontaneous NRS (Savitha et al 2007). A study from Japan reported 100% incidence of development of NRS (Kurata 1979). Among HWE patients from Turkey, Erdem et al (1992) reported 60% patients who developed generalized seizures and Bebek et al (2001) reported 62% patients in doing so. History of febrile convulsions (FCs), prior to development of HWE has been reported in various studies. From India, 11-27% of HWE patients reported history of FCs before development of HWE (Mani et al 1974, Subrahmanyam 1972, Satishchandra et al 1988). In Turkey, Bebek and colleagues (2001) observed 9.5% patients developing FCs.

1.4.2. Electroencephalography (EEG) features

In HWE patients, the interictal scalp EEG is usually normal and ictal EEG recordings are difficult to obtain due to technical limitations. Interictal EEG reveals no apparent cerebral lesions or characteristic EEG abnormalities, except in 15-20% patients who show diffused pattern (Mani et al 1972). Lateralized or localized spike discharges in the anterior temporal regions have been reported in few isolated cases (Syzmonowicz and Meloff 1978, Miyao et al 1982, Engel 2001). In a study of 21 cases from Turkey, interictal epileptogenic

abnormalities were documented in eight patients and normal EEG in another eight. In 40% of the patients, focal epileptogenic EEG abnormalities were mostly over the unilateral temporal regions (Bebek et al 2001). Resting EEG shows definite spikes, sharp waves, or spike and wave in 15% of the patients. In spite of the focal onset as clinical history, seizure discharges are predominantly generalized in nature (Bebek et al 2001). Interical EEG of 70 patients from India showed abnormal EEG in nine out of 45 HWE patients and six out of 25 patients having HWE with spontaneous seizure (Meghana et al 2012). Ictal EEG recordings obtained in few studies during provocation with hot water have demonstrated left temporal rhythmic delta activity (Stensman and Ursing 1971), sharp and slow waves in the left hemisphere (Parsonage et al 1976), bilateral spikes (Morimoto et al 1985), and temporal activity (Shaw et al 1988, Lenoir et al 1989). EEG recording in a patient (child) showed delta waves with a right hemisphere initial dominance, with rapid secondary generalization (Roos and van Dijk 1988). Another patient showed ictal EEG changes of increasing amplitude and decreasing frequency originating from the left centro-temporo-parietal electrodes (Lee et al 2000).

1.4.3. Neuroimaging

Neurological imaging studies in Indian HWE patients using computed tomography (CT) and magnetic resonance imaging (MRI) revealed no apparent structural abnormalities (Satishchandra et al 1988), whereas some cases indicate a temporal lobe or frontal lobe abnormality (Szymonowicz and Meloff 1978). An interictal and ictal single-photon emission (SPECT) scan in 10 HWE patients stimulated with hot water showed hypermetabolic tracer uptake in the medial temporal structures and hypothalamus in one hemisphere with spread to the opposite hemisphere. This demonstrated a functional involvement of these structures in prompting HWE (Satishchandra et al 2000). Recent advancements in imaging techniques have identified structural lesions in HWE patients; however, their involvement in the seizure mechanism is not well addressed. A case study of brain MRI on an HWE patient from Italy presented left parietal focal cortical dysplasia (Grosso et al 2004). A study from Turkey reported intercranial pathology like hippocampal sclerosis, dysplasia and a huge cystic lesion in their MRI study of five HWE patients (Tezer et al 2006).

1.4.4. Pathophysiology of HWE

Hot water epilepsy is an intriguing type of reflex epilepsy, pathogenesis of which remains largely unknown. There is variability to its pathophysiology. Typically, HWE in India is precipitated during bathing in hot water of 45-50°C poured over the head in quick successions that is sufficient to induce seizures (Satishchandra et al 1988). While the exact etiopathogenesis of this disorder is not known, there are several factors hypothesized to play a role which include genetic factors, environmental factors, consanguineous marriages and habit of taking bath with high temperature water (Satishchandra 2003). There is wide variation observed across populations in the age of onset, the region of contact with water, the water temperature, and the way of bathing, which triggers seizures. In most European countries, 'hot water' means warm water at a temperature around 37°C at which infants are usually bathed (Ceulemans et al 2008). In Turkey and India, however, temperature of hot water can go up to 50°C. Certain French patients experienced seizures with water at 37°C when immersed for water bath (Ioos et al 1999, 2000); another patient experienced seizures at temperature of 33°C which is lower than the core body temperature, only when water was poured over the head and/or face (Auvin et al 2006). In another report, seizures were experienced during water bath irrespective of the temperature, hot or cold (Seneviratne 2001). Some patients have a complicated 'triggering' procedure and some experience seizures even during body bath when water is not poured over the head. Triggering factors reported in isolation or as an accompaniment can include the heat of water, amount of water, duration of bathing, touching of water onto body or face, terminating the bath, bathing in one's own bathroom, hearing the sound of water, touching the water with hand or even having raindrops touch the face (Bebek et al 2001). A study in 71 HWE patients from India showed the variability in seizure triggering phenomenon. While all the 71 cases precipitated seizures upon hot water head bath, some of them had overlapping triggering phenomena. Eight of them experienced seizures also during washing face with hot water, 15, by pouring hot water over the body and 10, upon pouring cold water over their head (Savitha et al 2007). Although it is possible to trigger seizures under laboratory conditions by pouring hot water over the head of patients; hot water towels, sauna or blowing hot air on the head have usually been unsuccessful to induce seizures (Satishchandra et al 1988). This is suggestive of a complex triggering mechanism involving a combination of factors, like, contact of scalp with hot water, temperature of water, and specific cortical area of stimulation. Stensman and Ursing (1971) had suggested the involvement of complex tactile and temperature stimuli for precipitating such seizures. Tactile stimuli may indeed play a role in triggering seizures,

with temperature cues providing additive effect for somatosensory stimulation. This seems consistent with an interesting association of HWE with a common cultural custom of taking hot water bath following application of warm oil on the head, in southern parts of India (Mani et al 1972, Satishchandra et al 1988).

1.4.5. Differential diagnosis of HWE and their genetic etiology

Hot water epilepsy has overlapping features with a diversity of other conditions and a careful scrutiny is required to establish the diagnosis of the phenotype. Non-epileptic events of febrile seizures, startle reflex or vasovagal syncope confuse the diagnosis of HWE. However, with a thorough clinical history, it is easy to differentiate HWE. Unlike other reflex epilepsies, HWE has a longer latency period. Febrile seizures, one of the common non-epileptic events that occur among infants or children between six months to six years of age are associated with fever (Leviton and Cowan 1982). Febrile convulsions are sometimes seen in association with HWE (around 11 to 27% of Indian patients, as reported by Satishchandra 2003), but there is no causal link between the two, known so far. Genetic loci for febrile convulsions are known to map to 2q23-q24, 5q14-q15, 6q22-q24, 8q13-q21, 18p11.2, 19p13 (Audenaert et al 2006 for review); 5q34 (Audenaert et al 2006), 21q22 (Hedera et al 2006), 3p24.2-p23 (Nabbout et al 2007), 3q26.2-q26.33 (Dai et al 2008), 8q12.1-q13.2 (Salzmann et al 2012), indicating a significant genetic contribution to its etiology. In addition, at least two other human epilepsies, generalized epilepsy with febrile seizure plus (GEFS+) and severe myoclonic epilepsy in infancy (SMEI), are induced by high body temperature. In GEFS+, febrile seizures continue beyond six years and patients develop afebrile generalized seizures (Scheffer and Berkovic 1997). Genetic loci for GEFS+ map at 2q24.3 (Peiffer et al 1999, Singh et al 1999), 1p36 (Singh et al 1999), 2p24 (Audenaert et al 2005); 19q13, 2q24-q23, 5q34 (Audenaert et al 2006 for review); 8p23-p21 (Baulac et al 2008), 6q16.3-q22.31 (Poduri et al 2009), 12p13.33 (Morar et al 2011), 22q13.31 (Belhedi et al 2013), 16p11.2 (Schubert et al 2014), suggesting an appreciable genetic basis to the disorder. Mutations in *SCN1A*, *SCN1B*, *SCN2A*, *SCN9A*, *GABRD*, *GABRG2* (Piro et al 2011 for review), *PRRT2* (He et al 2014) have been reported in GEFS+ patients. Hot water induced seizures are also observed in an intractable form of epilepsy called severe myoclonic epilepsy in infancy (SMEI) or Dravet syndrome, mainly characterized by febrile convulsions before the age of one year, hemiconvulsions, partial seizures and myoclonic seizures (Dravet 1992). SMEI mainly arises due to *de novo* mutations in *SCN1A* (Selmer et al 2009) and *PCDH19* (De Jonghe 2011). Although hot

water-induced seizures coexist with FCs, SMEI and GEFS+, a thorough clinical investigation, mainly through questioning the patients, can distinguish HWE from other events.

1.4.6. Animal models of HWE: Kindling effects

Kindling involves a gradual and progressive increase in the neuronal activity upon repeated focal stimulation with electrical (Goddard et al 1969) and chemical stimuli (Wasterlain and Jonec 1983). The kindling effect was first described by Goddard in 1969 and since then has become an ideal experiment model for epilepsy. He showed that repeated exposure of constant non-polarizing electrical stimulus, just large enough to trigger a burst of epileptiform activity, can lead to fully generalized behavioural convulsions. Pathologies of direct damage, edema and metal toxicity as cause of kindling were ruled out; demonstrating seizures were result of repeated stimulation. Klauenberg and Sparber (1984) observed similar effect of kindling in rats by immersing them in 20 cm deep water at 45°C for 4 mins; which they called as 'hyperthermic kindling'. A phenomenon of hyperthermic kindling as a mechanism of HWE had been postulated (Satishchandra et al 1985; Satishchandra et al 1988). Histopathological evidence of hippocampal sprouting following hyperthermia induced stimulation was observed in rats (Jiang et al 1999). Rats exposed to hot water jet on the head (50-55°C) for a period of 8-10 minutes showed the hippocampus to be sensitive to hyperthermia and capable of producing seizure like discharges with the slight increase in water temperature (Ullal et al 1996). Depth electrode recording from the hippocampus revealed seizure discharges during the ictus, followed by low voltage indeterminate activity and a quiescent resting phase. In the susceptible animals, seizures occurred at a critical rectal temperature of 41.5°C (normal being 37°C) and a brain hippocampal temperature of 37°C (normal being 35.5°C). Intervention of hyperthermia by cooling after the ictus prevented subsequent seizure activity. However, no distinct evidence for kindling phenomenon had been observed in the study (Ullal et al 1996). In a later study of similar experimental pattern, histological examinations of brain showed anoxic changes in some neurons of hippocampus, brain stem and cerebellum in rats exposed to hot water jets. Further, a correlation in the degree of mossy fiber sprouting and number of stimulation with hot water was demonstrated (Ullal et al 2006).

1.4.7. Genetics of HWE

While the first case of HWE was reported in 1945 (Allen), the genetic underpinnings of this disorder has remained largely unexplored. In 1988, study of 279 HWE patients from India found 20 patients with positive family histories for HWE, and 20 patients with past histories of febrile convulsions (Satishchandra et al 1988). Positive family histories have been reported for 7-15% of the patients from India (Mani et al 1974, Satishchandra et al 1988) and in 10% of the patients from Turkey (Bebek et al 2001). In a French-Canadian family with X-linked focal epilepsy and reflex bathing seizures, mutation in the *synapsin I* gene has been reported. It was however debatable if these patients exhibited typical HWE or a distinct syndrome (Nguyen et al 2015). Studies involving two Indian HWE families with several of their members affected with the disorder, Ratnapriya and colleagues have reported HWE genetic loci at chromosomes 4q24-q28 (Ratnapriya et al 2009a) and 10q21.3-q22.3 (Ratnapriya et al 2009b). A loss of function mutation in *GPR56* was identified in a 5-year old Portuguese boy with bilateral frontoparietal polymicrogyria (BFPP) who presented with hot water epilepsy (Santos-Silva et al 2015). More recently, a study in a large HWE Indian family identified a locus at 9p24.3-p23 with rare variants in the gene, *SLC1A1* (Karan et al 2017).

1.5. Aims and scope of my thesis work

In this thesis, I present molecular genetic analysis of hot water epilepsy (HWE) in three multi-generation and multi-affected families from south India. I have employed genome-wide linkage analysis and whole exome sequencing analysis to identify the potential genes for hot water epilepsy among these families. The main objectives of my study are as follows:

(i) Genetic analysis of the 4q24-q28 locus and exploring the cell biological and functional aspects of *ZGRF1*, a proposed gene for the disorder (Chapter 2)

(ii) Genetic analysis of the 10q21.3-q22.3 locus and identification of potentially causative gene (Chapter 3)

(iii) Identification of potential candidate gene(s) in a family HWE307, with several of its members affected with the disorder (Chapter 4)

Towards the end, summary, limitations, and future perspective on this research work are discussed.

Chapter 2

Gene identification at hot water epilepsy locus at chromosome 4q24-q28

Summary

In this chapter, I describe identification of the potentially causative gene for hot water epilepsy transmitted in an autosomal dominant manner in a multiple-generation, multi-affected family, HWE227. A genome-wide linkage analysis of this four-generation family had provided parametric two-point LOD score of 3.50 at θ=0 for the marker D4S402, at chromosome 4q24-q28 (Ratnapriya et al 2009a). The critical genomic region was localized between D4S1572 and D4S2277 encompassing about 24 megabases of the genome sequence length. To examine the genes in this interval, whole exome sequencing was performed on an affected parent-offspring pair. Among 563 variants obtained, one rare variant c.1805C>T in *ZGRF1* (Zinc Finger GRF-Type Containing 1) was observed to co-segregate with the clinical phenotype in HWE227. ZGRF1 is an uncharacterized protein. The c.1805C>T transition changes a relatively conserved threonine residue to isoleucine (p.Thr602Ile). The gene was examined in 288 apparently unrelated individuals with hot water epilepsy, and five additional rare variants, c.977G>A (p.Arg326Gln), c.1979A>G (p.Glu660Gly), c.5584C>T (p.Arg1862*), c.5818T>C (p.Phe1940Leu) and c.5951A>G (p.Asp1984Gly) were identified. ZGRF1 is expressed in different regions of the human brain. Experiments in cultured HEK293 cells found ZGRF1 to localize in cell cytoplasm, nucleolus, centrosome, midbody and spindle poles across stages of cell division. The protein co-localizes with processing bodies (P-bodies) under conditions of stress. Upf1 (Up-frameshift 1), involved in nonsense-mediated mRNA decay (NMD) is the nearest known paralog of ZGRF1 and the two proteins co-localize across the different mitotic stages. *In vitro* functional consequences of the ZGRF1 variants suggested spindle pole defects during cell division and partially disrupted UPF1 localization at spindles. Additionally, a distinct cellular response of nuclear localization of ZGRF1 was observed under mitomycin C- induced DNA damage. This study suggests *ZGRF1* as a potential causative gene for hot water epilepsy with a possible involvement in nonsense-mediated mRNA decay, abnormal mitotic division, and DNA damage and repair pathways.

Previous studies involving whole-genome based linkage mapping in multi-generation families from south India with several of its members affected with hot water epilepsy have identified a novel genetic locus for hot water epilepsy at chromosome 4q24-q28. In this chapter, I present studies aimed at identifying potentially causative gene for the disorder at the 4q24-q28 locus.

2.1. Subjects and methods

2.1.1. Family ascertainment and clinical evaluation

A four-generation family with hot water epilepsy, HWE227 (Figure 2.1) was ascertained from the Department of Neurology, National Institute of Mental Health and Neurosciences (NIMHANS), Bangalore. It is a family of 28 individuals from Mandya District in Karnataka, India with 10 affected, 3 clinically asymptomatic (with at least one child manifesting HWE), and 15 apparently unaffected members. HWE in Family 227 transmitted in an apparently autosomal dominant mode with incomplete penetrance. This family was obtained through the proband III:7, diagnosed with hot water epilepsy at the age of 20 years. Since the age of 11, he had a history of seizures triggered by hot water head bathing, around once every two months. One year prior to examination at NIMHANS, seizures precipitated every time he took hot water head bath and during 40% of these episodes he lost consciousness with tonic clonic limb reflexes. He had no previous history of febrile convulsions or any other type of seizure. His seizure episodes were characterized by the usual features of hot water epilepsy, whereby, after pouring 5-6 mugs of hot water on the head, he experienced a sudden fear associated with lip smacking and dazed look with automatism lasting for around 45 seconds. The patient had no memory of the events beyond his initial sense of fear. He was overall intellectually sound, with no other neurological anomaly detected during clinical evaluation. His inter-ictal EEG showed runs of slow waves in the low theta range with sharp waves arising in the left anterior temporal region. Magnetic resonance imaging of the brain detected no hippocampal abnormality. He was treated with intermittent prophylactic clobazam therapy. All affected individuals satisfied the clinical criteria of seizure precipitation upon pouring hot water over their head at about 42-46°C. Information regarding age-at-onset, duration and type of seizure, time of seizure onset in relation to hot water bath, history of febrile seizures, precipitation of non-reflex seizure in HWE patients or other neurological disorders and family histories of epilepsy were obtained from patients and their surrogate respondents who had witnessed a number of

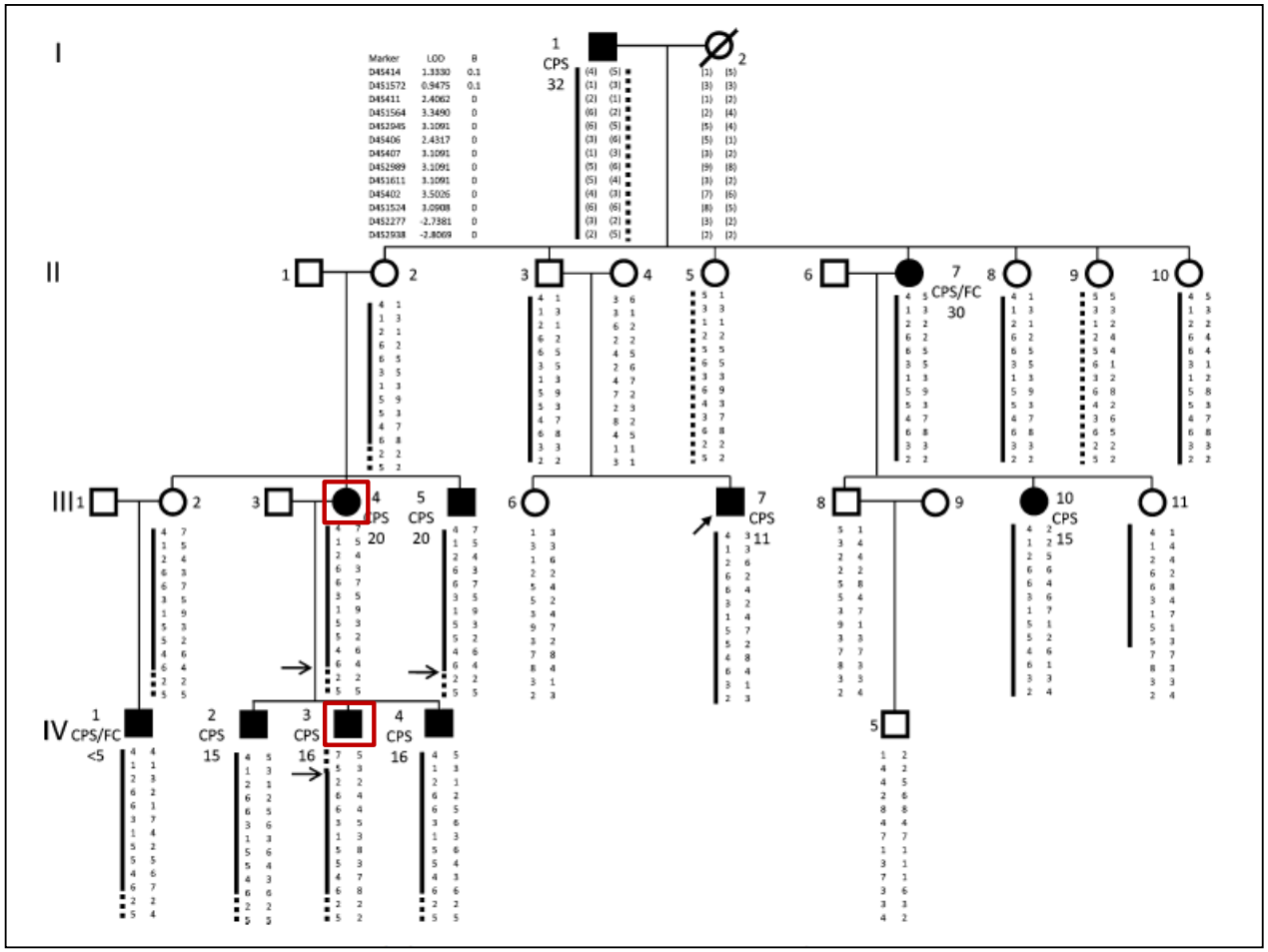


Figure 2.1: Pedigree chart of family HWE227. A blackened symbol represents a clinically affected individual and an empty symbol represents an unaffected individual. The proband III:7 is represented with an arrow. The Roman numbers to the left of the pedigree denote generations and Arabic numbers beside the symbols denote individuals. HWE-linked haplotype is represented as solid lines. Alleles in parentheses are inferred from the genotypes of other relatives. Key recombinant events in III:4, III:5 and IV:3, are indicated by arrows. Markers in the haplotype are in order from centromere to telomere and highest LOD score (Zmax) obtained for each marker is shown. Seizure types and age at onset are indicated beside symbols. CPS = complex partial seizures, FC = febrile convulsions. Individuals *boxed in red* were exome sequenced. (Modified from Ratnapriya et al 2009a).

such episodes in the affected individuals. The remaining members of the family were reported to be normal and were considered unaffected.

For genetic studies, 10 millilitres of peripheral blood sample was collected from all the participating members of HWE227. DNA was extracted from the white blood cells using a standard phenol chloroform method (Sambrook and Russell 2001, Appendix IVA). Written consent was obtained from all participating individuals. The study was approved by the institutional human bioethics committees.

2.1.2. Ascertainment of HWE and control cohorts

A cohort of about 288 apparently unrelated hot water epilepsy patients from south India was ascertained from NIMHANS, Bangalore. Clinical diagnoses of these patients were established according to ILAE guidelines for epilepsy classification (ILAE 2001). All patients satisfied the clinical criteria for hot water epilepsy. The control cohort comprised 288 healthy individuals from south India, also ascertained from NIMHANS, Bangalore. None of them had any reported history of neurological disorders. DNA was isolated from white blood cells from 10 millilitres of peripheral blood sample using a standard phenol chloroform method.

2.1.3. Whole-exome sequencing

As a first step towards searching for potential genes in HWE227, eight genes from the critical region were examined by Sanger-based sequencing. However, to examine a large number of genes in the region, whole exome sequencing experiment was undertaken. The whole exome sequencing experiment was performed on an affected parent-offspring pair, III:4 and IV:3 from HWE227 using SureSelect Human All Exon 50Mb Kit (Agilent Technologies, USA, Tables 2.1 and 2.2). This spans about 50Mb (1.5%) of human genome corresponding to the Consensus Coding Sequence Database (CCDS); comprising all protein coding exons annotated by the GENCODE project, with a set of curated protein-coding regions that are of high quality and consistently annotated, with 10bp flanking sequences on either side. It also includes exons of small non-coding RNAs from miRBase and Rfam. About 5µg of genomic DNA from individuals III:4 and IV:3 were fragmented to sizes between 100-400 bases by sonication (Bioruptor, Diagenode, Belgium), purified using Agencourt AMPure XP beads (Beckman Coulter, USA) and analyzed for size distribution on Agilent High Sensitivity Bioanalyzer Chip (Agilent Technologies). Subsequently, genomic DNA libraries were prepared according to Agilent's SureSelect Target Enrichment System

for Illumina Paired-End Sequencing Library v1.2 (Illumina Inc., USA). The DNA was subjected to a series of enzymatic reactions to repair the frayed ends; phosphorylated at 5'-end and adenylated at 3'-end according to a standard protocol; and then, ligated with 60bp adapter sequences (Illumina Inc.) to the 3'- adenylated ends (Illumina's protocols). Ligated products of the size range 150-300 bases were purified by gel excision with QIAquick purification reagents (Qiagen, Germany). The prepared library was enriched for adapter-ligated fragments through a PCR step and concentrated on a vacuum concentrator (Eppendorf, Germany) to ~500ng. The library was hybridized to SureSelect biotinylated RNA baits complementary to the exome targets at 65°C for 48 hours. The hybridized library fragments were isolated by magnetic capture using Dynabeads M-280 Streptavidin (Invitrogen, USA), followed by purification of the captured library-bead solution with AMPure XP beads and eluted in DNase-free water at 95°C (Agilent's guidelines). The captured library was enriched by PCR using adapter complimentary primers and the reaction products were purified with AMPure XP beads. The quality of the libraries was verified on Agilent High Sensitivity Bioanalyzer Chip and the concentration was measured using a spectrophotometer (NanoDrop Technologies, USA). Pre- and post-capture efficiencies of samples were assessed using Brilliant II SYBR Green QPCR Master Mix (Agilent Technologies) on Mx3005P Real-Time PCR System (Stratagene, USA). The library was used for cluster generation by bridge amplification on cBOT (Illumina Inc.) and sequenced with 100bp paired-end reads on Illumina Genome Analyzer (GAIIx).

Table 2.1: Targeted exome by Agilent 50Mb kit (Clark et al 2011)

Features	**Bases targeted by Agilent 50Mb kit**
Total CDS target size (whole genome)	51542882
RefSeq Coding	32326914
RefSeq UTR	3920825
Ensembl CDS	33472589
Ensembl all exons	38123201
miRBase	55249

Table 2.2: Attributes for whole exome capture of 4q24-q28

Reagent	Agilent SureSelect Human All Exon 50Mb kit
Capture design	
Probe length (bases)	100

Type of sequencing	Paired end
Targets	
Whole-exome	
CDS target size (bases)	39515328
Probes target size (bases)	51542882
Exome targeted at the 4q24-q28 interval	
4q24-q28 target size (bases)	35795000
CDS target size (bases)	275107
Probes target size	338579

2.1.4. Whole exome sequencing analysis

The whole exome sequencing reads were aligned to reference human genome (hg19/GRCh37) using BWA-0.6.0 (Li and Durbin 2009). Alignment was performed with high-quality reads having at least 70% of the sequence with a minimum Phred score of 20 obtained by quality control program SeqQC-V2.0. Duplicate reads particularly resulting from PCR artifacts, which would otherwise potentially inflate the observed sequencing depth, were removed using SAMtools v0.1.7a (Li et al 2009) and an *in-house* script. Variants were called by SAMtools at a Phred like SNP quality score of 20. The variants identified were annotated against dbSNP131 and dbSNP135 databases using *in-house* scripts. Variants determined were further validated against dbSNP144, Ensembl (http://www.ensembl.org/index.html), 1000 Genomes (http://www.1000 genomes.org/), Exome Variant Server (http://evs.gs.washington.edu/EVS/), and The Exome Aggregation Consortium (ExAC) (http://exac.broadinstitute.org/) datasets and their control frequencies were determined. The base and read statistics were obtained through SeqQC-V2.0. Alignment statistics were obtained from BEDtools v-2.12.0 (Quinlan and Hall 2010). Variants with minor allele frequency (MAF) $\leq$ 0.005 obtained from the analysis were ensured to be covered by at least 5x of read depth and preferably called from both the strands of DNA. In order to solicit probable variants that had gotten missed at low coverage, variants up to 2x read depth were manually analyzed on SAMtools and compared in the two whole exome sequenced (WES) individuals. The coding exons, 5'- and 3'-UTRs and intronic flanks up to 100bp for all the genes in the 4q24-q28 locus were manually examined to curtail the probability of missing out on new or rare variants. Any read with an expected variant in either or both the samples were taken into account. Any exons/regions within the span of the 4q24-q28 interval not covered by probes were manually sequenced by Sanger-based method (Appendix A2.1 and A2.2).

2.1.5. Sanger based variant validation and genetic analysis

Variants with a minor allele frequency (MAF) ≤ 0.005 obtained from the whole exome sequencing data were validated by Sanger sequencing. Gene sequences were obtained from Human Genome Mapviewer database (http://www.ncbi.nlm.nih.gov/mapview/, build 37.3, NCBI). The primer pairs spanning the variant-carrying exons/regions were designed using Primer3 Input (http://bioinfo.ut.ee/primer3-0.4.0/; Appendix A2.3) and OligoCalc: Oligonucleotide Properties Calculator to check for primer self complementarity (http://biotools.nubic.northwestern.edu/OligoCalc.html). Variants that were positive in both the exome sequenced individuals III:4 and IV:3 examined by single strand sequencing were confirmed by bi-directional sequencing. PCR amplification was carried out with 100ng of genomic DNA added to a cocktail of 1X reaction buffer containing 50mM KCl and 20mM tris-HCl (pH 8.4), 1.5mM $MgCl_2$, 800µM dNTPs, 0.25µM each of forward and reverse primers and 1U of *Taq* Polymerase (New England Biolabs, USA) in a 20µl reaction volume made up with deionised water. Amplification was carried out on a GeneAmp PCR9700 (Applied Biosystems) under the following conditions: initial denaturation at 94°C for 5 minutes, followed by 40 cycles of denaturation at 94°C for 30 seconds, annealing at 55-60°C for 30 seconds and extension at 72°C for 30 seconds, followed by final extension at 72°C for 10 minutes. The amplified products were purified by Multiscreen $PCR_{\mu 96}$ filter plate (EMD Millipore, USA). DNA sequencing was performed by cycle sequencing of the purified products in a solution of 3.5µl of 5x Sequencing Buffer (Applied Biosystems) containing 80mM Tris-HCl (pH 9.0) and 2mM $MgCl_2$, 1µl of ABI PRISM BigDye Terminator v3.1 Ready Reaction Mix (Applied Biosystems) containing fluorescent-labelled dideoxy-terminators, 0.04µM of forward or reverse primer, and 50-100ng of purified PCR product in a 20µl reaction volume made up with deionised water. The cycle sequencing was performed under the following conditions: initial denaturaion at 96°C for 1 minute, 25 cycles of denaturation at 96°C for 10 seconds, annealing at 50°C for 5 seconds, extension at 60°C for 4 minutes. The products were precipitated in 90% ethanol, washed with 70% ethanol and the pellet was dissolved in 10µl Hi-Di Formamide (Applied Biosystems). This was denatured at 96°C for 5 minutes, snap chilled in ice and electrophoresed on an automated 48-capilary ABI3730 DNA Analyzer (Applied Biosystems). Each amplicon was sequenced with respective primers and analyzed against their reference sequences (RefSeq database, NCBI) using SeqMan II 5.01 software (DNASTAR Inc., USA). The variants confirmed for co-occurrence in the two exome sequenced samples were examined for a complete family-based segregation to ascertain their co-segregation with the phenotype across generations.

Co-segregating variants were examined in an ethnically-matched control set of at least 288 normal individuals.

2.1.6. Screening for additional alleles in *ZGRF1* in HWE patients and control individuals

To gather further genetic evidence regarding the involvement of *ZGRF1* in hot water epilepsy, a screen for additional variants was performed. The gene was sequenced in 288 apparently unrelated probands. Twenty seven coding exons of the gene and their flanking intronic regions covering the splice junctions, and 5'- and 3'- untranslated regions (UTRs) were sequenced with specific primers designed using Primer3 Input (Koressaar and Remm 2007) (Appendix A2.4). *ZGRF1* was examined in 288 healthy (control) individuals from the southern parts of India. The gene was sequenced for all the regions as was done for the HWE patients.

2.1.7. Bioinformatic analysis and identification of ZGRF1 paralogs

The nucleotide and protein sequences for *ZGRF1* across different species were obtained from NCBI HomoloGene database (http://www.ncbi.nlm.nih.gov/homologene/). Multiple species sequence alignments for conservation analysis was carried out by Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/). DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST; Boratyn et al 2012) and pBLAST (protein-protein BLAST) were used to determine the domains of ZGRF1 and identify its paralogs and orthologs. Phylogenetic tree and cladogram were constructed using Simple Phylogeny (http://www.ebi.ac.uk/Tools/phylogeny/simple_phylogeny/). Domain-wise identities and similarities between ZGRF1 and its nearest paralog UPF1 were determined by pairwise alignment using EMBOSS Needle (http://www.ebi.ac.uk/Tools/psa/emboss_needle/).

2.1.8. Expression of *ZGRF1* in human brain regions

The presence of *ZGRF1* transcripts in different regions of the human brain was examined using reverse-transcriptase PCR (RT-PCR) in Marathon-Ready™ full-length brain cDNAs from cerebral cortex, cerebellum, hippocampus, hypothalamus and whole brain (Clontech, USA). Two primer sets were used to amplify cDNA from the N-terminal and C-terminal regions of *ZGRF1*. This was to ensure that the full length transcript (corresponding to the largest isoform) can be detected, if transcribed. The N-terminal primer pair 5'-TCCAAAGACACAGAAGCACA-3' and 5'-ACCTGCAAGAAGTCAATCTGC-3' located within exon 5 and the junction of exons 9 and 10 would amplify a product of 656 bases. The

C-terminal primer pair 5'-TCAGCCTAGGAGCAACATTGA-3' and 5'-TCCAACTACTCGAACCTGCT-3' located within exon 18 and junction of exons 21 and 22 would amplify a product of 572 bases. PCR was performed in a 20µl reaction volume containing 0.5ng of Marathon-Ready™ cDNA, 0.25µM of each primer, 800µM of dNTPs (200µM of each dNTP), 1.5mM of $MgCl_2$, 1X reaction buffer (50mM KCl, 20mM Tris-HCl pH 8.4) and 1U of *Taq* polymerase. PCR amplification was performed on GeneAmp® 9700 at the following conditions: initial denaturation (94°C, 5minutes), 40 cycles of denaturation (94°C, 30 seconds), annealing (60°C, 30 seconds) and extension (72°C, 30 seconds), followed by a final extension at 72°C for 10 minutes. The amplified products were verified with 1% agarose/TAE/EtBr gel electrophoresis. The products were gel purified using QIAprep gel extraction kit (Qiagen) and subjected to bi-directional DNA sequencing with the respective primer pairs on an ABI3730 DNA Analyzer.

2.1.9. Cloning and site-directed mutagenesis

The wild-type cDNA transcript of *ZGRF1* (NM_018392.4, NP_060862) cloned into mammalian pCMV6- entry vector with C-terminal Myc-DDK tag (RC229463) was obtained from Origene, USA. The identity of the clone was verified by Sanger-based sequencing with *ZGRF1* cDNA specific overlapping primers (Appendix A2.5). The open reading frame of the wild-type ZGRF1 was subcloned into pCMV6-AC-GFP with a C-terminal turbo GFP tag (Origene). The clone RC229463 was restriction digested with enzymes SgfI and MluI and the released insert was verified on 0.8% agarose/TAE/EtBr gel electrophoresis. The pCMV6-AC-GFP vector was digested with SgfI/MluI. The insert and vector, at a ratio of 1:3, were ligated overnight at 16°C in T4 DNA ligase buffer with T4 DNA ligase (New England Biolabs) along with a vector-only control ligation. 5µl of these ligation products was transformed into XL10-Gold competent cells (Agilent Technologies, Appendix IVC). Bacterial colonies obtained were screened by colony PCR. A small number of the positive colonies were grown for plasmid isolation using alkaline lysis method. The plasmids were insert released. Positive insert sizes were sequence verified by Sanger-based sequencing and the clone with the correct insert was prepared using QIAprep Spin Miniprep columns (Qiagen). Six variants were introduced in the wild-type *ZGRF1* cDNA with specific primers (Appendix A2.6) using QuikChange II XL Site-Directed Mutagenesis kit (Agilent Technologies). Plasmids carrying the *ZGRF1* variants were purified using QIAprep Spin Miniprep columns and the clones with the specific variants incorporated, were confirmed by Sanger sequencing.

2.1.10. Cell culture and transfections

To study endogenous protein expressions, human HEK293 (ATCC, USA), human SH-SY5Y neuroblastoma (ATCC) and rat C6 glioma (ECACC, Sigma-Aldrich, USA) cells were grown to about 80% confluence in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% heat-inactivated fetal bovine serum (FBS), 2mM L-glutamine, 100U/ml penicillin and 0.1mg/ml streptomycin (Sigma-Aldrich), in a humidified atmosphere of 5% CO_2 at 37°C. For protein overexpression study, all transfections were carried out in HEK293 cells grown up to 40% confluence under same conditions. Cells were grown either in 6-well plates or seeded on sterilized cover-slips coated with poly-L-lysine (0.1mg/ml, Sigma-Aldrich) in 35-mm dishes (BD Biosciences, USA) for Western blotting and microscopy-based experiments respectively. An hour before transfection, the culture medium was replaced with antibiotic- and serum- free DMEM. Mammalian cells were transfected with ZGRF1-wildtype cDNA, six ZGRF1-mutant cDNA with the incorporated coding variants, and vector control using calcium phosphate transfection (Appendix IVB). The transfection mix comprised of an equal volume of $CaCl_2$-DNA solution and 2x HBS (HEPES buffered saline) with a final concentration of calcium in the culture dish being 12.5mM. Six hours after transfection, the culture medium was replaced with complete DMEM supplemented with 10% FBS. Forty eight hours post-transfection, the cells were processed accordingly for immunoblotting or microscopy.

2.1.11. Immunoblotting

Whole cell protein lysates were prepared from HEK293 cells grown in 35mm dishes, 48 hours post-transfection. Untransfected cells and cells overexpressing the ZGRF1- wild- type and -mutant proteins were lysed in 200µl of RIPA (Radio ImmunoPrecipitation Assay) lysis buffer (50mM Tris-HCl pH 7.4, 150mM NaCl, 1mM EDTA, 0.5% sodium deoxycholate, 1% NP-40 and 0.1% SDS) supplemented with protease inhibitor cocktail at 1:1000 dilution (Sigma-Aldrich). Cells were lysed for three hours at 4°C in a rotating mixer. Protein lysate was obtained by spinning down lysed cells at 14,000g for 20 minutes at 4°C. Total protein concentration was quantified using bicinchoninic acid (BCA) assay (Sigma-Aldrich). 50µg of the total protein lysate mixed in 3µl of 6x SDS gel loading dye was loaded in each well of an 8% polyacrylamide gel (Laemmli 1970). Proteins from the gel were electro-transferred using a wet electroblotting system (Bio-Rad Laboratories, USA) to 0.45µm PVDF membrane (Millipore, USA) at a constant voltage of 100V for one hour in transfer buffer (25mM Tris-HCl, 192mM glycine, 20% methanol and 0.036% SDS) at 4°C. The membrane

was blocked in 5% skimmed milk (HiMedia, France) for 4 hours at 4°C followed by incubation in rabbit polyclonal anti-ZGRF1 antibody at 0.46µg/ml or 1:300 dilution (ab122126, Abcam, UK) in 1% BSA in 1x PBS (phosphate buffered saline) overnight at 4°C. The blot was washed gently 3 times in 1x PBS, for 5 minutes each, at room temperature. It was incubated for 1 hour at 25°C with a 1:10000 dilution of horseradish peroxidase (HRP)-conjugated, goat anti-rabbit IgG (Bangalore GeNei, India) in 1% BSA in 1x PBS. Post probing in secondary antibody, the blot was washed 4 times with 1x PBS containing 0.1% Tween-20, for 10 minutes each at room temperature. The protein bands were detected using an enhanced chemiluminiscent substrate for HRP (West Pico, Pierce, USA). γ-tubulin (T5326, Sigma-Aldrich) was used as a loading control, probed on the same blot. Brain lysates (50µg) from different regions of the brain: amygdala, basal ganglia, cerebellum, hippocampus, hypothalamus and temporal lobe were loaded on a 5-15% gradient gel, transferred on to a 0.45µm nitrocellulose membrane (GE Healthcare Life Sciences, USA) at constant voltage of 23V for one hour and thirty minutes using a semi-dry transfer system (Bio-Rad Laboratories), and immunoblotted using the same protocol as mentioned above.

2.1.12. Immunocytochemistry

Human HEK293, human SH-SY5Y neuroblastoma and rat C6 glioma cells were grown on poly-L-lysine coated coverslips in 10% DMEM up to 80% confluence. For determining the endogenous protein profile, the growth medium was removed, cells were washed twice at 25°C with 1x PBS for 5 minutes each. They were fixed with 2% PFA (paraformaldehyde) for 15 minutes, washed twice with 1x PBS for 5 minutes each; for permeabilization, treated with 0.1% Triton-X in PBS for 10 minutes at 25°C and blocked with 5% BSA and for four hours at 4°C. Cells were incubated with rabbit polyclonal anti-ZGRF1 antibody (ab122126, Abcam, UK) diluted to 1:100 in 1% BSA in PBS, overnight at 4°C. They were gently washed thrice at 25°C with PBS for 5 minutes each and incubated with Alexa Fluor 568-conjugated secondary antibody (Molecular probes, Invitrogen) diluted to 1:500 in 1% BSA in PBS for four hours at 4°C. The cells were washed with PBS four times for 10 minutes each and counterstained with DAPI (Sigma-Aldrich) at 1µg/ml. The coverslips were mounted in 70% glycerol (Merck Millipore, USA) on glass slides (Bluestar, India) and imaged on a LSM 880 Meta confocal laser scanning microscope (Carl Zeiss, Germany) under a 63x/1.4 oil immersion objective. Localizations of ZGRF1-wild type (WT) and -mutant proteins were examined in HEK293 cells transfected with ZGRF1-WT- and ZGRF1-

Arg326QGln-, Thr602Ile-, Glu660Gly-, Arg1862*-, Phe1940Leu- and Asp1984Gly- Myc-DDK plasmids. Transfected cells were processed as mentioned above with an anti-ZGRF1 antibody dilution of 1:1000. For dual antibody staining procedures, primary antibodies were mixed together in 1% BSA in their respective dilutions. Primary antibodies, rabbit polyclonal anti-ZGRF1 and mouse monoclonal anti-γ-tubulin (1:5000, T5326, Sigma-Aldrich) were co-stained at 4°C overnight. Anti-rabbit Alexa Fluor 568 (Molecular Probes, Invitrogen), anti-mouse Alexa Fluor 568 (Molecular probes, Invitrogen) and anti-rabbit Alexa Fluor 488 (Molecular probes, Invitrogen)-conjugated secondary antibodies were used in the same dilution (1:500) at 4°C overnight in co-localization experiments. For detecting γ-tubulin, cells were fixed with 100% methanol, else with 2% PFA in all other instances. For dual staining with anti-ZGRF1 antibody and rabbit monoclonal anti-UPF1 Alexa Fluor 488 (1:100, ab201761, Abcam), cells were sequentially stained; first with anti-ZGRF1 antibody and anti-rabbit Alexa Fluor 568 according to the above protocol, followed by incubation in anti-UPF1 Alexa Fluor 488 antibody overnight at 4°C.

2.1.13. Cell cycle synchronization and immunostaining

To achieve cell cycle synchronization, untransfected and transfected HEK293 cells for ZGRF1-WT- and the six mutants- Myc-DDK, were grown to 80% confluence and incubated in culture medium containing 2mM thymidine (Sigma-Aldrich) for 14 hours. To achieve mitotic enrichment, cells were washed thrice in plain DMEM and cultured further for 9 hours in culture medium containing 24μM deoxycytidine (Sigma-Aldrich). The cells were washed with 1x PBS, fixed in 100% methanol, permeabilized with 0.1% triton-X, blocked in 5% BSA and co-stained with rabbit polyclonal anti-ZGRF1 antibody (1:1000, Abcam) and mouse monoclonal anti-γ-tubulin antibody (1:5000, Sigma-Aldrich) at 4°C overnight. The cells were further washed and incubated with a 1:500 dilution of anti-rabbit Alexa 568 and anti-mouse Alexa 488 secondary antibodies each, at 4°C for four hours. The cells were counterstained with DAPI and mounted in 70% glycerol. The cells were imaged on a LSM 880 Meta confocal laser scanning microscope (Carl Zeiss, Germany) under a 63x/1.4 oil immersion objective. A total of 180 transfected, mitotic cells were counted for each construct. All statistical comparisons were done using GraphPad Prism5. Statistical test performed was one-way ANOVA followed by Dunnett's test for multiple comparisons. Results were shown as mean ± standard error of the mean (SEM). Differences between groups were considered statistically significant for $P \leq 0.05$ and $P \leq 0.001$.

2.1.14. P-body co-localization study

HEK293 cells were subjected to stress in two ways- a chemical, and an environmental factor: (i) osmotic and oxidative stress induced by treatment with 0.6M sorbitol for 4 hours (Dewey et al 2011) (ii) exposing cells to a high temperature (oxidative stress) (Teixeira et al 2005, Cowart et al 2010) of 42°C for 10 minutes. HEK293 cells transfected with ZGRF1-WT was treated in the same way. The cells were fixed and co-stained with anti-ZGRF1 antibody (1:100 for endogenous, and 1:1000 for transfected cells, Abcam) and mouse monoclonal GW182 (1:100, ab70522, Abcam) - a marker for processing bodies (P-bodies). They were detected with anti-rabbit Alexa Fluor 488 (Molecular Probes, Invitrogen) and anti-mouse Alexa Fluor 568 (Molecular Probes, Invirogen) respectively. UPF1, the nearest paralog of ZGRF1, was treated in the same way and served as a positive control in the experiment. All the sets were experimented with respective untreated controls.

2.1.15. γH2AX DNA damage assay and laser microirradiation

To assay the response of ZGRF1 to DNA damage, cells were treated with mitomycin C and laser microirradiation. Human HEK293, human SH-SY5Y neuroblastoma and rat C6 glioma cells were grown on poly-L-lysine coated coverslips in 10% DMEM up to 80% confluence. To induce DNA damage, the cells were treated with 1μM mitomycin C (Sigma-Aldrich) for 24 hours (Smogorzewska et al 2010). Post treatment, the cells were washed twice with 1x PBS and fixed with 2% PFA. The cells were co-stained for the endogenous proteins with rabbit polyclonal anti-ZGRF1 antibody (1:100, Abcam) and for γH2AX as a marker for DNA damage, with mouse monoclonal to γH2AX antibody (1:300, ab22551, Abcam), to determine mitomycin induced damaged cells. Cells were similarly treated and co-stained for UPF1 and γH2AX. HEK293 cells were transfected with ZGRF1-WT and ZGRF1-Arg326QGln-, Thr602Ile-, Glu660Gly-, Arg1862*-, Phe1940Leu- and Asp1984Gly- Myc-DDK and treated with 1μM mitomycin C, 36 hours post transfection for 24 hours. The cells were co-stained as described, with ZGRF1 antibody at 1:1000 and γH2AX antibody at 1:300.

For laser microirradiation experiment, HEK293 cells were grown in glass-bottomed dishes and transfected with ZGRF1-WT-tGFP and empty vector. Forty eight hours post-transfection, specific regions in the cell nuclei were marked and induced for DNA damage by laser irradiation. Laser-induced damage was performed with 405nm laser diode at 100% laser power for a single pulse of 50 iterations and the cells were monitored for 10 minutes to one hour (Tampere and Mortusewicz 2016).

2.2. Results

2.2.1. Whole exome sequencing reveals a new rare variant co-segregating with HWE in Family 227

The genomic region delimited by markers D4S1572 and D4S2277 at 4q24-q28 was defined by the highest parametric two-point LOD score of 3.50 at θ=0 for D4S402, at 60% penetrance value and 1% phenocopy rate. The whole exome sequencing experiments in III:4 and IV:3 provided a coverage of 95.78% and 95.70% respectively at the 4q24-q28 locus (Tables 2.3 and 2.4). This experiment yielded a total of 563 variants at the region (Figure 2.2). Of the 563 variants, a total of 538 variants were found to occur at a high frequency (MAF > 0.005) across different databases and 25 variants were reported at a MAF ≤ 0.005 (Table 2.5). Variants that were heterozygous and present in any one of the individuals, or homozygous and present in any one or both the individuals that could be confidently determined at >10X read depth, were not considered further. Among the 25 variants, 20 were non protein-coding (10 in UTR and 10 in intron regions) and 5 were protein-coding variants. These comprised 20 SNVs and 5 insertion/deletions (Indels). Of these, eight variants were confirmed by Sanger sequencing to be present in a heterozygous state in both the exome sequenced individuals and examined for segregation with the disease phenotype in the family. Non-segregating variants were not considered further. A single non-synonymous, heterozygous, segregating variant, c.1805C>T located in the exon 5 of *ZGRF1* (previously known as *C4orf21*) was obtained (Figure 2.3). This variant codes for the corresponding amino acid change p.Thr602Ile. Of the available species sequences, the residue is conserved among higher mammals (Figure 2.4).

Table 2.3: Summary statistics for exome sequencing experiment

Alignment features	Individual III:4		Individual IV:3	
Read length in bases (paired)	100		100	
Total number of reads[a]	79978638		84909810	
Total reads for alignment (percentage high quality reads)[b]	74872876	(93.6)	79634096	(93.8)
Total reads aligned (percentage reads aligned)[c]	74575432	(99.60)	79332908	(99.62)
Reads aligned to whole exome (percentage reads aligned)	51581554	(69.17)	54971128	(69.29)
4q24-q28 target length	275107		275107	
4q24-q28 target covered (percentage target covered)	263491	(95.78)	263277	(95.70)

[a] Number of raw reads, [b] Number of reads yielded post processing of raw reads. Reads with >70% bases with Phred score >20, [c] All reads aligned, including exome and other regions of the genome.

Table 2.4: Sequence coverage summary of the 4q24-q28 target region

Coverage attributes	Individual III:4		Individual IV:3	
	Chromosome 4q 24-q28	Whole exome	Chromosome 4q24-q28	Whole exome
%Total target covered with at least 5X read depth	93.40	90.33	93.39	90.68
%Total target covered with at least 10X read depth	92.71	86.56	92.79	86.99
%Total target covered with at least 15X read depth	92.08	83.40	92.20	83.92
%Total target covered with at least 20X read depth	91.24	80.37	91.37	81.04

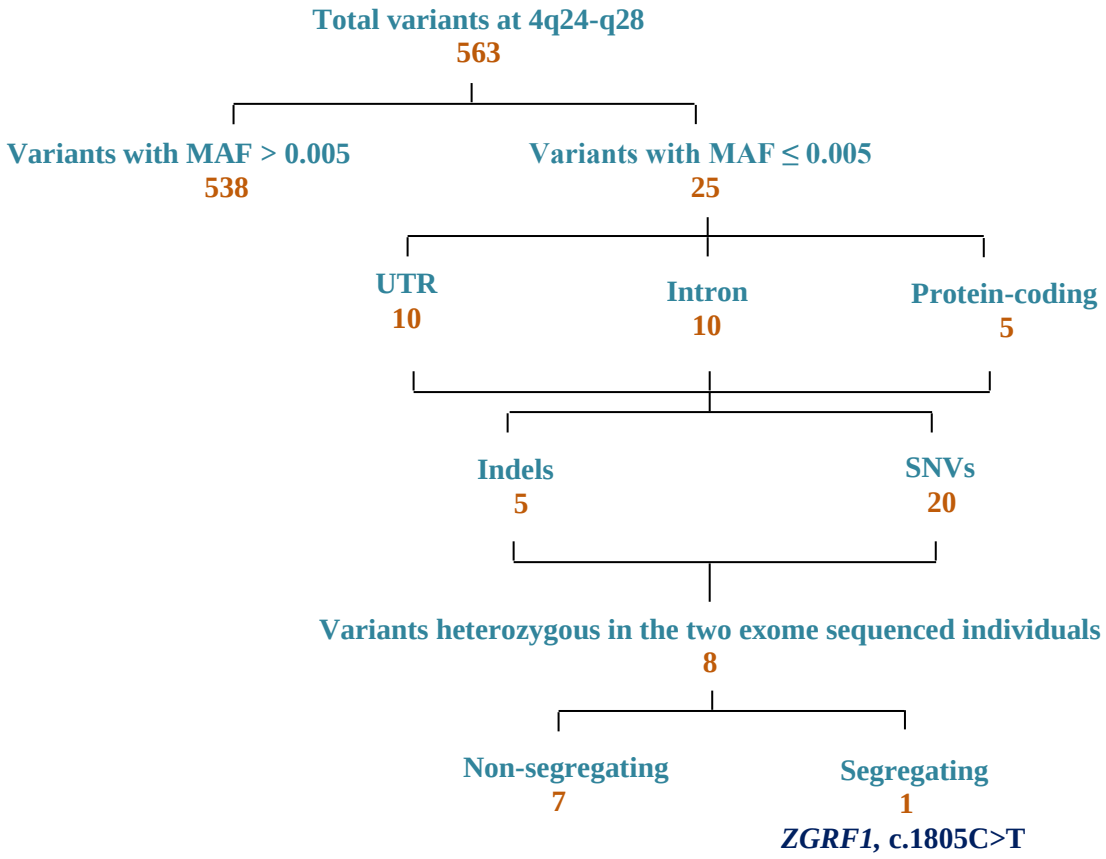


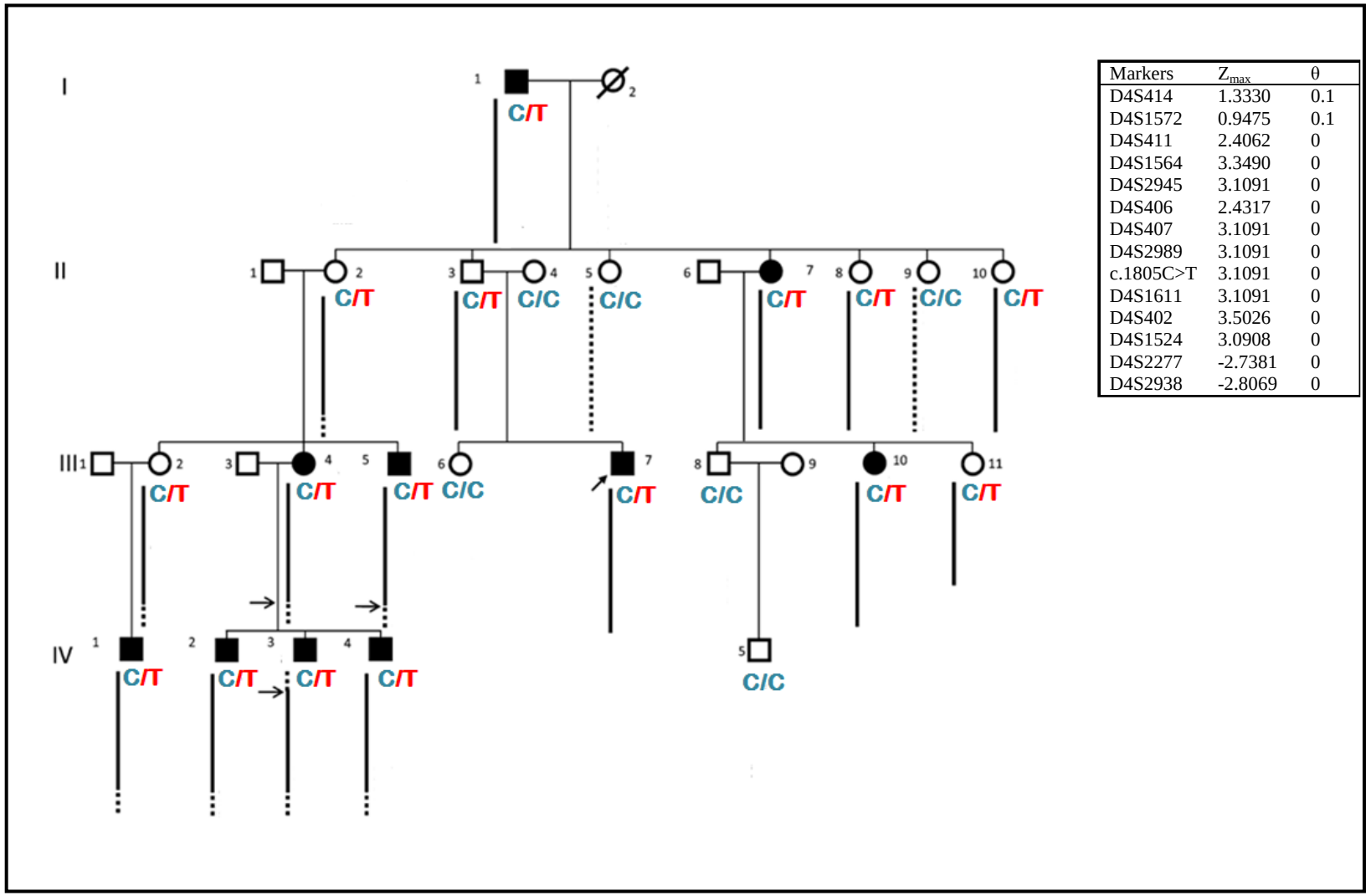
Figure 2.2: Targeted analysis of the 4q24-q28 region in whole exome sequencing experiment. The schematic represents the work flow for screening and filtering the gene variants in the 4q24-q28 target region to identify potentially causative one.

Table 2.5: Twenty five novel or rare variants with MAF ≤ 0.005 from whole exome sequencing analysis

Genomic position	Gene	Sequence variant	Hom in either or both (Y/-)[a]	Het and present in either sample (Y/-)[b]	Het and present in both samples (Y/-)[c]	Effect on protein	Variant ID	Frequency in control databases[d]				Segregation (Y/N)[e]
								ExAC	1000G	ESP	dbSNP144	
103808458[f]	*CISD2*	NM_001008388.4:c.319-40T>C	-	Y	-	-	rs201682974	0.0010	-	-	-	-
103822063[h]	*SLC9B1*	NM_139173.3:c.*211C>T	-	-	Y	-	rs200321062	-	-	-	-	N
103822082[h]	*SLC9B1*	NM_139173.3:c.*192C>T	-	-	Y	-	rs62327290	-	-	-	-	N
103822089[h]	*SLC9B1*	NM_139173.3:c.*185T>A	-	-	Y	-	rs62327291	-	-	-	-	N
103822298[h]	*SLC9B1*	NM_139173.3:c.1524G>A	-	-	Y	p.Gln508=	-	0.00003	-	-	-	N
103822492[h]	*SLC9B1*	NM_139173.3:c.1333-3C>T	-	-	Y	-	rs3974500	0.0029	-	-	-	N
104640282[f]	*TACR3*	NM_001059.2:c.548+3T>C	-	Y	-	-	-	-	-	-	-	-
106055705[f]	*TET2*	NM_017628.4:c.606C>T	-	Y	-	p.Asn202=	rs201865755	0.0028	0.0038	-	0.0038	-
106068079[f]	*TET2*	NM_017628.4:c.-193+836_-193+839delAGAA	-	Y	-	-	rs572309712	-	0.0034	-	0.0034	-
106474096[f]	*ARHGEF38*	NM_017700:c.174C>T	Y	-	-	p.Thr58=	rs6533206	0.00007	0.0002	-	0.0002	-
107237722[f]	*AIMP1*	NM_001142416.1:c.18T>C	-	Y	-	p.Ala6=	-	-	-	-	-	-
109543638[f]	*RPL34*	NM_033625.3:c.166-33C>T	-	Y	-	-	rs200789602	0.0008	0.0010	0.0002	0.0002	-
109578572[f]	*OSTC*	NM_021227.3:c.234-34G>A	Y	-	-	-	rs2851379	0.0044	0.0018	0.0047	0.0018	-
109578850[f]	*OSTC*	NM_021227.3:c.431+47T>C	-	Y	-	-	rs375863579	0.0023	0.0028	0.00007	0.0028	-
110384047[h]	*SEC24B*	NM_006323.4:c.134-10_134-9insCTTTT	-	-	Y	-	rs70948092	0.0002	-	0.0164	-	N
110635489[f]	*PLA2G12A*	NM_030821.4:c.*43_*44insA	-	N	-	-	rs57182211	-	-	-	-	-
113539393[h]	*ZGRF1*	NM_018392.4:c.1805C>T	-	-	Y	p.Thr602Ile	rs201904239	0.0012	0.0012	-	0.0012	Y

Genomic position	Gene	Sequence variant	Hom in either or both (Y/-)[a]	Het and present in either sample (Y/-)[b]	Het and present in both samples (Y/-)[c]	Effect on protein	Variant ID	Frequency in control databases[d]				Segregation (Y/N)[e]
								ExAC	1000G	ESP	dbSNP144	
113568126[h]	*LARP7*	NM_016648.3:c.552+15_552+16insA	-	-	Y	-	rs747753883	-	-	-	-	N
114267023[f]	*ANK2*	NM_020977.3:c.4249-33G>A	Y	-	-	-	rs9968405	0.0002	0.0002	0.0002	0.0002	-
114304735[f]	*ANK2*	NM_020977.3:c.*2108C>T	-	Y	-	-	rs190513908	-	0.0008	-	0.0008	-
114822867[g]	*ARSJ*	NM_024590.3:c.*561_*562insAA	-	Y	-	-	rs397732178	-	-	-	-	-
123816124[f]	*FGF2*	NM_002006.4:c.*2573T>C	Y	-	-	-	rs2168915	-	0.0030	-	0.0030	-
123843750[f]	*NUDT6*	NM_145207.2:c.-548G>T	-	Y	-	-	rs200477292	0.0018	0.0024	0.0009	0.0024	-
123977466[f]	*SPATA5*	NM_145207.2:c.2080-76T>A	-	Y	-	-	rs558517923	-	0.0002	-	0.0002	-
124240031[f]	*SPATA5*	NM_145207.2:c.*4812G>C	-	Y	-	-	rs538982379	-	0.0026	-	0.0026	-

[a] Any homozygous variant present in either or both of the exome sequenced individuals was not considered further for segregation analysis in the family. 'Y' represents yes for presence of the same.
[b] Any heterozygous variant present in either of the two exome sequenced individuals were not taken forward for segregation analysis in the family.
[c] Heterozygous variants present in both the exome sequenced individuals were confirmed by Sanger sequencing and analyzed for segregation in the family.
[d] Frequency of all variants obtained from the analysis were determined in four databases. Variants with a MAF ≤ 0.005 in either of the databases were taken into consideration.
[e] Variants that satisfied the condition of being heterozygous in both the exome sequenced individuals and presence in proband, were considered for segregation analysis in the complete family.
[f] Variants examined and eliminated by sequence analysis at >10X read depth in SAMtools.
[g] Rare variant identified from UTR regions covered by Sanger sequencing.
[h] Variants confirmed and examined by Sanger sequencing.
Y = Yes, N = No.



Markers	Z_{max}	θ
D4S414	1.3330	0.1
D4S1572	0.9475	0.1
D4S411	2.4062	0
D4S1564	3.3490	0
D4S2945	3.1091	0
D4S406	2.4317	0
D4S407	3.1091	0
D4S2989	3.1091	0
c.1805C>T	3.1091	0
D4S1611	3.1091	0
D4S402	3.5026	0
D4S1524	3.0908	0
D4S2277	-2.7381	0
D4S2938	-2.8069	0

Figure 2.3: Segregation of *ZGRF1*, c.1805C>T in HWE227. Family pedigree of HWE227 representing the segregation of c.1805C>T with the disease haplotype. The maximum LOD score obtained for the markers and for the variant is represented in the box.

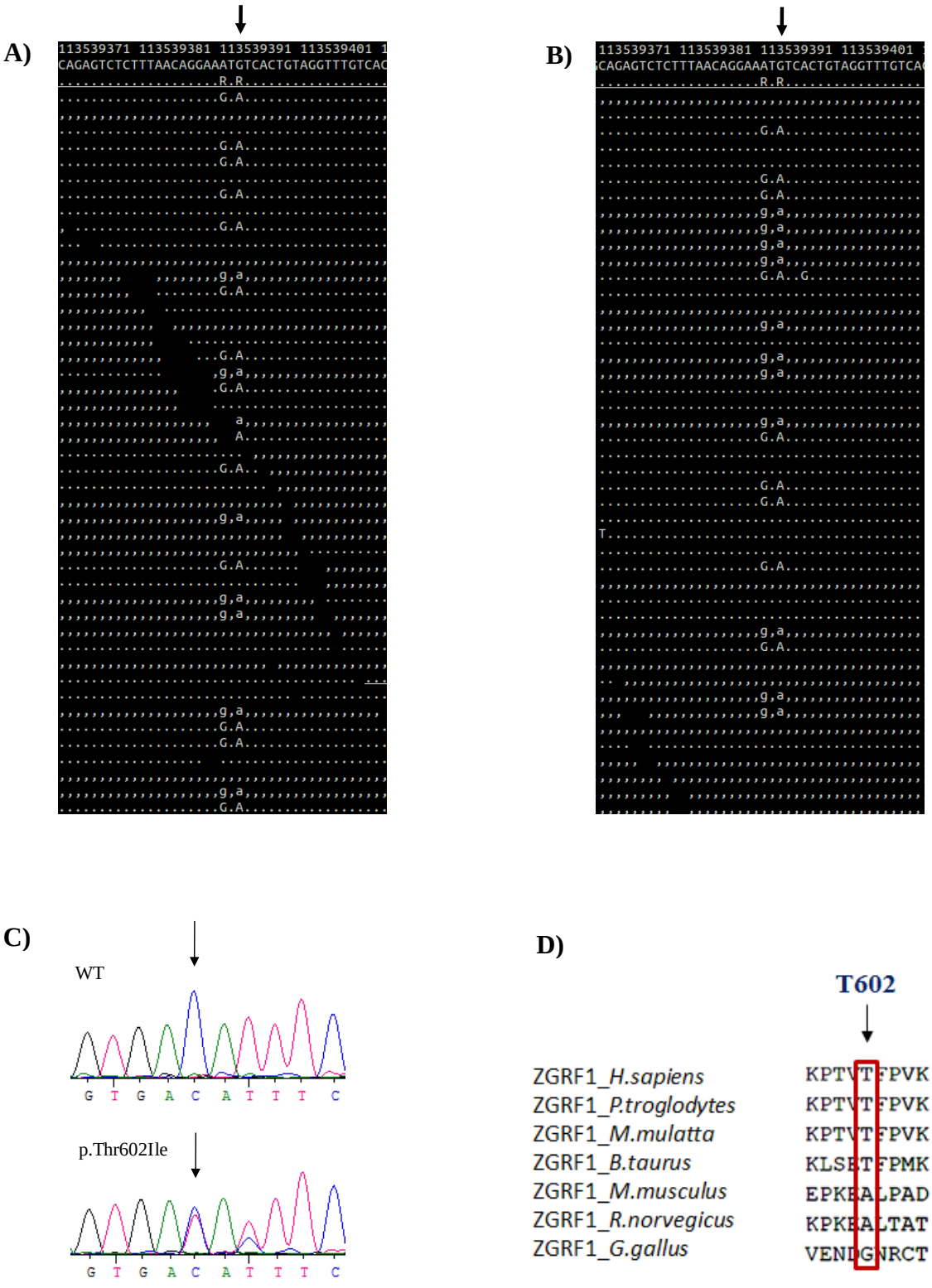


Figure 2.4: *ZGRF1*, c.1805C>T. (A) SAMtools alignment snapshots of c.1805C>T variant in individual III:4 and (B) individual IV:3. The variant is marked with an arrow in both the individuals as G>A in the forward strand alignment to the reference genome. The variant is present at high read depth and in both forward (marked by dots) and reverse (marked by comma) strands of the two individuals. (C) Electropherogram representation of the variant in an unaffected individual and the proband in the family. (D) Conservation of the amino acid residue T602 across different species. The residue is conserved across higher mammals.

2.2.2. Sequence analysis of *ZGRF1* in south Indian HWE patients reveals six heterozygous rare variants in the coding exons

The complete transcript structure of *ZGRF1* was sequenced using Sanger based sequencing in 288 HWE probands from southern India. Thirty five single nucleotide variations (SNVs) were identified comprising coding and non-coding variants (Table 2.6). All the variants were examined for their frequencies in 288 controls from south India. From the SNVs identified, 17 were non-coding and 18, coding. Of the 17 non-coding variants, seven were unreported in databases. They were found at high frequencies in the south Indian control population with comparable frequency in the HWE patients and the control cohort. The reported non-coding variants were found at high frequency in Indian controls. Among the 18 coding variants, four were unreported in databases, of which two were found in high frequency in south Indian controls and two were absent in these controls. Of the 14 reported coding variants, 11 were at high frequency in databases, two were rare, and one did not have reported frequency. The three variants (two rare and one with unreported frequency) were either absent or rare in the south Indian controls examined. Taken together, this study identified five rare missense variants in the gene, c.977G>A (p.Arg326Gln), c.1979A>G (p.Glu660Gly), c.5584C>T (p.Arg1862*), c.5818T>C (p.Phe1940Leu) and c.5951A>G (p.Asp1984Gly). Variant c.1805C>T (p.Thr602Ile) had been identified at 4q24-q28 to be segregating with the phenotype in HWE227. It was found again, in two apparently unrelated families (Tables 2.7 and 2.8, Figures 2.5 and 2.6).

Table 2.6: *ZGRF1* variants among HWE patients

Exon/Intron	Nomenclature	Amino acid change	Ref SNP ID	Allele frequency in Indian controls (n=576)	Allele frequency in control databases
5'UTR	c.-284C>T	-	≠	0.03	-
5'UTR	c.-176C>A	-	≠	0.01	-
5'UTR	c.-53T>A	-	rs13104310	0.035	0.42
Intron 1	c.22-99G>A	-	≠	0.07	-
Intron 2	c.102+26G>T	-	≠	0.02	-
Exon 3	c.142C>A	p.Leu48Met	rs61745597	0.012	0.0204
Intron 3	c.162+13G>C	-	rs775648125	0.002	0.00012
Exon 4	c.170C>T	p.Pro57Leu	rs775504533	0.02	-

Exon/Intron	Nomenclature	Amino acid change	Ref SNP ID	Allele frequency in Indian controls (n=576)	Allele frequency in control databases
Intron 4	c.351+59G>T	-	rs1017468137	0.012	-
Intron 4	c.351+28T>C	-	rs753126322	0.03	0.00034
Intron 4	c.351+49G>T	-	≠	0.01	-
Exon 5	c.977G>A	p.Arg326Gln	rs535674967	0	-
Exon 5	c.1035G>C	p.Gly345=	rs188866585	0.03	0.0005
Exon 5	c.1071T>C	p.Asp357=	rs575464722	0.035	0.00012
Exon 5	c.1229A>G	p.Asn410Ser	rs7696816	0.03	0.4137
Exon 5	c.1085C>T	p.Thr602Ile	rs201904239	0.0034	0.001
Exon 5	c.1677G>A	p.Glu557Lys	≠	0.02	-
Exon 5	c.1087T>C	p.Phe603Leu	rs78381215	0.02	0.026
Exon 5	c.1979A>G	p.Glu660Gly	≠	0	-
Intron 8	c.2802+92T>C	-	rs62316615	0	0.098
Exon 10	c.3060A>G	p.Val1020=	≠	0.01	-
Exon 15	c.4434C>T	p.Ser1478=	rs561737889	0.005	0.0002
Intron 15	c.4439-91G>A	-	rs186067473	0.001	0.0004
Intron 15	c.4439-73A>G	-	≠	0.022	-
Intron 15	c.4439-41G>A	-	rs182036970	0.012	0.008
Intron 17	c.4697+81G>T	-	rs6838559	0.022	0.4284
Exon 21	c.5289T>C	p.Tyr1763=	rs141688719	0.0012	0.003
Intron 22	c.5474+83C>G	-	rs78439826	0.007	0.022
Exon 23	c.5475G>A	p.Arg1825=	rs78336146	0.026	0.0262
Exon 23	c.5584C>T	p.Arg1862*	≠	0	-
Exon 25	c.5816C>T	p.Thr1939Met	rs150416544	0.012	0.0004
Exon 25	c.5818T>C	p.Phe1940Leu	rs773566516	0	0.00005
Exon 26	c.5951A>G	p.Asp1984Gly	rs780748235	0.0017	0.00008
3'UTR	c.*17T>G	-	≠	0	-
3'UTR	c.*52A>G	-	rs564397944	0.012	0.001

≠ Variants not reported in any of the control databases.
Allele frequency where not available is represented with '-'.

Table 2.7: Rare non-synonymous *ZGRF1* variants identified in the HWE patients examined

Nucleotide change	Amino acid change	Family ID	Type of family	dbSNP ID	MAF in control chromosomes (n=576)	Minor allele frequency (MAF) in control databses			
						dbSNP135	1000G	ESP	ExAC
c.977G>A	p.Arg326Gln	HWE 026	Familial	rs535674967	0	0.0002	0.00	-	0.00004
c.1805C>T	p.Thr602Ile	HWE 011 HWE 160 HWE 209	Familial, branch 1 Familial Familial	rs201904239	0.0034	0.0012	0.001	-	0.0012
c.1979A>G	p.Glu660Gly	HWE 035	Familial	-	0	-	-	-	-
c.5584C>T	p.Arg1862*	HWE 050 HWE 041	Familial Trio	-	0	-	-	-	-
c.5818T>C	p.Phe1940Leu	HWE 011 HWE 223	Familial, branch 2 Familial	-	0	-	-	-	0.00003
c.5951A>G	p.Asp1984Gly	HWE 124	Familial	-	0.0017	-	-	-	0.00008

A)

B)

Domains	Amino acids
DUF2439	4-73
Zf-GRF	1347-1391
AAA_11	1614-1844
DNA2	1739-2057
recD	1799-1842
AAA_12	1855-2039

Figure 2.5: ZGRF1 protein with location of the six rare variants. (A) Representation of the ZGRF1 protein structure showing the positions of the six rare variants. Three of the variants lie in a non-domain containing region of the protein and the other three lie within the P-loop NTPase/UvrDC2 domain of the protein. (B) Amino acid positions of the different proposed domains of ZGRF1.

Table 2.8: Bioinformatic analyses of the six rare variants obtained in *ZGRF1*

Mutation	Conservation Score[a]	SIFT Score[b]	PolyPhen-2 Hum Var Score[c]	PROVEAN Score[d]	Mutation Taster Probability Value[e]	Mutation Assessor FI Score[f]	FATHMM Score[g]
p.Arg326Gln	0.015	Tolerated 0.26	Probably damaging 0.998	Neutral -1.391	Polymorphism 0.923	Low 1.2	Damaging -1.96
p.Thr602Ile	0.601	Damaging 0.03	Benign 0.012	Neutral -0.969	Polymorphism 0.999	Low 1.04	Damaging -1.64
p.Glu660Gly	-0.817	Damaging 0.05	Benign 0.013	Neutral -2.021	Polymorphism 0.999	Medium 2.25	Damaging -2.02
p.Arg1862*	-1.202	Damaging 0.01	-	-	Disease causing 0.999	-	-
p.Phe1940Leu	-1.094	Tolerated 0.11	Benign 0.284	Deleterious -3.321	Disease causing 0.967	Neutral 0.065	Damaging -2.86
p.Asp1984Gly	1.562	Tolerated 0.35	Benign 0.108	Neutral -2.222	Polymorphism 0.999	Neutral 0.67	Tolerated -1.13

[a] The Consurf Server; conservation scores for the amino acids based on the phylogenetic relationship from 41 homologous sequences (Ashkenazy et al 2010).
[b] Sorting Intolerant From Tolerant [SIFT] (Kumar et al 2009) : SIFT predicts for amino acid substitutions with scores ranging from 0-1. Variant scores ranging from 0.0-0.05 are considered deleterious, with scores closer to 0.0 being more confidently predicted as deleterious. Scores between 0.05-1.0 are considered tolerated, with scores closer to 1.0 having more confidence for tolerated prediction.
[c] Polymorphism Phenotyping v2 [PolyPhen-2] (Adzhubei et al 2010) : PolyPhen predicts for amino acid substitutions with scores ranging from 0-1 as being benign to damaging.
[d] Protein Variation Effect Analyzer [PROVEAN] (Choi and Chan 2015) : Variants with a score equal to or below -2.5 are considered 'deleterious'. Variants with a score above -2.5 are considered 'neutral'.
[e] Mutation Taster (Schwarz et al 2014) : Probability of prediction; values close to 1 indicates a high security of the prediction.
[f] Functional impact score (FIS) (Reva et al 2011) : Neutral impact ($FIS \leq 0.8$), Low impact ($0.8 < FIS \leq 1.9$), Medium impact ($1.9 < FIS \leq 3.5$), High impact ($FIS > 3.5$).
[g] Functional Analysis through Hidden Markov Models (FATHMM) (Shihab et al 2013) : Combines sequence conservation with hidden Markov models to determine 'pathogenecity weights'

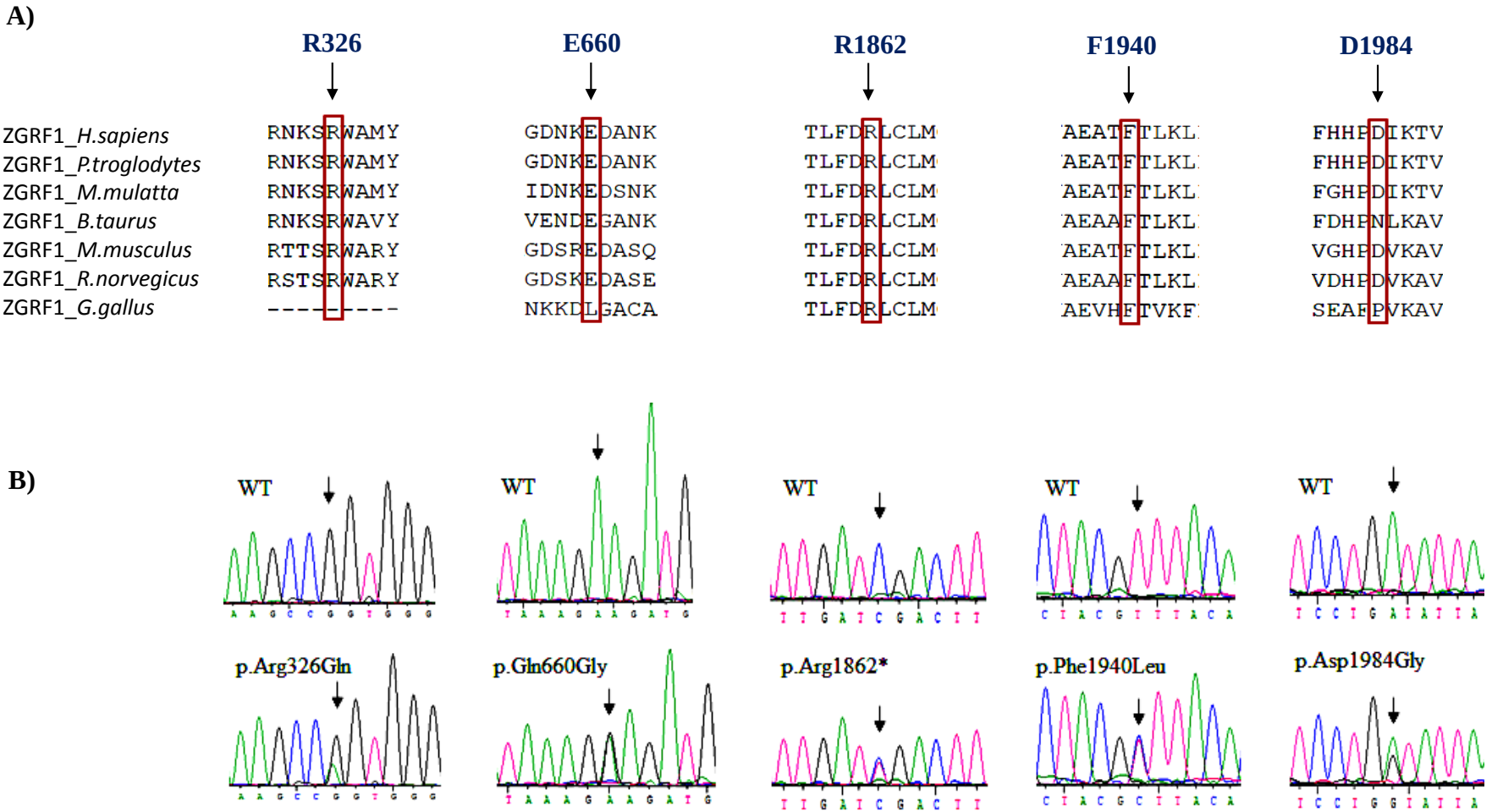


Figure 2.6: Sequence of rare variants in *ZGRF1*. (A) Schematic representation of the five variants in ZGRF1 and their conservation profile across different species; *H.sapiens*: NP_060862.3, *P.troglodytes*: XP_517401.3, *M.mulatta*: XP_002804225.1, *B.taurus*: XP_002688121.1, *M.musculus*: NP_932114.2, *R.norvegicus*: XP_006233355.1 and *G.gallus*: XP_426326.4. (B) Representative electropherogram sequences of the variants in a normal and affected individual.

2.2.3. Gene and protein architecture of ZGRF1

The *ZGRF1* (zinc finger GRF-type containing 1) gene comprises 97.66kb of genome sequence length encoding the 5'UTR, 27 coding exons, introns and 3'UTR (Figure 2.7). *ZGRF1* located at 4q25 encodes a protein of 2104 amino acids length (Figure 2.8). According to HGNC (HUGO Gene Nomenclature Committee), ZGRF1 belongs to the families of zinc-fingers GRF-type and UPF1 like RNA helicases. The UPF1 like RNA helicases family comprises 11 genes, *AQR* (aquarius intron-binding spliceosomal factor), *DNA2* (DNA replication helicase/nuclease2), *HELZ* (helicase with zinc finger), *HELZ2* (helicase with zinc finger 2), *IGHMBP2* (immunoglobulin mu binding protein 2), *MOV10* (Mov10 RISC complex RNA helicase), *MOV10L1* (Mov10 RISC complex RNA helicase like 1), *SETX* (senataxin), *UPF1* (up-frameshift 1, RNA helicase and ATPase), *ZNFX1* (zinc finger NFX1-type containing 1) and *ZGRF1*. DELTA-BLAST search against protein data bank (PDB) proteins yielded significant alignment at the C-terminal region of ZGRF1 with highest E-values for Sen1 helicase chain A and Upf1-RNA complex chain A (Figure 2.9A and B). With data curated from NCBI-blastp (protein-protein BLAST), DELTA-BLAST non-redundant proteins and Ensembl, ZGRF1 was found to consist of six domains.

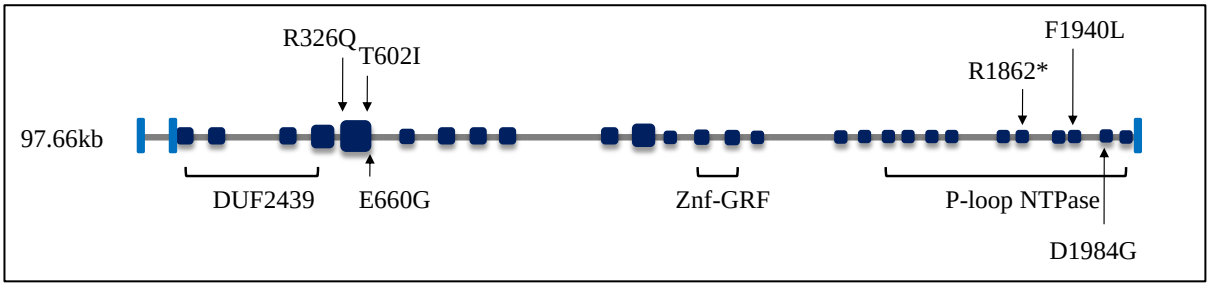


Figure 2.7: *ZGRF1* structure. Representation of *ZGRF1* gene of 97.66kb sequence length comprising 27 coding exons. Shown here are the corresponding protein domains and the arrows pointing the exons/domains with the six rare HWE variants. The exons are marked in dark blue. The light blue bars represent the 5'- and 3'- UTRs.

An *N-terminal domain* DUF2439 (domain of unknown function 2439) encompasses 70 amino acids. Orthologs of this domain are found in DNA break localizing Dbl2, ATP-dependent DNA helicase Rdh54 (in *Schizosaccharomyces pombe),* Mph1-associated telomere maintenance protein1 Mte1, Rdh54 (in *Saccharomyces cerevisiae*), AT4G10890 (in *Arabidopsis thaliana*) and DNA repair and recombination protein RAD54B (*Homo sapiens*) (Polakova et al 2016). DUF2439 truncation mutants of Rdh54 in *Saccharomyces cerevisiae* fail to interact with Rad51 and lose ability to appropriately dissociate Rad51-dsDNA

(double-stranded DNA) complexes (Santa Maria et al 2013, Chi et al 2006, Raschle et al 2004). Dbl2 have been implicated in meiotic chromosome segregation (Polakova et al 2016). Mte1 is reported to interact with Mph1 (serine threonine protein kinase) for crossover recombination and telomere maintenance (Silva et al 2016).

A *second domain* zf-GRF (zinc finger-GRF type) is a presumed zinc binding domain of 45 amino acids named after three conserved residues glycine, arginine and phenylalanine. This domain occurs in six other genes namely, *APEX2/ZGRF2* (apurinic/apyrimidic endodeoxyribonuclease 2), *NEIL3/ZGRF3* (nei like DNA glycosylase 3), *ZCCHC4/ZGRF4* (zinc finger CCHC-type containing 4), *ERI2/ZGRF5* (ERI1 exoribonuclease family member 2), *TTF2/ZGRF6* (transcription termination factor 2) and *TOP3A/ZGRF7* (DNA topoisomerase III alpha).

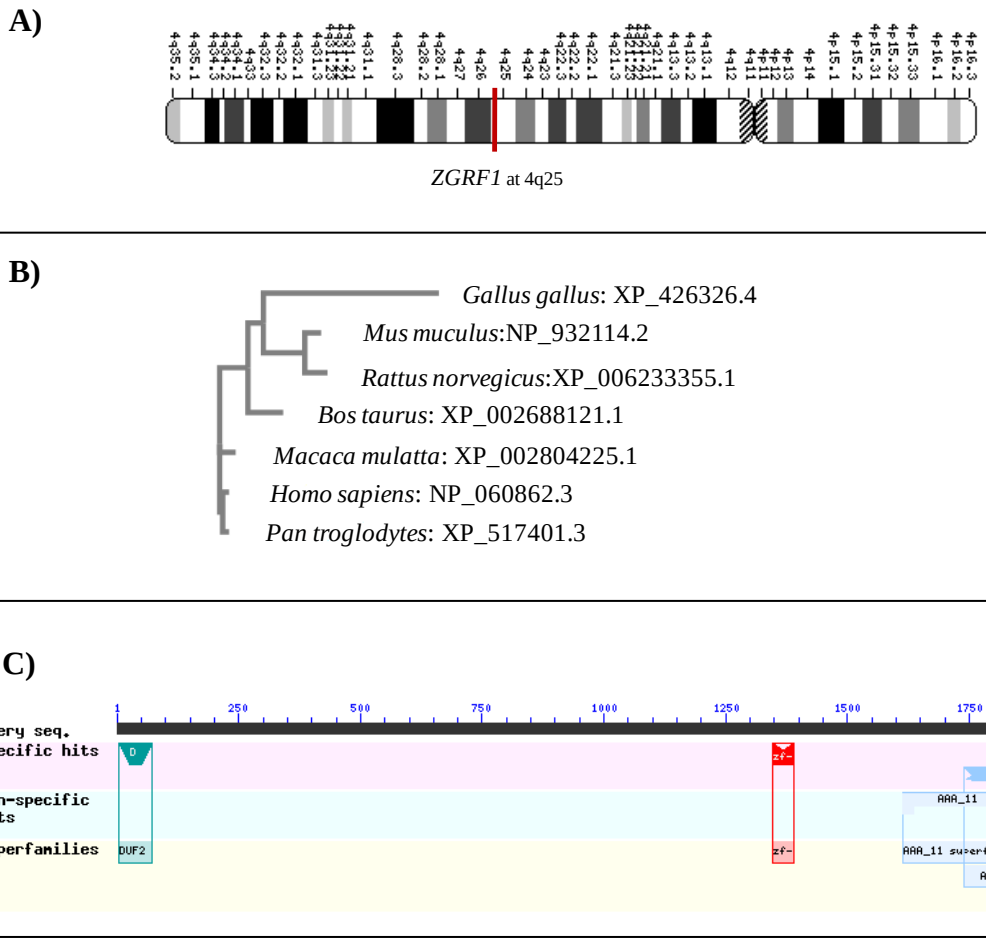


Figure 2.8: ZGRF1 bioinformatic analysis. (A) Chromosome 4 ideogram representing the location of *ZGRF1* at 4q25. (B) Phylogenetic tree (with distance correction) of ZGRF1 across different species. (C) DELTA-BLAST output of ZGRF1 representing the different domains and their amino acid positions.

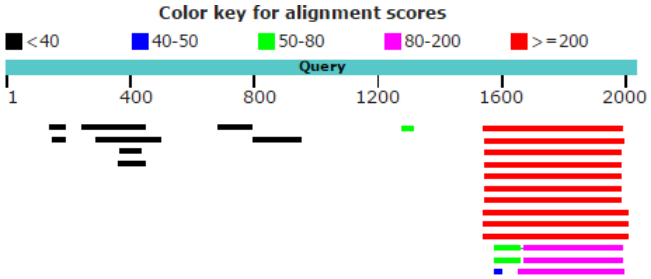
The *third and fourth domains* AAA_11 and AAA_12/UvrDC2 span 231 and 185 amino acids respectively in the C-terminus of the protein. These domains belong to a large and functionally diverse AAA superfamily (ATPase Associated with diverse cellular Activities) of ring-shaped P-loop NTPases that function as molecular chaperons (Iyer et al 2004) through energy-dependent unfolding of macromolecules (Frickey and Lupas 2004). AAA superfamily belongs to the second major structural group of P-loop NTPases, namely, additional strand catalytic E (ASCE) group, that commonly function as oligomeric rings with a hexameric arrangement (Iyer et al 2004). P-loop NTPases contain a Walker A motif or P-loop which is a conserved nucleotide phosphate binding motif that binds to β/γ phosphate of ATP/GTP, and a Walker B motif that binds to magnesium ion (Walker et al 1982). UvrDC2 is found at the C-terminus of several DNA and RNA helicases.

A *fifth domain*, DNA2, of 319 amino acids belong to the superfamily I (SF1) DNA and/or RNA helicase. This is characterized by the P-loop motifs. Structural data is available for a single SF1 RNA helicase, UPF1 (up-frameshift 1), a protein involved in nonsense-mediated mRNA decay (Cheng et al 2007). Biochemical analysis of SF1 RNA helicases (Upf1, SETX, IGHMBP2) have revealed dual RNA- and DNA- stimulated ATPase and helicase activities (Kim et al 1999, Bhattacharya et al 2000, Guenther et al 2009). A *sixth domain* recD, spanning 44 amino acids is a family of exodeoxyribonuclease V. This domain usually has nuclease/helicase activity.

2.2.4. Conservation of domains among ZGRF1 and its nearest paralog UPF1

As mentioned earlier, ZGRF1 belongs to the family of SF1-Upf1-like RNA helicases (http://www.rnahelicase.org/sf1.html). UPF1 is a DNA/RNA helicase, a core nonsense-mediated mRNA decay (NMD) factor, and an RNA surveillance protein. It degrades transcripts with premature nonsense mutations and prevents the translation of truncated proteins (Avery et al 2011). ZGRF1 has protein domains similar to UPF1 at its C-terminal region (Figure 2.9C). Both proteins comprise of DNA2 and recD domains. They have zinc finger domains and helicase regions comprised of P-loop NTPase domain. The pairwise alignment of ZGRF1 to the PDB UPF1-RNA complex (2XZL_A, Chakrabarti et al 2011) represents 29% identity and 46% similarity (Figure 2.9D).

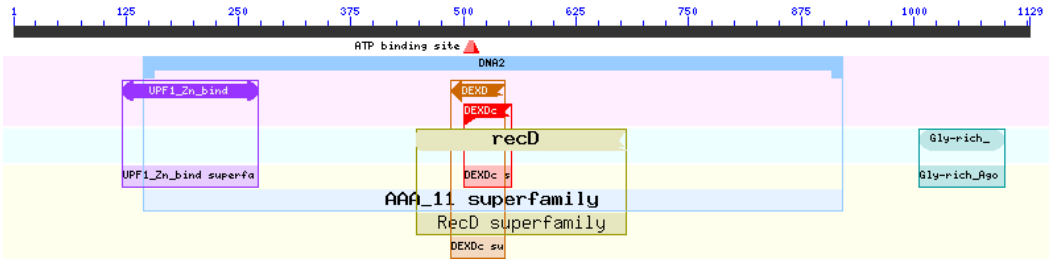
A)



B)

Description	Max score	Total score	Query cover	E value	Ident
Chain A, Helicase Sen1	430	430	22%	5e-131	26%
Chain A, Upf1-Rna Complex	415	415	22%	4e-125	29%
Chain A, Upf1 Helicase - Rna Complex	405	405	21%	7e-124	29%
Chain A, Structural And Functional Insights Into The Human Upf1 Helicase Core	405	405	21%	7e-124	29%
Chain A, Crystal Structure Of The Complex Between Human Nonsense Mediated Decay Factors Upf1 And Upf2	405	405	21%	9e-122	29%
Chain X, Crystal Structure Of 1qhmbp2 Helicase	383	383	21%	9e-116	24%
Chain A, Crystal Structure Of Iqhmbp2 Helicase In Complex With Rna	383	383	21%	9e-116	24%
Chain A, Crystal Structure Of Dna2 Nuclease-helicase	315	315	23%	1e-88	24%
Chain A, Crystal Structure Of Dna2 In Complex With An Ssdna	315	315	23%	1e-88	24%
Chain A, Crystal Structure Of Dna2 In Complex With A 5' Overhang Dna	315	315	23%	1e-88	24%

C)



D)

Chain A, Upf1-Rna Complex

Sequence ID: 2XZL_A Length: 802 Number of Matches: 1

Range 1: 360 to 795 GenPept Graphics

Score	Expect	Method	Identities	Positives	Gaps
415 bits(1067)	4e-125	Composition-based stats.	138/471(29%)	219/471(46%)	65/471(11%)

```
Query  1534  KLNKDQATALIQIAQMMASHESIEEVKELQTHTFPITIIHGVFGAGKSYLLAVVILFFVQ  1593
             +LN  Q+ A+  + Q                P+++I G  G GK+   A ++
Sbjct  360   QLNSSQSNAVSHVLQR-----------------PLSLIQGPPGTGKTVTSATIVYHL--  399

Query  1594  LFEKSEAPTIGNARPWKLLISSSTNVAVDRVLLGLLSLGFENFIRVGSVR--KIAKPILP  1651
                          ++L+ + +NVAVD +   L         +R+ +     +   +
Sbjct  400   ----------SKIHKDRILVCAPSNVAVDHLAAKL-RDLGLKVVRLTAKSREDVESSVSN  448

Query  1652  YSLHAGSENESEQLKELHALMKEDLTPTERVYVRKSIEQHKLGTNRTLLKQVRVVGVTCA  1711
              +LH      ++   +    +K+++          K   +   T   +L +  VV   TC
Sbjct  449   LALHNLVGRGAKGELKNLLKLKDEVG-ELSASDTKRFVKLVRKTEAEILNKADVVCCTCV  507

Query  1712  ACPFPCMNDLKFPVVVLDECSQITEPASLLPIARFECEKLILVGDPKQLPPTIQGSDAAH  1771
                    ++  KF  V++DE +Q +EP  L+PI +   +++ILVGD +QL P I    AA
Sbjct  508   GAGDKRLDT-KFRTVLIDESTQASEPECLIPIVK-GAKQVILVGDHQQLGPVILERKAAD  565

Query  1772  ENGLEQTLFDRLCLMGHKPILLRTQYRCHPAISAIANDLFYKGALMNGVTEIERSPLLEW  1831
               GL+Q+LF+RL  +GH PI L  QYR +P +S   +++FY+G+L NGVT  +R+
Sbjct  566   -AGLKQSLFERLISLGHVPIRLEVQYRMNPYLSEFPSNMFYEGSLQNGVTIEQRTVPNSK  624

Query  1832  LP------TLCFYNVKGLEQ-IERDNSFHNVAEATFTLKLIQSLIASGIAGSMIGVITLY  1884
              P       + F+   G E+      SF N  EA    ++I  L   G+    IGVIT Y
Sbjct  625   FPWPIRGIPMMFWANYGREEISANGTSFLNRIEAMNCERIITKLFRDGVKPEQIGVITPY  684

Query  1885  KSQMYKLCHLLSAVDFHHPDIK-TVQVSTVDAFQGAEKEIIILSCVRTRQ---VGFIDSE  1940
             + Q   +   +         D+   V+V++VDAFQG EK+ IILSCVR   +  +GF+
Sbjct  685   EGQRAYILQYMQMNGSLDKDLYIKVEVASVDAFQGREKDYIILSCVRANEQQAIGFLRDP  744

Query  1941  KRMNVALTRGKRHLLIVGNLACLRKNQLWGRVIQH-------CEGREDGLQ  1984
             +R+NV LTR K  L+I+GN   L +N LW  ++ H        EG  D LQ
Sbjct  745   RRLNVGLTRAKYGLVILGNPRSLARNTLWNHLLIHFREKGCLVEGTLDNLQ  795
```

Figure 2.9: BLAST analysis of ZGRF1. (A) BLAST alignment output for ZGRF1 full-length protein. (B) BLAST output for the first 10 sequences with highest E-values better than threshold. (C) DELTA-BLAST domain alignment of UPF1. (D) Pairwise alignment of the ZGRF1 query sequence with the PDB entry 2XZL_A showing identity and similarity of 29% and 46%, respectively, between the amino acid residues.

2.2.5. Expression of *ZGRF1* transcripts in human brain regions

In different regions of the human brain, the N-terminal primer pair amplified a product of 656 bases and the C-terminal primer pair amplified 572 bases (Figure 2.10A). Nucleotide sequences of the gel purified products presented a complete sequence match with the *ZGRF1* transcript (NM_018392.4) in the regions spanned by the primer pairs. This observation suggests expression of *ZGRF1* at a transcript level in the human brain regions: cerebral cortex, cerebellum, hippocampus, hypothalamus and whole brain. The full length of *ZGRF1* transcript is 6.3kb. My initial attempts to amplify this sequence length did not succeed in the cDNA pool. However, portions of the transcript could be successfully amplified to indicate that the full-length transcript is expressed. We procured the cDNA clone for ZGRF1 (Origene, NM_018392.4). The RNA expression profile of *ZGRF1* across various tissues was determined from the HPA RNA-seq of normal tissues (Human Protein Atlas RNA-sequencing, https://www.proteinatlas.org/ENSG00000138658-ZGRF1/tissue). The transcript is moderately expressed across the different human tissues (Figure 2.10B).

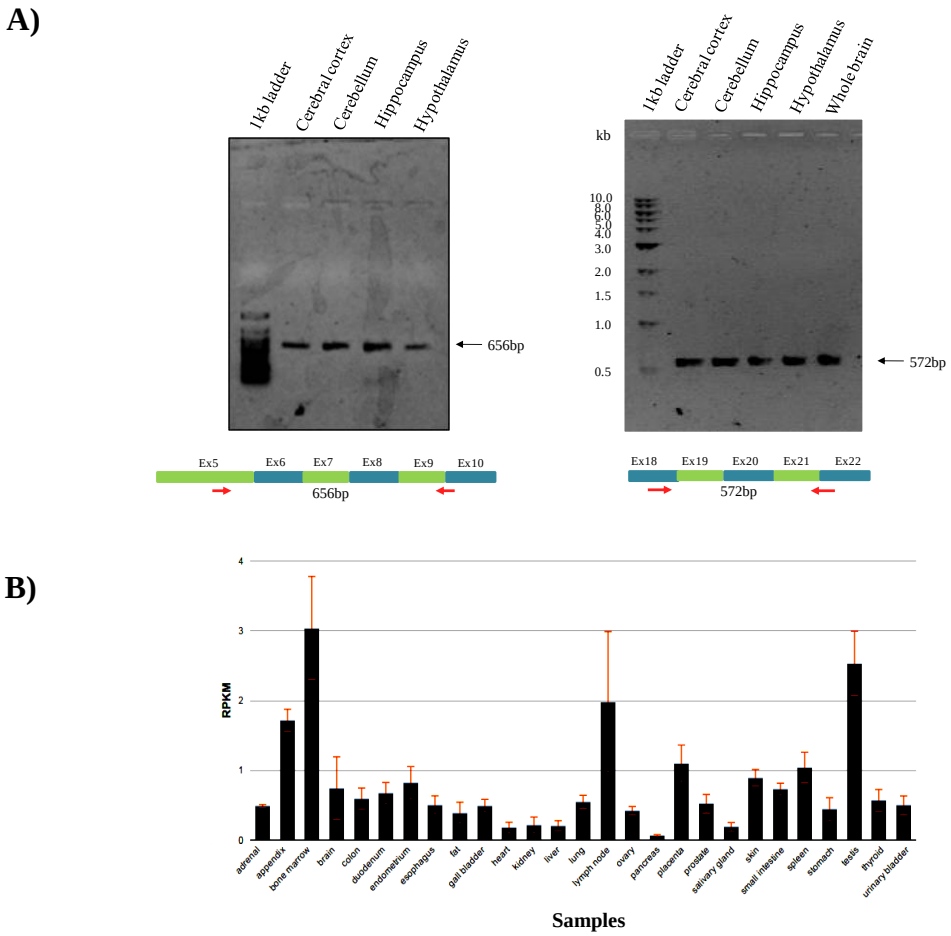


Figure 2.10: *ZGRF1* transcript expression. (A) Expression of *ZGRF1* transcript amplified from different regions of the human brain cDNA. (B) HPA RNA-seq profile of *ZGRF1* transcript across different tissues. RPKM: reads per kilobase million.

2.2.6. Expression analysis of ZGRF1 from cultured mammalian cells and human brain regions

Western analysis was performed on lysates from untransfected HEK293 cells and those transiently transfected with plasmids carrying the wild-type ZGRF1 or variants in the cDNA. The predicted molecular weight of ZGRF1 is 236.6kDa (http://web.expasy.org/compute_pi/). In this study, the untransfected and transfected proteins appeared at above the 250kDa marker. No significant difference in expression levels of the variants with respect to the wild- type ZGRF1 protein was observed (Figure 2.11A). ZGRF1 was found to be expressed in amygdala, basal ganglia, cerebellum, hippocampus, hypothalamus and temporal lobes of the brain. The protein migrated around the range of 250kDa (Figure 2.11B). Another band of lower intensity is visualized a little below 150kDa. This might correspond to products of the transcripts that remain undetected so far. Alternatively, it could represent non-specific cross-reactivity.

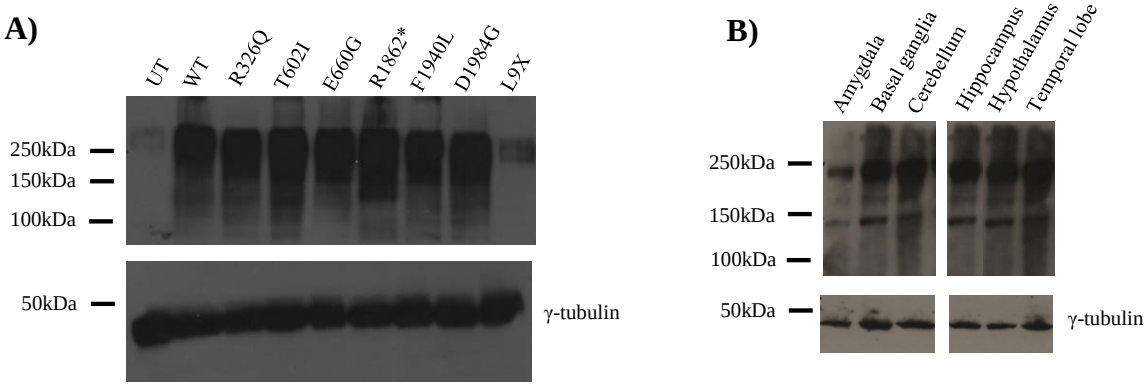


Figure 2.11: ZGRF1 expression in cultured mammalian cells and in brain regions. (A) Untransfected and over-expressed ZGRF1 wild-type and mutant proteins in HEK293 cells. (B) Expression of ZGRF1 in different regions of the human brain. The size of the ladder is represented on the left of the images.

2.2.7. ZGRF1 localize to cytoplasm, spindle poles and midbody in cultured mammalian cells

Immunolocalization analysis of ZGRF1 in cultured HEK293 cells showed an endogenous localization of the protein in the cytoplasm in non-dividing cells with nucleolar localization in a few of the cells. In mitotic cells, the protein was observed at the centrosome in the prophase stage and at the spindle poles in the metaphase stage. No discrete localization pattern was observed in the anaphase and cytokinesis stage cells where the protein appeared to be distributed throughout the cytoplasm. The protein localized at the midbody in the

telophase cells (Figure 2.12). In SH-SY5Y neuroblastoma cells and rat C6 glioma cells, ZGRF1 localized in the cytoplasm and could be detected at the spindles poles in dividing cells (Figure 2.13A). Localization of the protein to the spindle poles was confirmed by localization analysis with γ-tubulin with which it co-localized (Figure 2.13B and C).

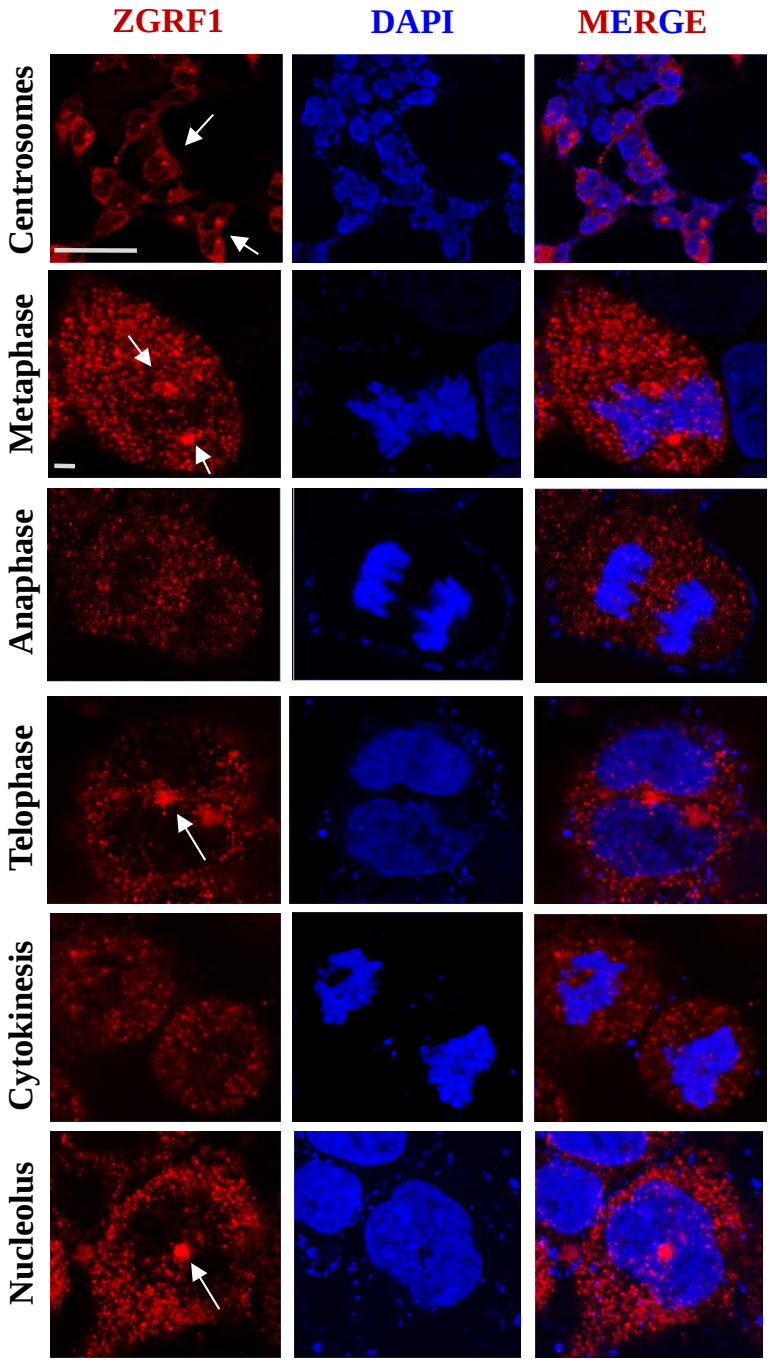


Figure 2.12: Endogenous localization of ZGRF1 in HEK293 cells. ZGRF1 is seen to localize in the cytoplasm and centrosomes of cells. In dividing cells, in metaphase stage it is localized at spindles poles. No distinct localization pattern was detected during anaphase and cytokinesis. It localized to midbody in the telophase stage. In some cells, nucleolar localization was observed. Scale bar=20µm, 5µm.

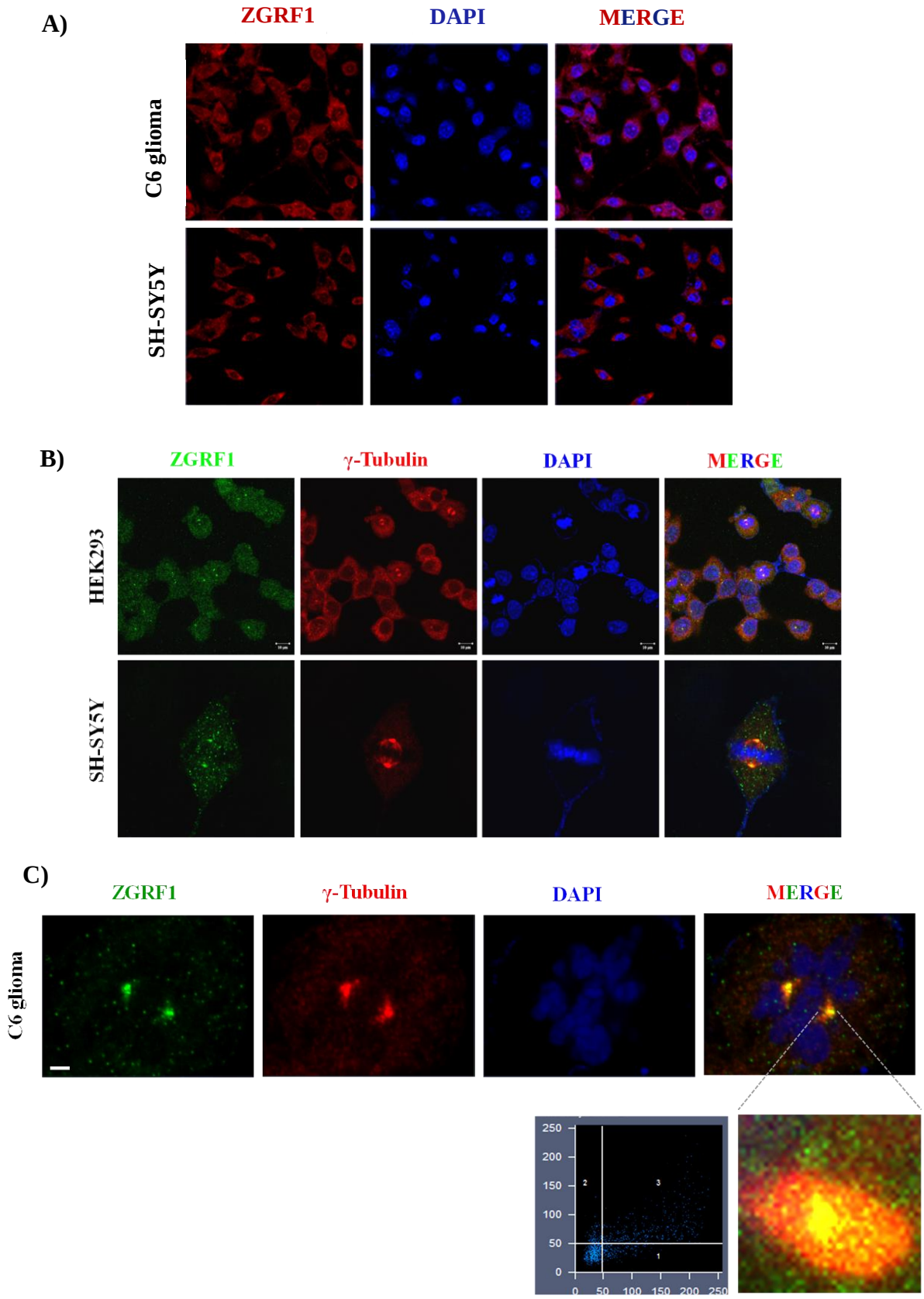


Figure 2.13: ZGRF1 and tubulin. (A) Endogenous localization of ZGRF1 in C6 glioma and SH-SY5Y cells. (B) Co-localization of ZGRF1 with γ-tubulin at spindle poles. (C) Zoomed image of the spindle pole slowing co-localization of ZGRF1 and γ-tubulin with the corresponding scatter plot for co-localization. Scale bar=1µm.

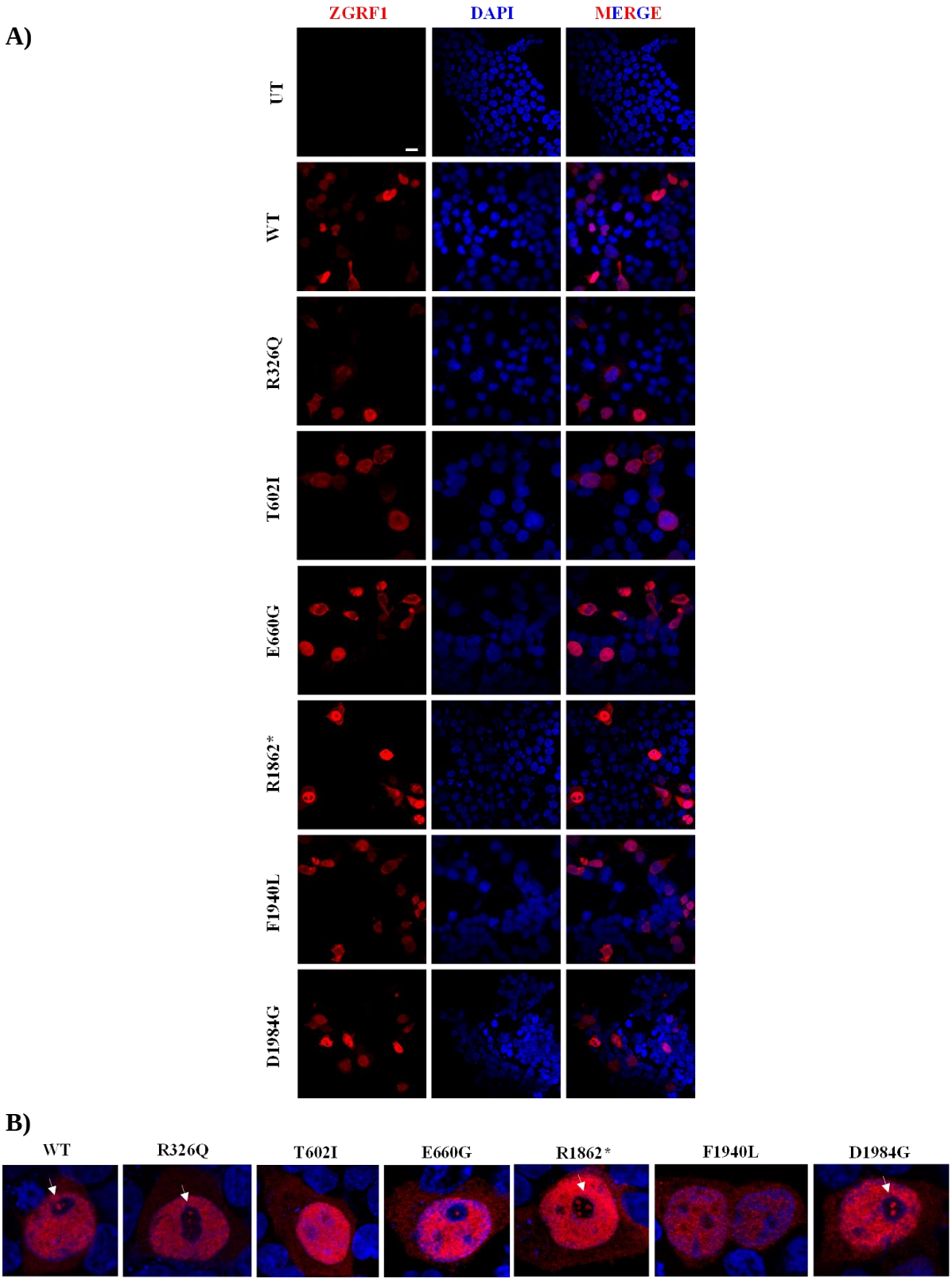


Figure 2.14: Localization of over-expressed ZGRF1 in HEK293 cells. (A) A field view of ZGRF1 localization for the wild-type and mutant proteins. The protein is localized majorly to the nucleus with a small fraction of cells exhibiting expression in cytoplasm and nucleolus. Scale bar=20µm. (B) Zoomed-in images of cells represented in part A, showing localization of the protein in the cell cytoplasm, nucleus and nucleolus for the different WT and mutant proteins. Arrows point towards the protein localization in the nucleolus.

2.2.8. ZGRF1 over-expression in HEK293 cells show no differential localization

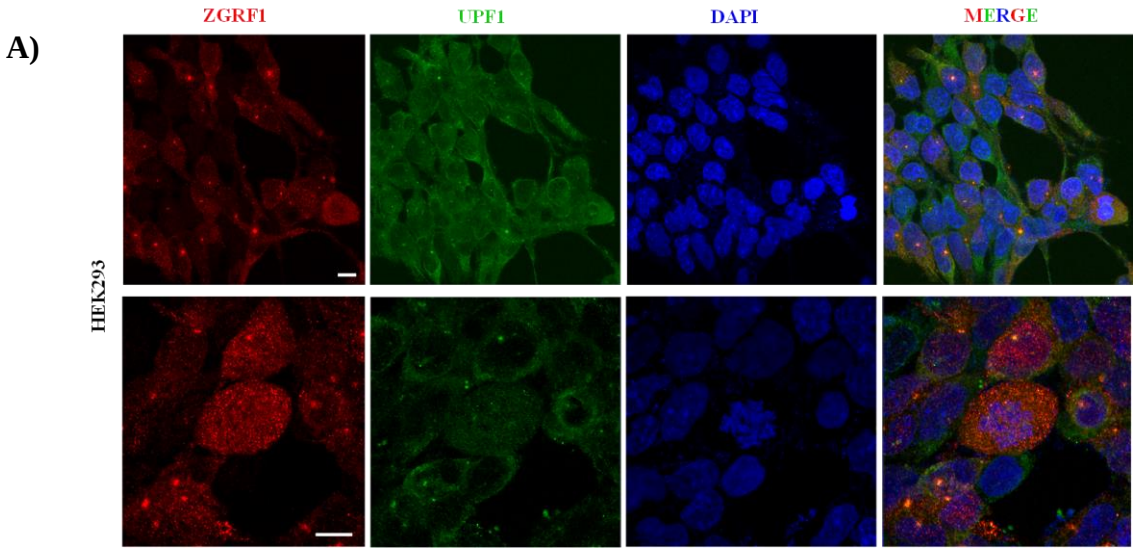
To examine the expression of ZGRF1 in a transient transfection system, HEK293 cells transfected with the wild type and variant- carrying ZGRF1 plasmids were fixed and immunostained. In the antibody staining experiments, the endogenous protein was detected at a 1:100 dilution of anti-ZGRF1 antibody. Localization of the overexpressed protein was assessed at 1:1000 dilution of the antibody. At this dilution, the endogenous protein in the untransfected (UT) control could not be detected. The transfected wild-type protein localized in a heterogeneous manner within the cell; majorly in the nucleus (237/250) with nucleolar speckles in some cells (4/250). In a smaller fraction of cells, it localized exclusively to the cytoplasm (9/250). The mutant proteins showed a similar pattern of localization with no distinct disturbance in the cell morphological features (Figure 2.14). This observation indicates that ZGRF1 mutant proteins present the same expression pattern as that of the wild-type protein, and their sub-cellular localizations do not seem to be impaired in any apparent manner.

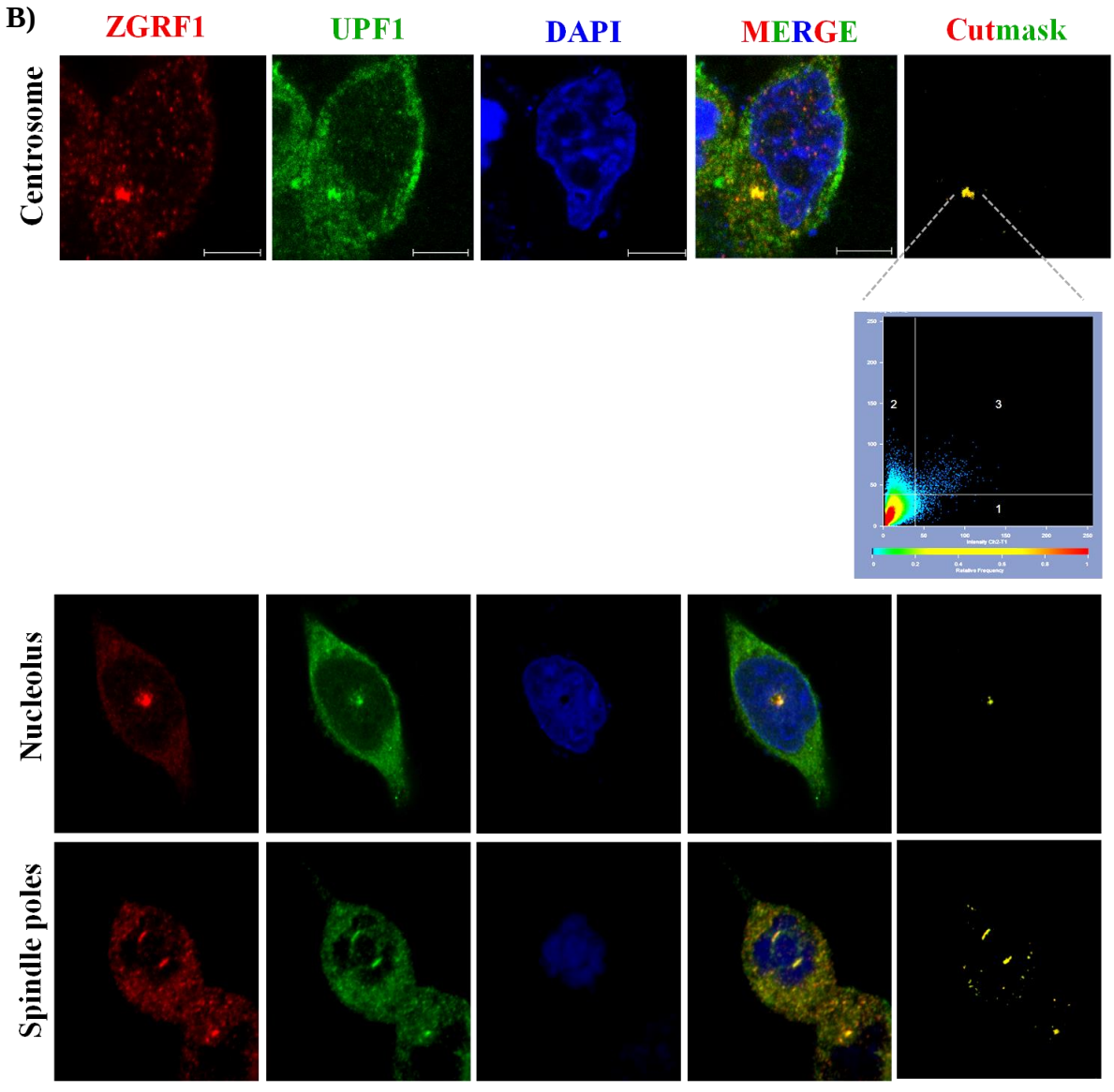
2.2.9. ZGRF1 co-localizes with its paralog UPF1

Immunostaining of HEK293 cells with anti-ZGRF1 antibody and anti-UPF1 Alexa 488 suggested the two proteins co-localize. Earlier studies have shown that hUPF1 (a DNA/RNA helicase) is a cytoplasmic protein (Applequist et al 1997) which is consistent with the fact that nonsense mediated decay (NMD) is a cytoplasmic process (Singh et al 2007). Both the proteins are present in the cell cytoplasm with a distinct co-localization at centrosomes in prophase cells (Figure 2.15A). They also co-localized in the cell nucleolus of some cells. Similar to its co-localization with γ-tubulin, ZGRF1 in dividing cells co-localized with UPF1 specifically at the spindle poles (Figure 2.15B). In dividing cells, UPF1 localized to spindle poles as well as spindles and co-localized with γ-tubulin (Figure 2.15C). UPF1's localization to spindles has not been previously reported. ZGRF1 and UPF1 co-localized in a similar pattern in the nucleolus and spindles of dividing C6 glioma and SH-SY5Y cells (Figures 2.15D and E).

The localization pattern of ZGRF1 independently and in alliance with UPF1 provides an indication that ZGRF1 might be an RNA- and DNA- binding protein that associates with UPF1, if not always, but to perform certain cellular functions. The specific localization of ZGRF1 across cell cycle stages and at spindle poles suggests that it might be involved in cell division and associate with DNA. Zinc fingers are known to bind both DNA as well as RNA. On the other hand, nucleolus is the centre for RNA-related activities. Detecting ZGRF1

in the nucleolus is an indication that it might be involved in RNA-related cellular functions.





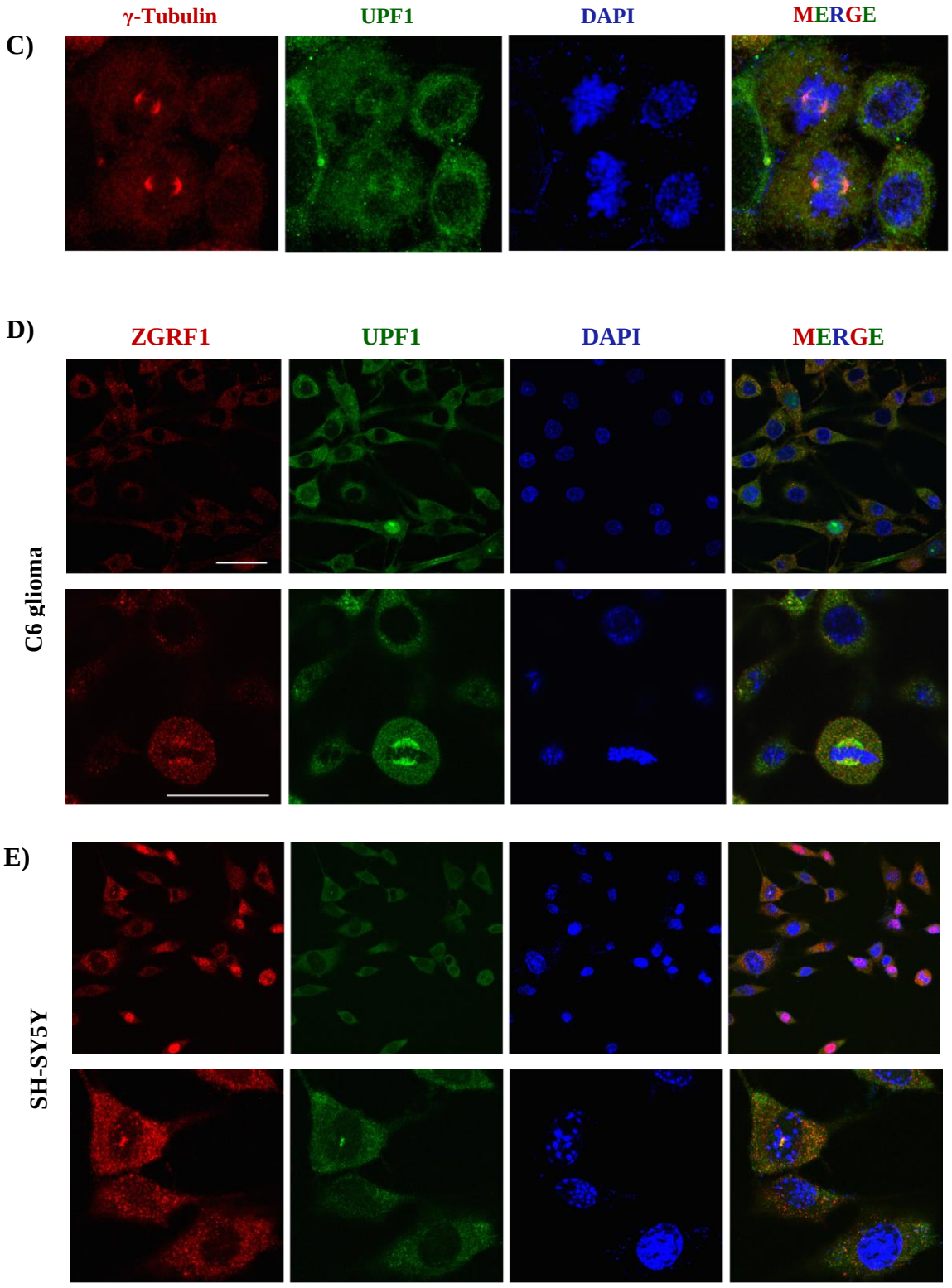
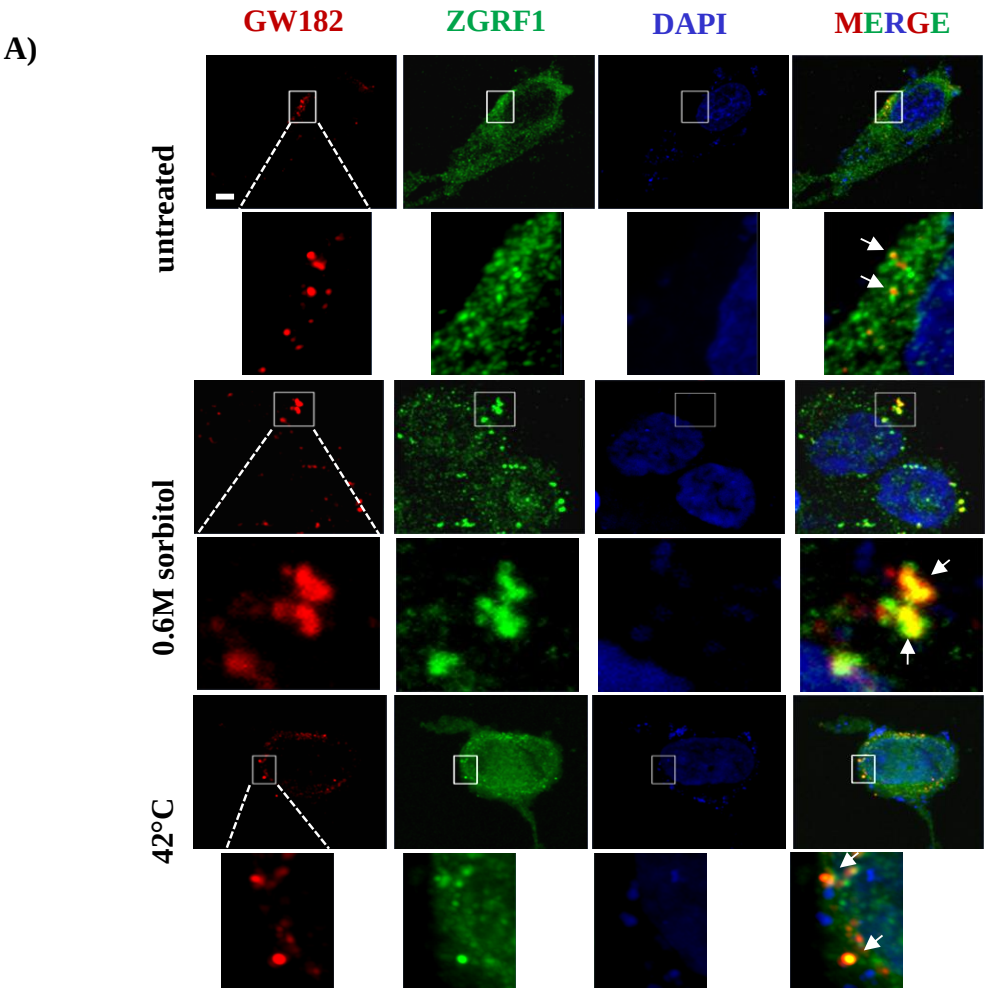


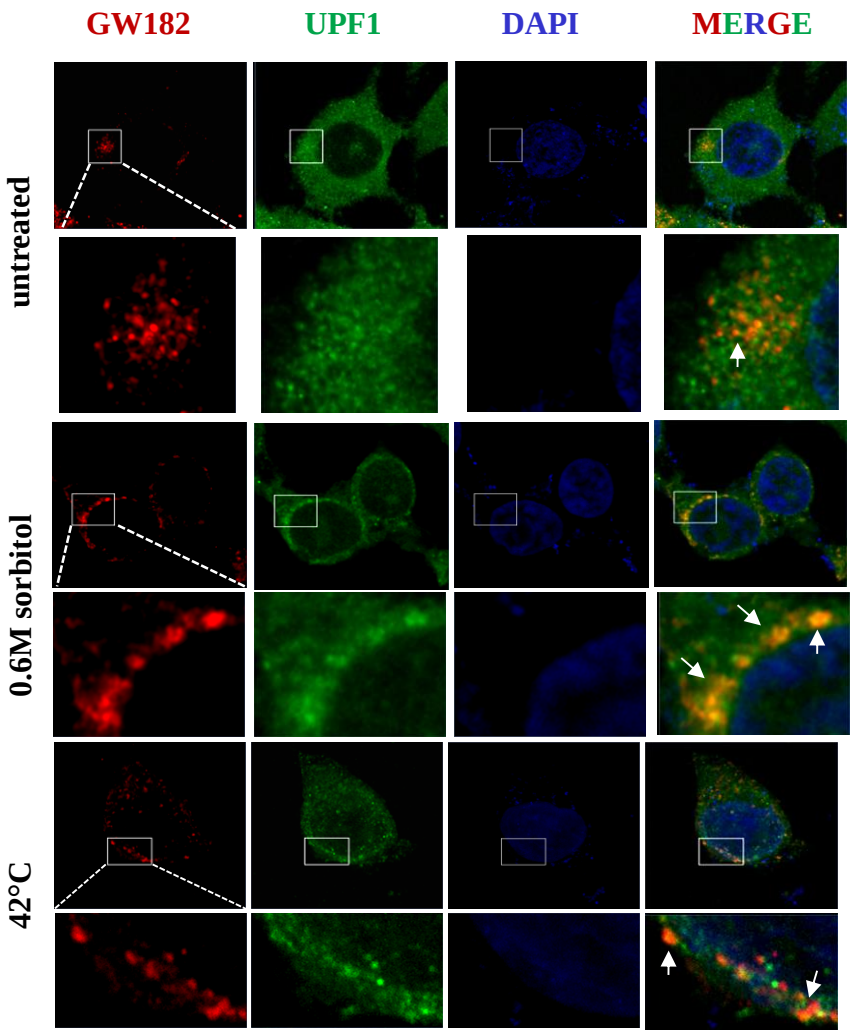
Figure 2.15: ZGRF1 and UPF1 co-localization. (A) A representative image of the endogenous pattern of ZGRF1 and UPF1 in HEK293 cells. Scale bar=10µm. (B) ZGRF1 and UPF1 in HEK293 cells are seen to co-localize in the nucleolus, at centrosomes in the prophase and at the spindle pole in metaphase stages. The co-localization is represented by a scatter plot. Scale bar=5µm. (C) Co-localization of UPF1 with γ-tubulin. (D) Endogenous ZGRF1 and UPF1 in C6 glioma cells. The two panels represent the two proteins co-localizing at lower and higher magnifications. The lower panel shows co-localization at the spindle pole in a dividing cell. (E) Endogenous proteins in SH-SY5Y cells. The upper panel represents co-localization at a lower magnification. The lower panel represents cytoplasmic and nucleolar co-localization of the two proteins. Scale bar=10µm.

2.2.10. ZGRF1 co- localize with P-bodies

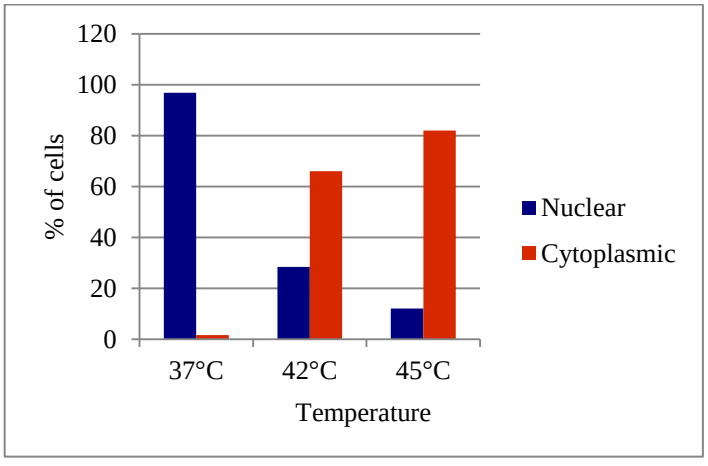
Cells when exposed to stress, redistribute cytoplasmic mRNA into cytoplasmic granules where they are either degraded or stored in a non-translatable form till the stress is removed (Bond 2006 for review). These foci or P-bodies (processing bodies) are known to be the sites for NMD. Sorbitol induces osmotic and oxidative stress (Dewey et al 2011). Such conditions of stress lead to increase in number of the P-bodies (Teixeira et al 2005). Increased temperature or heat stress also induces P-body formation (Cowart et al 2010). Upon stress, ZGRF1 co-localized with the processing bodies (Figure 2.16A). UPF1, the core NMD protein, used as a positive control in this experiment also co-localized with the P-bodies (Figure 2.16B). ZGRF1-wild-type over-expressing HEK293 cells when exposed to high temperature, showed an increase in the cytoplasmic localization of the protein (Figure 2.16C). The wild-type over-expressed protein showed partial co-localization with the P-bodies (Figure 2.16D). However, high levels of the over-expressed protein in the cytoplasm under stress, made it difficult to assay for co-localization of the two proteins. Hence, ZGRF1 mutants could not be assayed for their co-localization with P-bodies.



B)


GW182
UPF1
DAPI
MERGE
untreated
0.6M sorbitol
42°C

C)


120
100
80
60
40
20
0
% of cells
37°C
42°C
45°C
Temperature
Nuclear
Cytoplasmic

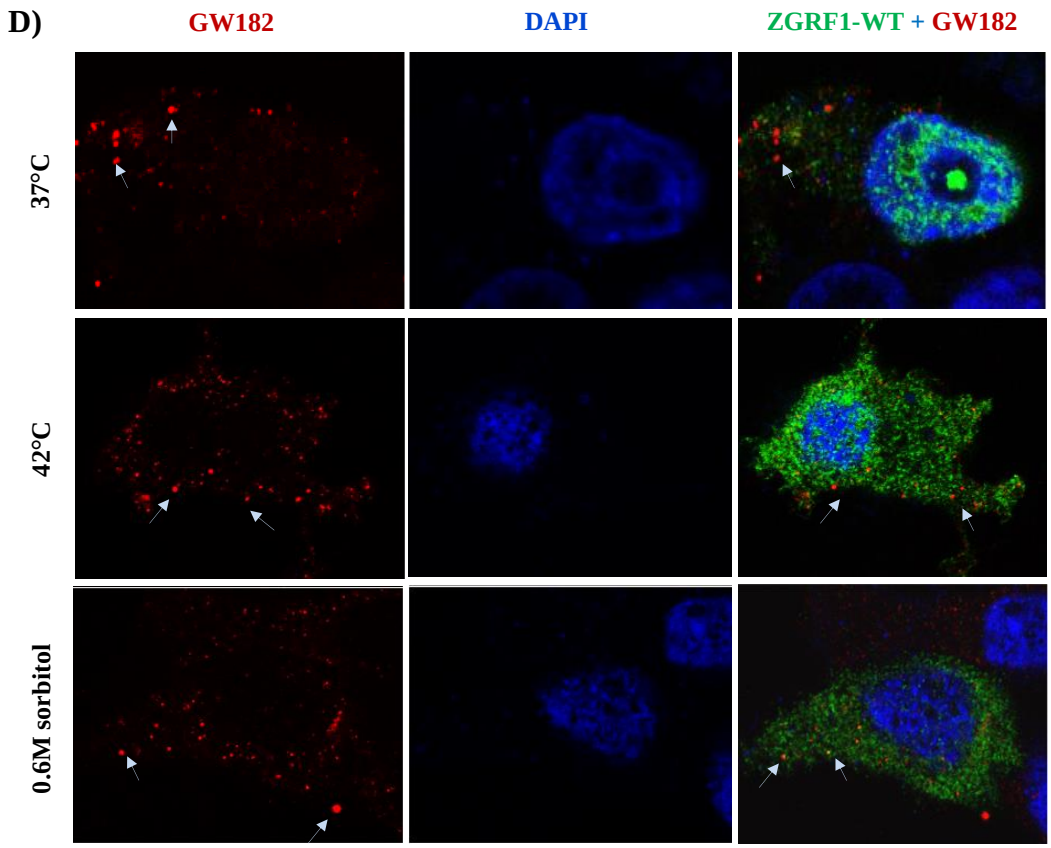


Figure 2.16: ZGRF1 and UPF1 localization with P-bodies. (A) Localization of endogenous ZGRF1 and P-bodies (marker GW182) under untreated, 0.6M sorbitol treated and high temperature conditions. Under untreated conditions, the P-bodies are smaller in size and of lesser number and have partial co-localization with ZGRF1. Under sorbitol and high temperature treatments, the P-bodies are bigger in size and ZGRF1 is seen to form cytoplasmic granular structure that co-localize with P-bodies. Scale bar=10µm. (B) Localization of endogenous UPF1 with P-bodies. Under untreated conditions, there is partial localization of UPF1 with the P-bodies. Upon sorbitol and high temperature treatments, the protein co-localizes distinctly with these cytoplasmic bodies. (C) ZGRF1-wildtype (WT) over-expressed cells when exposed to higher temperature, show greater localization of the protein in cytoplasm. (D) Co-localization assay for ZGRF1-WT over-expressed protein with P-bodies under different conditions of stress, including sorbitol and high temperature treatment. High levels of expression of ZGRF1-WT protein and increased cytoplasmic lozalization upon stress, made it difficult to assay for co-localization of the over-expressed proteins with P-bodies.

2.2.11. ZGRF1 mutants cause partial disruption in UPF1 localization to spindles

The endogenous co-localization study provided evidence for ZGRF1 to co-localize with UPF1. It has been previously reported that *Drosophila* cells mutant for UPF1 are unable to grow due to increased cell death and arrest in cell division (Avery et al 2011). This provides support for the involvement of UPF1 in cell division, noting that in our study, UPF1 was found to co-localize at spindles with γ-tubulin. With this in view, I was keen to examine if the ZGRF1 mutants had an influence on the localization of UPF1 at spindles. HEK293 cells transfected with wildtype and mutant ZGRF1 constructs were analysed in dividing cells co-stained with ZGRF1 and UPF1. A total of 10 cells were imaged per construct and 50 were visualized. In this experiment, only cells with two spindle poles have been shown to maintain a consistency in the data representation. The mutants Arg326Gln and Glu660Gly behaved like the wild-type protein and did not appear to influence the localization pattern of UPF1. Mutants Arg1862* and Phe1940Leu although stained for UPF1 in the poles and spindles, a faint disruption was observed in the overall staining for UPF1, especially in the microtubules. The mutants Thr602Ile and Asp1984Gly showed substantial disruption in the UPF1 staining at the pole and microtubules (Figure 2.17). These observations suggest that ZGRF1 associates with UPF1, and the mutants Thr602Ile and Asp1984Gly negatively influences the normal localization of UPF1 in a dividing cell.

2.2.12. ZGRF1 mutants induce spindle defects during mitosis

An observation was made during the study of the effect of ZGRF1 mutants on UPF1 localization to spindles. It was found that a number of cells expressing the mutated proteins exhibit defects related to the spindles, mostly multipolar and disorientation of poles. This initial observation prompted to examine the percentage of mitotic defects in cells with the wild-type and mutant proteins (n=180). Any mitotic cellular defects including increased or decreased number of poles, disoriented poles, and abnormal cytokinesis were taken into account. The most common defects were increased number of spindle poles and misalignment of poles with respect to the chromosomes (Figure 2.18A). Quantification of mitotic defects revealed higher proportion of mitotic defects in the cells carrying ZGRF1 mutant proteins as compared to the wild-type ZGRF1 (Figures 2.18B and C). The mutant Thr602Ile showed significant mitotic defects, although mitotic defects were observed in all the mutants. These results demonstrate cell-division defects in the presence of mutant ZGRF1, suggesting a role of this protein in normal cell division process.

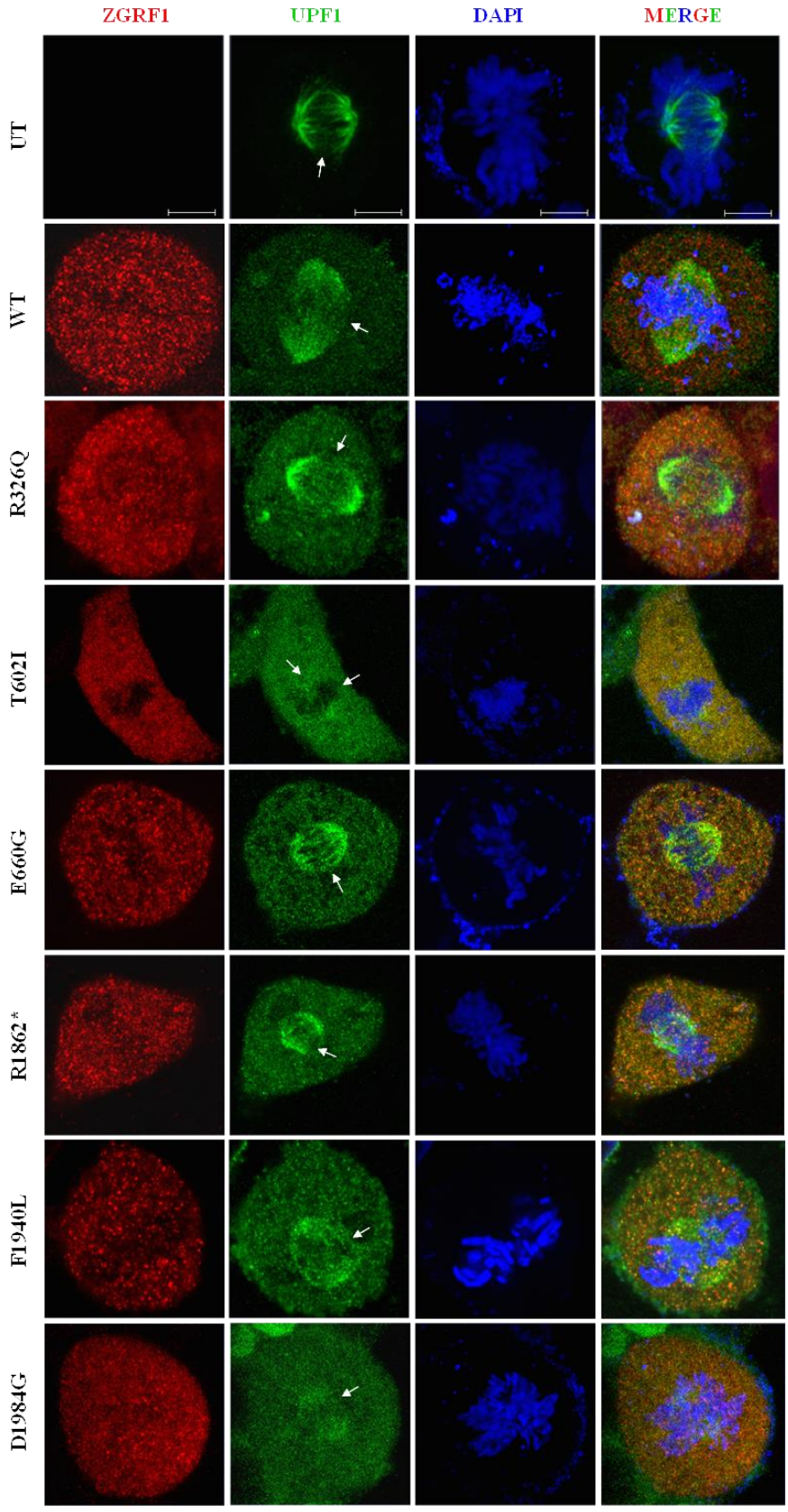
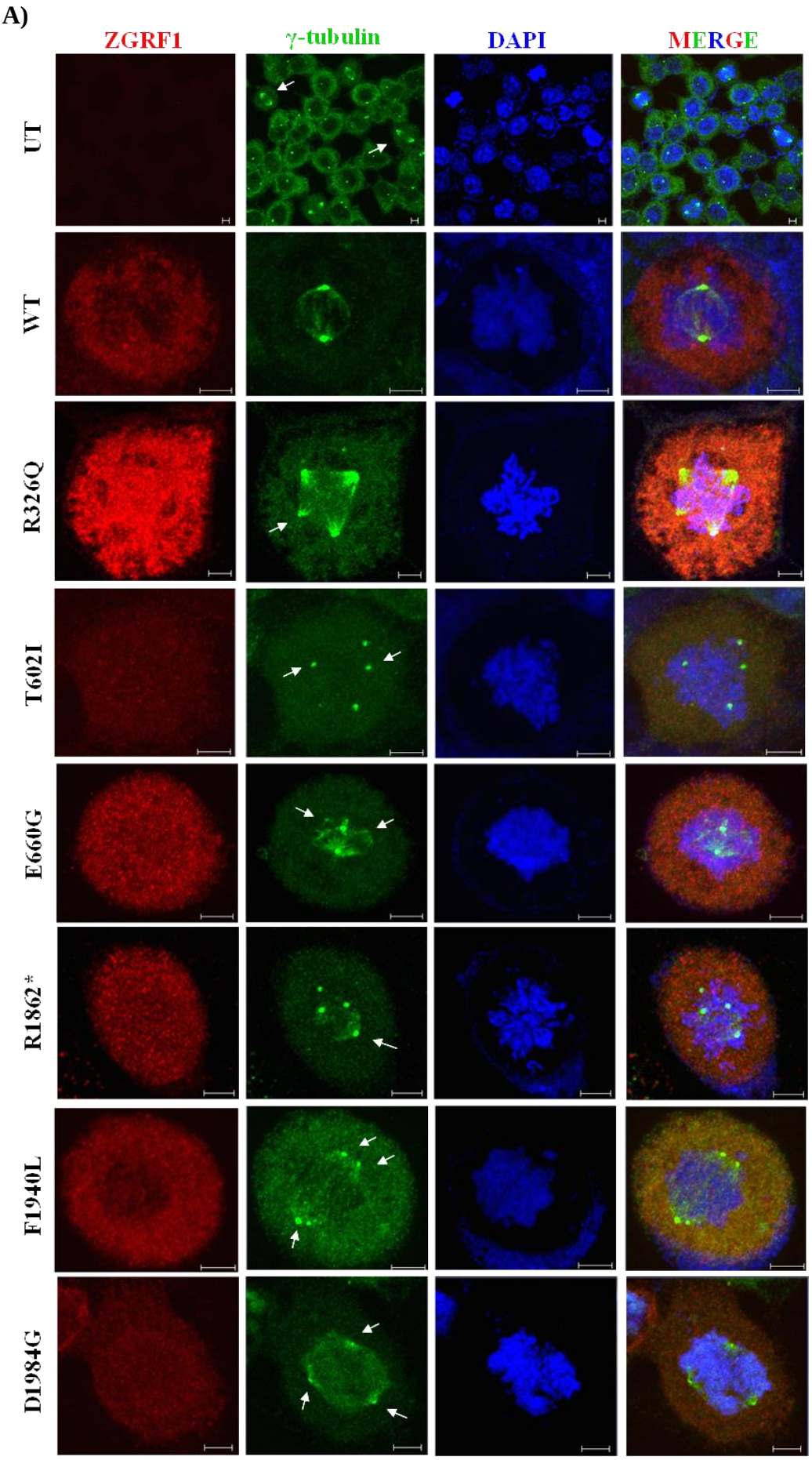
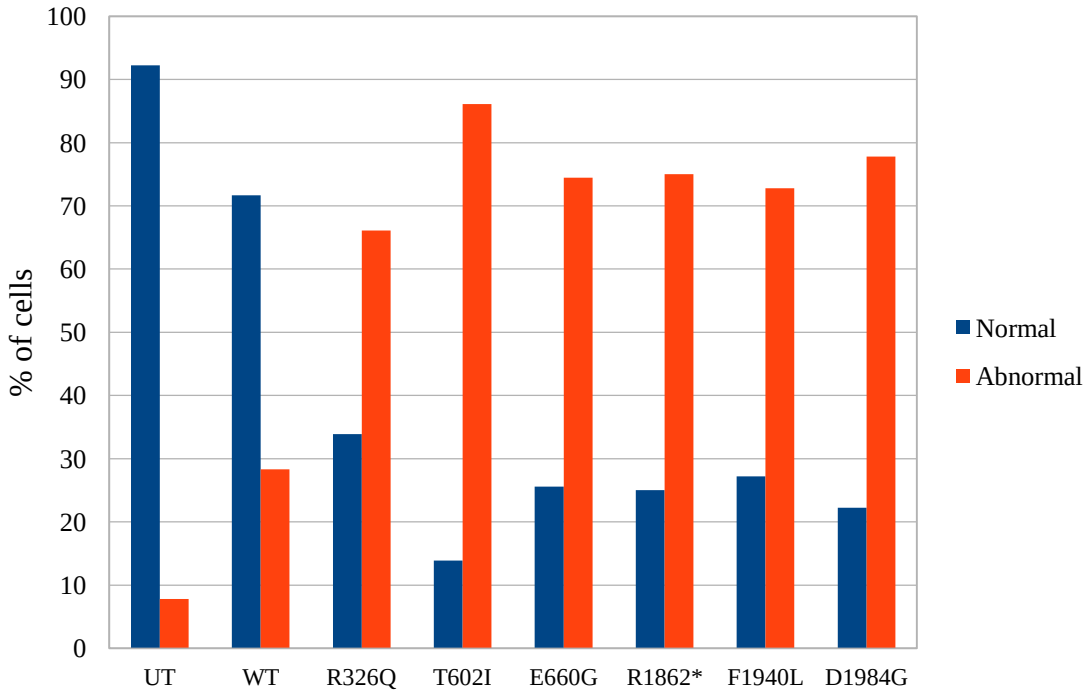


Figure 2.17: Effect of ZGRF1 mutants on UPF1 cellular localization. In the background of over-expressed ZGRF1-wildtype and -mutant proteins, UPF1 localization to spindles was unaffected by R326Q and E660G, partially disrupted by R1862* and F1940L and substantially affected by T602I and D1984G. Arrows point towards spindle poles and microtubules. Scale bar=5µm.


A)
ZGRF1
γ-tubulin
DAPI
MERGE
UT
WT
R326Q
T602I
E660G
R1862*
F1940L
D1984G

B)



C)

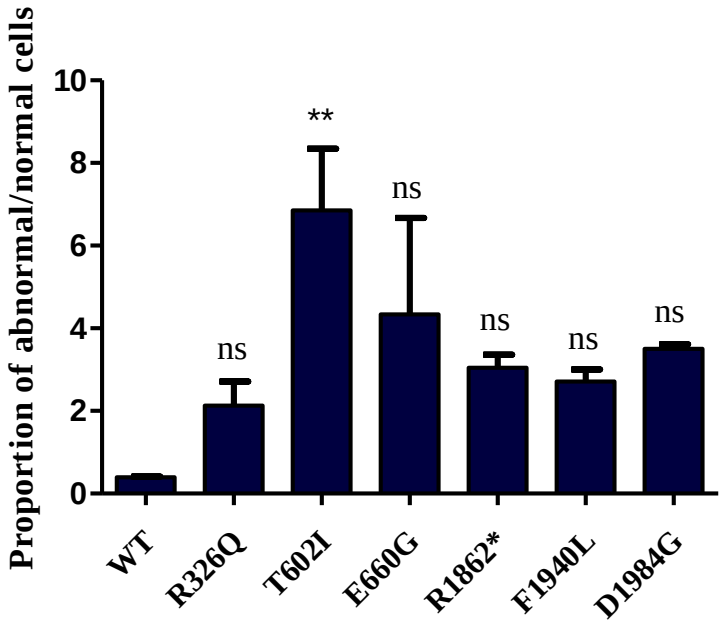


Figure 2.18: Mitotic defects in cells with ZGRF1 mutants. (A) Representative image showing the defects in the spindles of cells carrying ZGRF1 mutants. Arrows point towards spindle poles. Scale bar=5µm. (B) Percentage of normal and abnormal cells (with mitotic defects) in the untransfected, wild-type and the six ZGRF1 mutants examined (n=180). (C) Statistical quantification of the number of abnormal cells (with mitotic defects) to normal cells in the wild-type and mutants (n=180). ** $P \leq 0.001$. ns = not significant.

2.2.13. ZGRF1 transports to the nucleus upon DNA damage

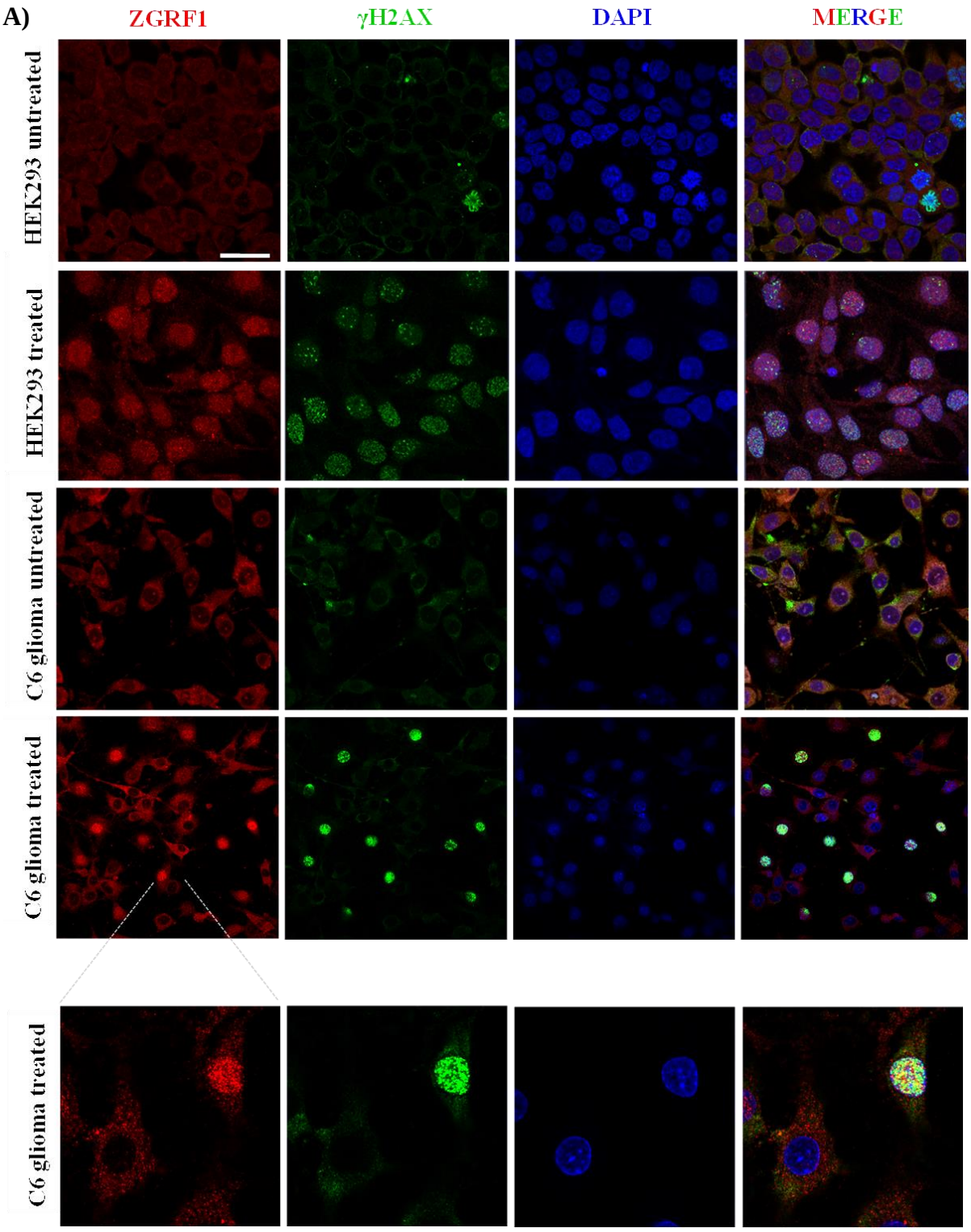
A genome-wide shRNA screen for sensitivity to mitomycin C had identified ZGRF1 as a protein involved in DNA crosslink resistance (Smogorzewska et al 2010). Another study of a genome-wide siRNA screen identified ZGRF1 as a regulator of the mammalian homologous recombination machinery (Adamson et al 2012). The candidate genes pursued in these studies were identified as components of DNA-damage response. Proteins involved in cellular processes of crosslink resistance and homologous recombination integrate into DNA damage response and repair pathways.

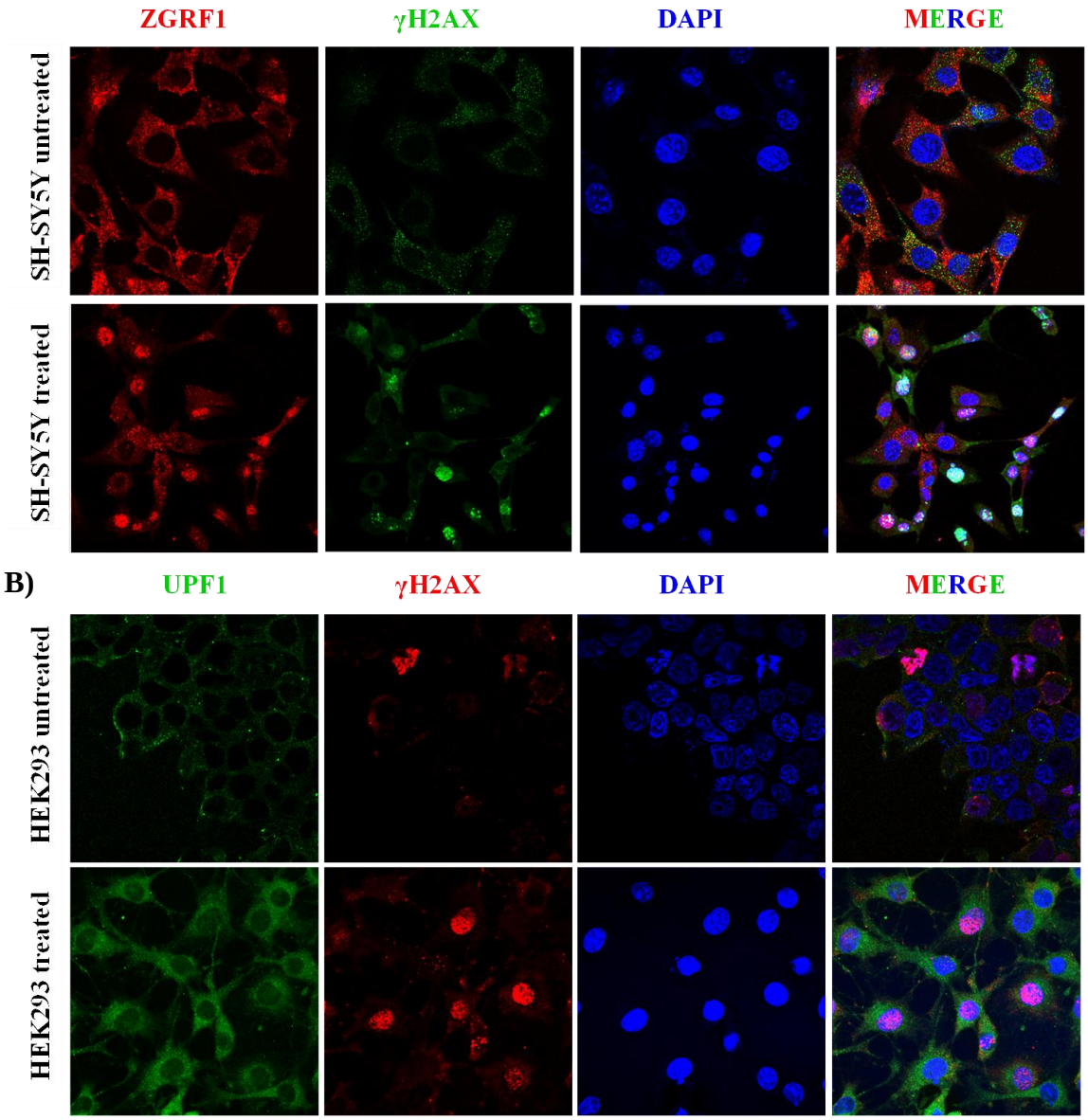
To determine any effect of DNA damage on ZGRF1 localization, HEK293 cells were treated with mitomycin C. Upon inducing damage, a distinct staining was observed for the protein in the nucleus, while a low level of staining was seen in the cytoplasm (Figure 2.19A). Although under the conditions of inducing damage randomly in the cells, some did escape damage. In these cells, the protein remained localized to cytoplasm. This indicates that translocation of majority of the protein to nuclei is in response to DNA damage. Similar observations were made in C6 glioma and SH-SY5Y cells (Figure 2.19A). UPF1 on the other hand did not show this phenotype (Figure 2.19B). A total of 300 cells were counted and all the damaged cells accounted for ZGRF1 localization in the nucleus (Figure 2.19C).

HEK293 cells over-expressed with the wild-type or mutant proteins did not show any distinct phenotype, as the over-expressed protein majorly localizes to the nucleus. In the wild-type over-expression, the protein appeared in the nucleus with foci for damage stained by γH2AX. There was no difference in the localization of the protein upon inducing damage (Figure 2.19D). I aimed to score for an increased number of damage foci or their increased intensity in the mutant proteins with respect to the wild-type over-expression, with an idea of a condition where the mutant's crosslink resistance ability is compromised. However, as the event of DNA damage is random, these foci could not be scored.

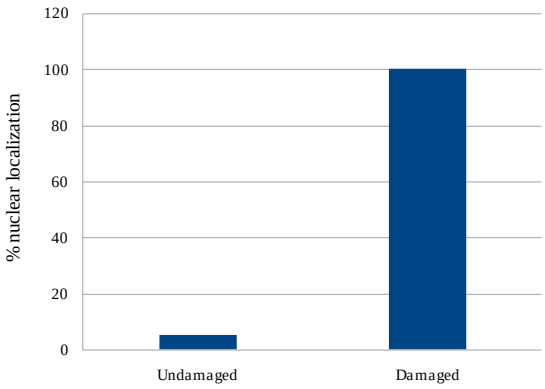
HEK293 cells transfected with ZGRF1-wild-type-tGFP and empty vector-GFP were micro-irradiated in specific regions of the cell nucleus to examine if the protein localized to these regions of DNA damage. In the ZGRF1-WT, the protein did not accumulate at the regions of irradiation. But, the GFP signal intensity which is the readout of GFP-tagged protein levels in the cell, gradually increased in the nucleus from 2 minutes post damage induction to 20 minutes, and then the intensity remained constant up to 60 minutes (Figures 2.20A, B). On the other hand, the GFP intensity did not change post induction of damage in the empty

vector control (Figures 2.20C and D). This suggests that the increase in GFP signal intensity in the nucleus is a readout of the increase of ZGRF1 protein in the cell nucleus in response to damage. This experiment adds evidence to the role of this protein in the cell nuclei upon induction of DNA damage.





C)



D)

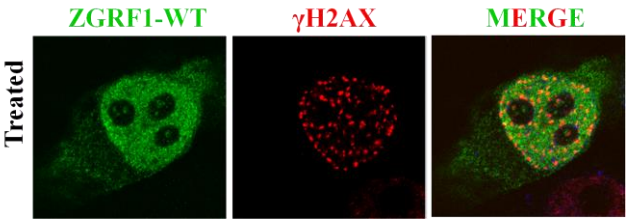


Figure 2.19: ZGRF1 response to damage.
(A) Representation of HEK293, C6 glioma and SH-SY5Y cells untreated and treated with mitomycin C, showing the nuclear localization of ZGRF1 in damaged cells. Scale bar=10μm. (B) UPF1 in cells shows no similar phenotype. It remains cytoplasmic after induction of damage. (C) Graphical representation of the percentage of undamaged and damaged cells with nuclear localization of ZGRF1. (D) ZGRF1-WT overexpressed cells treated with mitomycin C.

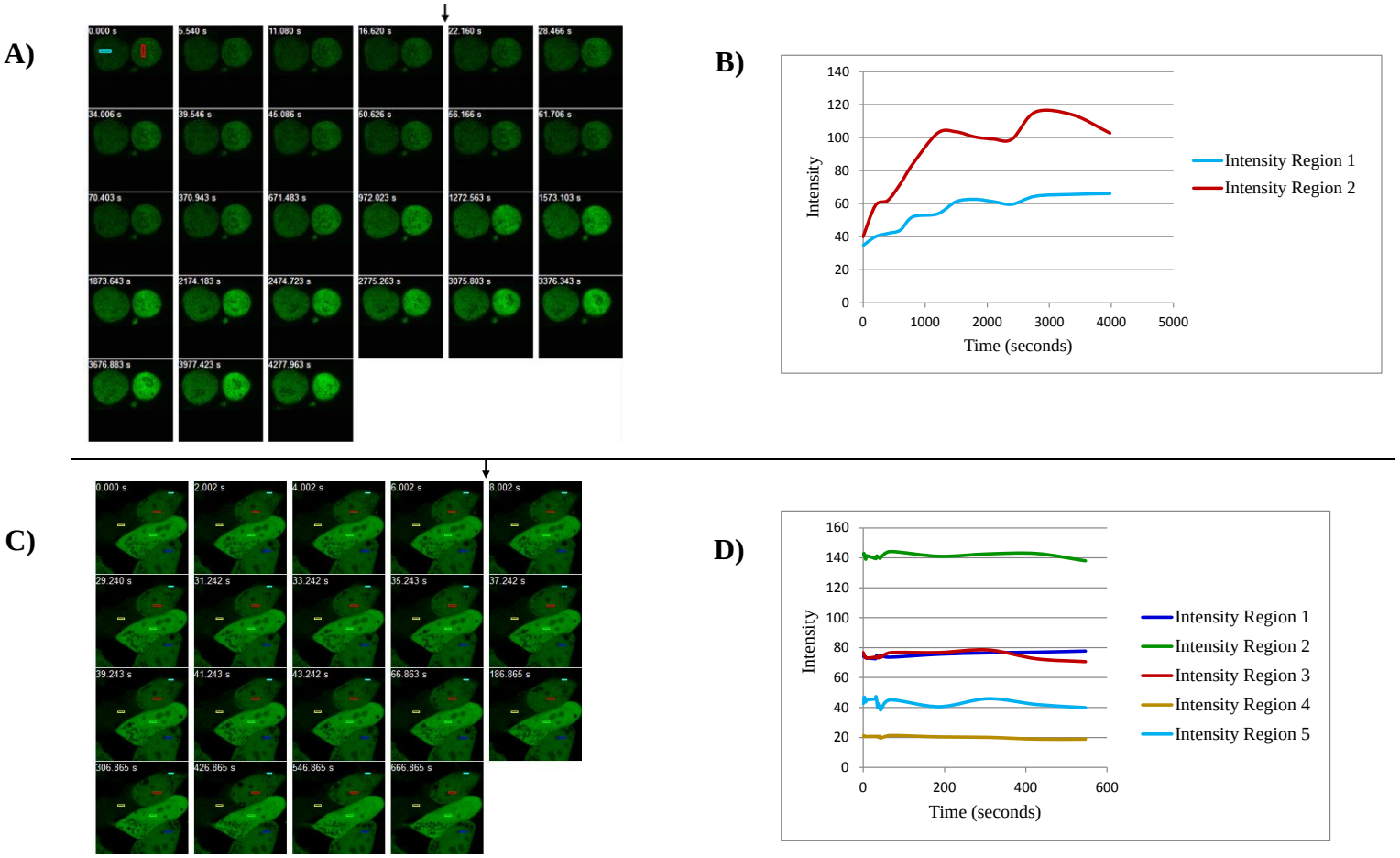


Figure 2.20: Laser irradiation of ZGRF1-GFP and empty vector containing cells. (A) Representation of ZGRF1 cells with regions marked for irradiation represented for a 60-minutes time interval. The arrow at 22-seconds of the image represent the point of irradiation. The GFP intensity as a readout of ZGRF1 levels in the cells increase in the cell nuclei upon irradiation. (B) Graphical representation of change in the intensity profiles in cells after irradiation. The intensity gradually increases up to 1200 seconds in the nuclei post irradiation, and is maintained almost constant at high levels in the nuclei. (C) Empty vector cells with regions marked for irradiation represented for a 10-minutes time interval. The arrow represents the time at irradiation. (D) Graphical representation of change in the intensity profiles in cells after irradiation. No significant change was observed in the GFP intensity profile.

2.3. Discussion

In this chapter, I have presented the whole exome-based genetic analysis of a family HWE227 with several of its members affected with hot water epilepsy, to identify the potential causative gene for the disorder at 4q24-q28. In addition, certain functional aspects of the protein ZGRF1, found to carry a potentially pathogenic variant in the family, were studied.

Whole exome-based sequencing was carried out for two affected individuals of HWE227- with a previously reported locus between markers D4S1572 and D4S2277 at chromosome 4q24-q28 (Ratnapriya et al 2009a). A variant, c.1805C>T in *ZGRF1* was identified to be segregating with the epilepsy phenotype in the family. The complete transcript analyses of *ZGRF1* in 288 HWE patients helped identify five additional rare variants. Bioinformatic analyses of ZGRF1 provided clues to its homology, at the C-terminus end, with domains of various DNA/RNA helicases. It was found to have the highest homology to UPF1 (Up-frameshift 1), a protein known to be involved in nonsense-mediated mRNA decay. ZGRF1 currently belongs to the family of UPF1-like RNA helicases. No paralogs could be identified for the N-terminal portion of the protein.

Immunostaining studies showed localization of ZGRF1 in cytoplasm, nucleoli, centrosomes and spindle poles where it co-localized with UPF1. From this observation, it was speculated that ZGRF1 might follow UPF1's functional footprint. Localization of ZGRF1 to nucleoli which is the centre for RNA activity, and spindle poles which are associated with cell division, suggested that it might exhibit a DNA and RNA-binding property, similar to its paralog. It was then of interest to see if ZGRF1 localized to the hub for mRNA decay: the cytoplasmic P-bodies. UPF1, the nearest paralog of ZGRF1, is known to associate with P-bodies (Brogna et al 2008, Mocquet et al 2012, Moreno et al 2013). Under oxidative and osmotic stress, ZGRF1 exhibited an increased localization at P-bodies, suggesting possible involvement of the protein in NMD pathway. With an initial evidence for possible association of ZGRF1 and UPF1, it was interesting to examine if ZGRF1 variants influence UPF1 localization in dividing cells. Two variants Thr602Ile and Asp1984Gly were found to abrogate the spindle localization of UPF1, suggesting that UPF1 localization at spindles might be influenced by ZGRF1 to an extent. ZGRF1 variants exhibited microtubule-based cell cycle defects indicating a role of the protein in the cell division pathway. Two studies in the literature provided us hints for possible involvement of ZGRF1 in the DNA damage response and DNA repair pathways. Based on these cues, a DNA damage experiment was performed which indicated that ZGRF1 distinctly responded to genotoxic stress by translocating to the

nuclei of damaged cells. This indicates a nucleocytoplasmic shuttle activity of the protein and its role in the nucleus, upon damage. Taken together, these studies provide prima facie evidence of involvement of ZGRF1 in the NMD pathway, cell division pathway and DNA damage/ repair pathways.

The biological functions of ZGRF1 remain largely unexplored. Only a few pointers in the literature are available regarding this gene and protein. Two candidate variants in *ZGRF1*, rs76187047 (c.4087G>A, p.Glu1363Lys) and rs61745597 (c.142C>A, p. Leu48Met) have been identified in a multigenerational family with childhood apraxia of speech (Peter et al 2016). ZGRF1 has been linked to alcohol dependency, psychological and personality disorders (Kalsi et al 2010). *ZGRF1*, c.1504A>G (p.Met502Val) has been identified in three schizophrenic patients in a large-scale exome study (Need et al 2012). These reports suggest a role of ZGRF1 in the human brain.

Similarity of domain architecture with UPF1 provides clues for ZGRF1's possible functional similarity to UPF1, and co-localization with P-bodies added evidence to its potential involvement in NMD. The NMD pathway aids or aggravates numerous neurological and psychiatric disorders. Premature translation-termination codons (PTCs) that direct mRNAs to NMD are responsible for about one-third of inherited disorders including genetic epilepsies (Kang et al 2009). Disorders due to UPF1 mutations reported to date are mostly attributed to the involvement of NMD pathway. In a patient with severe case of idiopathic intellectual disability (ID), brain structural defects, and seizures, double heterozygous mutations in *UPF1* and *SQSTM1* (*Sequestosome-1*) is found; UPF1 is speculated to aggravate the phenotype probably caused by *SQSTM1* (Zahir et al 2017). UPF1 highly expresses in resected hippocampus of patients with intractable temporal lobe epilepsy (Mooney et al 2017). Increased UPF1 levels was found in hippocampus of mice post-induction of status epilepticus, and inhibition of NMD resulted in reduced spontaneous seizures post-induction (Mooney et al 2017); suggesting the role of NMD and UPF1 in epileptogenesis. Premature termination codons (PTCs) in GABA receptor that direct to the NMD pathway have been associated with genetic epilepsy (Kang et al 2010) and a spectrum of epilepsy syndromes ranging from generalized absence epilepsy to Dravet syndrome (Kang and Macdonald 2016).

ZGRF1 identified in HWE227 to be potentially causative, is an interesting gene belonging to the superfamily of SF1 RNA helicases and family of UPF1-like RNA helicases. SF1 RNA helicases have roles in almost all aspects of nucleic acid metabolism (Singleton et al 2007).

Mutations in *SETX* (*Senataxin*), a UPF1-like RNA helicase causes neurological disorders such as amyotrophic lateral sclerosis (Hirano et al 2011) and spinocerebellar ataxia 1 (Saracchi et al 2014). Mice haploinsufficient for an RNA-binding protein NOVA (Neuro-Oncological Ventral Antigen), involved in splicing and local translation in the neurons, exhibit spontaneous epilepsy (Eom et al 2013).

ZGRF1 was one of the candidates found in two large-scale studies to identify genes involved in resistance to DNA cross-linking agents (Smogorzewska et al 2010), and genes that regulate the mammalian homologous recombination (HR) machinery by repairing double strand breaks (Adamson et al 2012). In our experiments, ZGRF1 was seen to respond to genotoxic stress by shuttling from the cytoplasm to the nucleus. While this observation suggests ZGRF1's role in DNA repair pathway, it does not suggest how ZGRF1 functions in DNA repair. The importance of DNA damage and repair pathway in the brain has been implicated in several neurological disorders. Mutations in *PNKP* (Bifunctional polynucleotide phosphatase/kinase) with defects in DNA repair is involved in *m*icro*c*ephaly, early-onset, intractable seizures and developmental delay (MCSZ), and mild epilepsy (Shen et al 2010, Poulton et al 2013). Mutations in *ATXN3* (*Ataxin 3*) causing spinocerebellar ataxia type 3 (SCA3) interacts with PNKP and integrates in the DNA damage and repair pathway (Kawaguchi et al 1994, Araujo et al 2011). Additionally, mRNA processing factors including surveillance factors play a role to prevent DNA damage (Paulsen et al 2009).

Several proteins in the brain play a role in its development by regulating cell cycle through microtubule- based cytoskeletal dynamics. Distinct localization of ZGRF1 across different stages of cell division represents its role in mitosis/ cell cycle. A number of cell division proteins are known to cause epilepsy. Mutations in *EFHC1* (EF-Hand Containing 1) results in mitotic defects in juvenile myoclonic epilepsy (Raju et al 2017). Mutations in *DCX* (Doublecortin), a microtubule- associated protein results in severe epilepsy in humans (Kerjan and Gleeson 2007). The processes of NMD, cell division and DNA damage/ repair pathways are integral to the functioning of a normal brain. Mutations in proteins involved in these pathways cause several neurological disorders, including epilepsy. Protecting the brain from mRNAs with premature stop codons, DNA damages including those that arise during a seizure event (chapter by Crowe and Kondratyev 2009, Qiang et al 2016), and miscreants during cell division are essential for maintaining the brain physiology.

Hot water epilepsy is a sensory epilepsy syndrome wherein seizures are triggered by the stimuli of contact with hot water. Seizures are mostly focal, possibly of temporal lobe origin and results in impaired awareness. In about one-third of the cases, these seizures evolve to bilateral or tonic-clonic seizures (Karan et al 2017). It has been found to also have a psychological aspect to it (Venkataramiah and Ghorpade 2013). Patients manifest visual and auditory hallucinations, sense of fear, dazed appearance, irrelevant speech, and complex repeated behaviour (Satishchandra 2003). Genes involved in the causation of HWE could be expected to play a role for neurological and neuropsychiatric phenotypes. *ZGRF1* is an interesting gene, evidences for which are becoming available in literature for its role in the brain. In this study, DNA damage experiments conducted suggest ZGRF1's role in the nuclei of damaged cells. Determining UPF1 as the nearest paralog of ZGRF1 opened an avenue for me, to examine the role of ZGRF1 in brain physiology. Co-localization of ZGRF1 and UPF1 provided evidence for a probable functional similarity among the two proteins. This directed examining ZGRF1 in the P-bodies, to which it co-localizes; indicating its potential role in NMD. Two of the ZGRF1 variants disrupted UPF1 localization to the spindles, suggesting ZGRF1 could possibly influence UPF1's localization during cell division. ZGRF1 variants were found to exhibit microtubule-based cell defects. This finding is useful to guide our thinking regarding the role of ZGRF1 in brain and its involvement in epilepsy. This study has been limited to assays in cultured cells and microscopic observations. Functional interactions employing biochemical techniques have not been possible to attempt. Experiments with P-bodies for NMD and mitomycin-induced DNA damage could not be attempted in ZGRF1 variants due to technical limitations. It would be worthy to perform globin gene based NMD assay and comet assay for DNA damage/repair in ZGRF1 wild-type and mutant cells to further the understanding of any comprises in mutant protein function. A transcriptome-based study may help to examine ZGRF1's regulatory and interactive network. A comparative analysis can be further undertaken among wild-type and mutant proteins to demonstrate differences, if any.

In summary, the findings presented in this chapter provide initial evidence for the role of *ZGRF1* variants in HWE. ZGRF1 is a multifunctional protein potentially involved in three molecular pathways. Functional consequences of the *ZGRF1* variants in cell-based assays suggest subtle cellular defects in neuronal cells may contribute to pathophysiology of the disorder.

Chapter 3

Towards gene identification at hot water epilepsy locus at 10q21.3-q22.3

Summary

In this chapter, I present experiments conducted for identification of potentially causative gene for hot water epilepsy in a family, HWE150 wherein, the genetic locus is known to map to chromosome 10q21.3-q22.3. HWE150 is a four-generation family with 10 affected individuals. A genome-wide linkage study carried out previously to identify the genetic locus for the disorder provided highest parametric lod score of Z_{max}=3.17 at θ=0 for the marker D10S412 at 10q21. The critical genomic interval in Family 150 was located between D10S581 and D10S201 spanning about 15Mb sequence length at 10q21.3-q22.3. I carried out whole exome sequencing analysis for an affected father-son pair to identify the gene responsible for the disorder in the family. Among the 534 variants obtained, three rare variants, c.1171G>A in *FUT11* (fucosyltransferase 11), c.5165-58C>T in *DLG5* (Discs Large Homologue 5) and c.1248+57C>T in *NDST2* (N-deacetylase and N-sulfotransferase 2) were observed, segregating with the epilepsy phenotype in the family. *In-silico* analysis suggested *FUT11,* c.1171G>A to be a damaging variant. To gather additional evidence for the involvement of *FUT11* in HWE, the gene was sequenced in 288 unrelated, ethnicity-matched hot water epilepsy patients. This search found one new, rare coding variant c.1448A>G in the gene. No apparent defects in cellular localization were observed in the wild-type and mutant FUT11 proteins transiently overexpressed in cultured HEK293 cells.

This chapter describes the identification of potentially causative gene for hot water epilepsy in a multi-affected four generation family with the disease locus mapping to chromosome 10q21.3-q22.3. Whole exome sequencing analysis in this 15Mb critical genomic region led to the identification of a new, rare variant in the *FUT11* gene.

3.1. Materials and methods

3.1.1. Family ascertainment and clinical evaluation

HWE150 (Figure 3.1), a four-generation family with hot water epilepsy was ascertained from the Department of Neurology, National Institute of Mental Health and Neurosciences

(NIMHANS). This family comprises 10 affected and 20 apparently unaffected members. The family was recruited through the proband III:10 who was diagnosed with HWE at the age of 15 years. His interictal EEG showed generalized discharges arising from the left temporal region of the brain. He was treated with intermittent prophylactic clobazam therapy. Complex partial seizures (CPS) were reported in all the affected members of the family. HWE in the Family 150 transmitted in an apparently autosomal dominant mode with incomplete penetrance. Informed written consent was obtained from all the members participating in the study. This study had approval of bioethics committees of both the institutions. A ten millilitre blood sample was collected from all participating members, and DNA was isolated from peripheral white blood cells using a standard method (Sambrook and Russell 2001).

3.1.2. Ascertainment of HWE and control cohorts

A group of 288 apparently unrelated hot water epilepsy patients were ascertained at the NIMHANS, Bangalore. All patients satisfied the clinical criteria for hot water epilepsy. The control group comprised 288 healthy individuals from southern parts of India. None of them had any apparent history of neurological disorders.

3.1.3. Sequence analysis

As a first step towards identification of causative gene for epilepsy in HWE150, six genes were examined by Sanger sequencing for their exonic-, flanking intronic- and untranslated regions using DNA from the proband, III:10. Subsequently, to examine a large number of genes at the same time in the locus, whole exome sequencing experiment was undertaken. The whole exome sequencing was carried out for two affected individuals, I:5 and III:10 from the family, with TruSeq Exome Enrichment kit (Illumina Inc., USA).

The exome enrichment kit spans 62Mb (1.7%) of the human genome corresponding to the RefSeq, Consensus Coding Sequence Database (CCDS), ENCODE/GENCODE protein coding exons, 5'UTRs, 3'UTRs, and microRNA targets from miRBase including their predicted targets. It uses 95mer probes constructed against the human GRCh37/hg19 reference genome. About 5µg of genomic DNA from each of I:5 and III:10 was fragmented to generate dsDNA of sizes 180-280bp by sonication (Covaris, ThermoFisher Scientific, USA). The DNA was enzymatically polymerized and phosphorylated at the 5'-end, purified using Agencourt AMPure XP beads (Beckman Coulter, USA) and adenylated at the 3'-end. It was indexed and ligated with adapter sequences of 60 bases using a standard protocol and

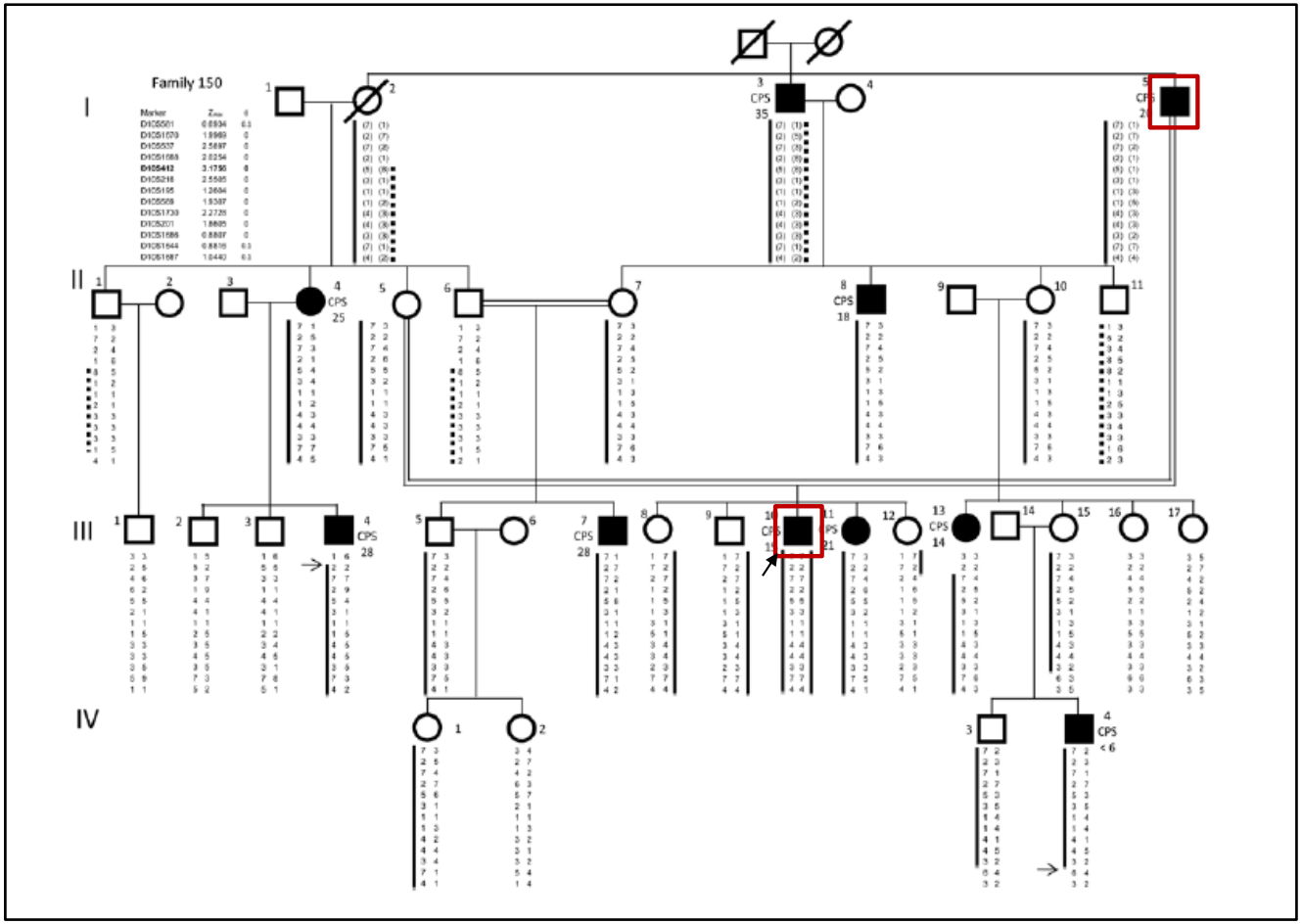


Figure 3.1: Pedigree chart of Family 150. The filled symbols represent clinically affected individuals and empty symbols represent unaffected individuals. The proband III:10 is represented with an arrow. The Roman numbers to the left of the pedigree denote generations and Arabic numbers beside the symbols denote individuals. A 13-marker haplotype is shown below the symbols. The HWE-associated haplotype is represented as solid lines. Alleles in parentheses are inferred. Key recombinant events in individuals, III:4 and IV:4, are indicated by arrows. The highest LOD score (Zmax), obtained for each marker in the haplotype is shown on the top left. Seizure types and age at onset are indicated beside symbols. CPS = complex partial seizures. Individuals *boxed in red* were exome sequenced. (Modified from Ratnapriya et al 2009b).

cleaned using AMPure XP beads. Adapter-ligated products of the size range 200-300 bases were purified by gel excision with MinElute Gel Extraction Kit (Qiagen, Germany). The prepared library was enriched for adapter-ligated fragments through a PCR step, purified with AMPure XP beads and checked for size distribution using an Agilent High Sensitivity Bioanalyzer Chip (Agilent Technologies, USA). The library was denatured, hybridized to biotin-labelled probes, bound to streptavidin beads and magnetically enriched by pulling down biotinylated DNA fragments bound to streptavidin. The captured library was eluted from the beads in DNase-free water at 95°C, and hybridized for a second PCR enrichment reaction. Cluster generation was performed on cBOT (Illumina Inc.) and sequencing on Illumina Genome Analyzer (GAIIx) following the manufacturer's guidelines.

3.1.4. Whole exome sequencing analysis

The whole exome sequencing reads were aligned to reference human genome (hg19/GRCh37) using BWA 0.6.0 (Li and Durbin 2009). Alignment was performed with high quality reads having at least 70% bases. Duplicate reads resulting from PCR artifacts were removed using SAMtools v0.1.7a (Li et al 2009) and an *in-house* script. Variants were called by SAMtools at a Phred like SNP quality score greater than 20. The variants identified were annotated against dbSNP131 and dbSNP135 databases using *in-house* scripts. Unreported variants were verified against dbSNP138, Ensembl (http://www.ensembl.org/index.html), Exome Variant Server (http://evs.gs.washington.edu/EVS/), 1000 Genomes database (http://www.1000 genomes.org/) and The Exome Aggregation Consortium (http://exac.broadinstitute.org/) datasets. Coverage statistics for aligned reads and base statistics were obtained through custom scripts. The novel and rare variants obtained from the analysis were ensured for at least 5x of read depth coverage and preferably called from both the DNA strands. Variants up to 2x read depth were manually analyzed and compared across the two whole exome sequencing (WES) samples. Protein coding exons with intronic flanks up to 100bp, 5'UTRs, and 3'UTRs for genes situated at the 10q21.3-q22.3 locus were manually examined to minimize the possibility of missing out new variants. All suspected variants at low read depth, and those not called in either of the samples, were taken into consideration for validating using Sanger based sequencing. All regions in the target exome, not covered by the TruSeq probes, were Sanger sequenced to extensively assess any region that might contain a potential new variant (Appendix A3.1 and A3.2).

3.1.5. Genetic analysis and Sanger based sequence validation

Novel variants and those with low frequency of occurrence (MAF ≤ 0.005, as reported in databases) obtained from the whole exome sequencing datasets for the locus were considered for further analysis. Novel variants and rare variants whose zygosity could not be conclusively determined from the whole exome sequence alignment were examined and validated by Sanger-based sequencing as previously described in the section 2.1.5. The variants were first examined for co-occurrence in the two exome sequenced individuals. The affected parent, I:5 carried a single copy of the segregating haplotype, and the offspring III:10, inherited the haplotype from both its parents (Figure 3.1). III:10 was therefore expected to be homozygous for the variant segregating with the haplotype in the family. This criterion was used to rule out rare/novel variants that were heterozygous in III:10, if any, from further examining its segregation across the family. Post-confirmation, a complete family-based segregation analysis was carried out for the variants that were present in both the exome sequenced individuals: heterozygous in I:5 and homozygous in III:10, to ascertain their co-segregation with the phenotype across generations. Co-segregating variants were examined in an ethnically-matched control set of 288 healthy individuals.

The gene sequences were obtained from Human Genome Mapviewer (build 37.3, NCBI, http://www.ncbi.nlm.nih.gov/mapview/) database. The primer pairs spanning the variant-carrying exons/regions were designed using Primer3 Input (http://bioinfo.ut.ee/primer3-0.4.0/; Appendix A3.3). Each Sanger sequenced amplicon was compared with the reference sequence available in the GenBank database (http://www.ncbi.nlm.nih.gov/genbank/) using SeqMan II 5.01 (DNASTAR Inc., USA).

3.1.6. Screening for additional alleles in *FUT11* in HWE patients

FUT11 was screened among additional hot water epilepsy patients to try to identify more variants of potential functional significance. The gene was sequenced in a group of 288 unrelated probands. The entire transcript structure was examined for the coding exons with their flanking intronic regions covering the splice junctions and 5'- and 3'- untranslated regions (UTRs), with specific primers designed using Primer3 Input (Appendix A3.4, Koressaar and Remm 2007). In addition, *FUT11* was examined in 288 control individuals.

3.1.7. Bioinformatics analysis

The nucleotide sequence of *FUT11* corresponds to 4148 bases at 10q22.2 according to the GRCh37.p13 assembly. The protein sequences for FUT11 across different species were

obtained from NCBI HomoloGene database (http://www.ncbi.nlm.nih.gov/homologene/). The protein structure and location of rare HWE variants was predicted using Protter v1.0. Multiple species sequence alignments for amino acid conservation analysis was carried out by Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/). Probable functional effects of missense variants were predicted using PolyPhen-2 (Adzhubei et al 2010), SIFT (Kumar et al 2009), Mutation Taster (Schwarz et al 2010), Mutation Assessor (Reva et al 2011) and FATHMM (Shihab et al 2013). The human FUT11 protein sequence NP_775811 from NCBI was used as input for these tools. Splice site prediction for the intronic variants obtained in genes *NDST2* and *DLG5* were performed using Mutation Taster, NNSplice (Jian et al 2014), NetGene2 (Hebsgaard et al 1996) and Human Splicing Finder (Desmet et al 2009). The gene sequences obtained from NCBI Human Genome Mapviewer were used as inputs for these tools.

3.1.8. Expression of *FUT11* transcripts in human brain regions

The presence of *FUT11* transcript in different regions of the human brain was examined using reverse-transcriptase PCR (RT-PCR) in Marathon-Ready full-length brain cDNAs from cerebral cortex, cerebellum, hippocampus, hypothalamus and whole brain (Clontech, USA). A primer pair with the forward primer 5'- CAACTTCTTGCTGAGCCACG-3' located within the first exon and reverse primer 5'-GATTGTTCGGCATCCAGTCC-3' within the second exon respectively of *FUT11* was used to amplify the transcript region. The amplified product of 500 bases was verified on a 1% agarose/TAE/EtBr gel electrophoresis and confirmed by bi-directional DNA sequencing with the respective primer pairs on an ABI3730 Genetic Analyzer.

3.1.9. Cloning and site-directed mutagenesis

The wild-type cDNA transcript of *FUT11* (NM_173540.2, NP_775811.2) cloned into mammalian pCMV6- entry vector with C-terminal GFP tag (RG206082) was obtained from Origene, USA. The identity of the clone was verified by Sanger sequencing with *FUT11* cDNA specific overlapping primers (Appendix A3.5). The cDNA was cloned into a pcDNA3.1(+) vector with primer pairs 5'-TGGTACCGAGCCACCATGGCGGCC GGCCCCATTAGGGTG-3' and 5'-GAGCGGCCGCTTAGAGATGTTGCCTCTTCATGA AGATTTC-3'. Two variants were introduced in the wild-type *FUT11* cDNA with specific primers (Appendix A3.6) using QuikChange II XL Site-Directed Mutagenesis reagents (Agilent Technologies, USA). All the plasmids were purified using QIAprep Spin Miniprep

(Qiagen) columns and the specific variants incorporated were confirmed by Sanger sequencing.

3.1.10. Cell culture and transfections

To study endogenous protein expressions, human HEK293 (ATCC, USA), human SH-SY5Y neuroblastoma (ATCC) and rat C6 glioma (ECACC, Sigma-Aldrich, USA) cells were grown to about 80% confluence in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% heat-inactivated fetal bovine serum (FBS), 2mM L-glutamine, 100U/ml penicillin and 0.1mg/ml streptomycin (Sigma-Aldrich), in a humidified atmosphere of 5% CO_2 at 37°C. Transfections for protein over-expression studies were carried out in HEK293 cells grown up to 40% confluence under same conditions. The cells were transfected with plasmids containing FUT11-wildtype and FUT11-mutant cDNA with the incorporated coding variants, and vector control using calcium phosphate transfection. Forty eight hours post-transfection, the cells were processed accordingly, for immunoblotting or microscopy.

3.1.11. Immunoblotting

Cell lysates were prepared from HEK293 cells over-expressing FUT11- wildtype and -mutant proteins using 200µl of chilled RIPA lysis buffer (10mM Tris-HCl pH 7.5, 150mM NaCl, 1% TritonX-100, 1% sodium deoxycholate, 0.1% SDS, and 5mM EDTA) supplemented with protease inhibitor cocktail (Sigma-Aldrich). The lysates were quantified by bicinchoninic acid (BCA) assay (Sigma-Aldrich). 50µg of protein was mixed with 6x SDS gel loading dye, boiled for 5 minutes and loaded in each well of an 8% SDS-polyacrylamide gel (Laemmli 1970). The proteins on the gel were subsequently electrotransferred on to a 0.45µm nitrocellulose membrane (GE Healthcare Life Sciences, USA) at constant voltage of 22V for one hour and fifteen minutes using a semi-dry transfer apparatus (Bio-Rad Laboratories, USA). The blot was blocked for 1 hour at room temperature in blocking solution [5% BSA (Sigma-Aldrich) in 1x PBS]. It was then incubated overnight at 4°C with rabbit anti-FUT11 polyclonal antibody (ab121411, Abcam, UK) at 1:1000 dilution in 1% BSA in 1x PBS. The blot was gently washed three times with wash solution containing 1x PBS and 0.1% Tween-20 (Sigma-Aldrich) at room temperature for 5 minutes each. The blot was further incubated for 4 hours at 4°C with a 1:10000 dilution of horseradish peroxidase (HRP)-conjugated, goat anti-rabbit IgG (Bangalore GeNei, India) in 1x PBS containing 1% BSA. This was then washed four times for 10 minutes each with the wash solution, at room temperature. The protein bands were detected

using an enhanced chemiluminiscent substrate for HRP (West pico, Pierce, USA). γ-tubulin (T5326, Sigma-Aldrich) or vinculin (ab129002, Abcam) was used as a loading control, probed on the same blot.

For the human brain regions, 50µg of protein lysates from different regions of the brain were loaded into gel, electrophoresed and immunoblotted using the same protocol as described above.

3.1.12. Immunocytochemistry

Human HEK293, human SH-SY5Y neuroblastoma and rat C6 glioma cells were grown on poly-L-lysine coated coverslips in 10% DMEM up to 80% confluence. Immunostaining was performed on cells for endogenous expression and those over-expressing wild-type FUT11-pcDNA3.1 and mutants Val391Ile and His483Arg- pcDNA3.1 after 48 hours of transfection. The cells were fixed with 2% PFA (paraformaldehyde) for 15 minutes, washed twice with 1x PBS for 5 minutes each, permeabilized with 0.1% Triton-X in PBS for 10 minutes at 25°C and blocked with 5% BSA for four hours at 4°C. Cells were incubated with rabbit polyclonal anti-FUT11 antibody diluted to 1:100 for endogenous expression and 1:1000 for overexpression in 1% BSA in PBS, overnight at 4°C. They were gently washed thrice at 25°C with PBS for 5 minutes each and incubated with anti-rabbit Alexa Fluor 568-conjugated secondary antibody (Molecular probes, Invitrogen, USA) diluted to 1:500 in 1% BSA in PBS for four hours at 4°C. For dual antibody staining procedure, primary antibodies, mouse polyclonal anti-58K Golgi protein antibody (1:100, ab27043, Abcam) and rabbit polyclonal anti-FUT11 (1:100, Abcam) were combined at 1:1 in 1% BSA and added onto the cells on coverslips and incubated at 4°C overnight. Secondary antibodies, anti-mouse Alexa Fluor 568 (Molecular probes, Invitrogen) and anti-rabbit Alexa Fluor 488 (Molecular probes, Invitrogen) were mixed in the same ratio at 1:500 dilution for the co-localization experiment. For dual antibody staining with endoplasmic reticulum marker calnexin, cells were stained subsequently; first with primary and secondary antibodies against FUT11 (1:100), followed by blocking in 5% BSA, and then staining with rabbit polyclonal anti-calnexin antibody (1:100, ADI-SPA-860-D, Enzo Life Sciences, USA).

3.2. Results

3.2.1. Genome-wide linkage analysis

A previous study on Family 150 employing the whole genome-wide linkage analysis had provided a highest parametric two-point LOD score of 3.17 at θ=0 at 60% penetrance value

and 1% phenocopy for D10S412, suggesting evidence for linkage at 10q21.3-q22.3. The critical genomic interval at chromosome 10 was defined by D10S581 (centromere-proximal) and D10S201 (centromere-distal) spanning a 15Mb sequence length at the 10q21.3-q22.3 region (Ratnapriya et al 2009b).

3.2.2. Whole exome sequencing analysis

The whole exome sequencing experiment provided a sequence coverage of 89.94% in I:5 and 89.99% in III:10 with high quality reads (Table 3.1). Whole exome sequencing analysis of the 15Mb target region at 10q21.3-q22.3 yielded a total of 534 variants in the two affected individuals I:5 and III:10, examined (Figure 3.2). The 534 variants included both common and rare/novel single nucleotide polymorphisms (SNPs) and insertion/deletions (InDels). Upon variant analysis, a total of 488 variants were found to be reported at high allele frequencies in dbSNP138, Ensembl, Exome Variant Server, 1000 Genomes and The Exome Aggregation Consortium (ExAC) databases. These variants were not considered for further examination. Variants that were heterozygous or homozygous and present in any one of the individuals or both, that could be confidently determined at >10X read depth, were not considered further (Table 3.2). Forty six variants were found to be novel or rare with a minor allele frequency (MAF) ≤ 0.005 (Table 3.3). Variants whose zygosity could not be conclusively determined, and those that were heterozygous in I:5 and homozygous in III:10 were examined by bi-directional Sanger sequencing. The disease haplotype segregates in a heterozygous and homozygous manner respectively in the two exome sequenced individuals, I:5 and III:10. Five among the 46 variants were present in I:5 as heterozygous and in III:10 as homozygous, according to the segregating haplotype.

These five variants were examined further in the family for segregation with the disease phenotype. Four of these variants, *USP54* (Ubiquitin specific peptidase 54), c.4999G>A (p.Asp1606Asn); *FUT11* (fucosyltransferase 11), c.1171G>A (p.Val391Ile); *NDST2* (N-deacetylase and N-sulfotransferase 2), c.1248+57C>T and *DLG5* (Discs Large Homologue 5), c.5165-58C>T segregated with the disorder in this family (Figure 3.3). Analysis of the variants in ethnically matched controls found c.4999C>T in *USP54* to be a common variant and was, therefore, excluded from further analysis. The sequence variants in *FUT11*, c.1171G>A; *NDST2*, c. 1248+57C>T and *DLG5*, c.5165-58C>T occurred at frequencies of 0.008, 0 and 0.003 respectively, among 288 healthy controls.

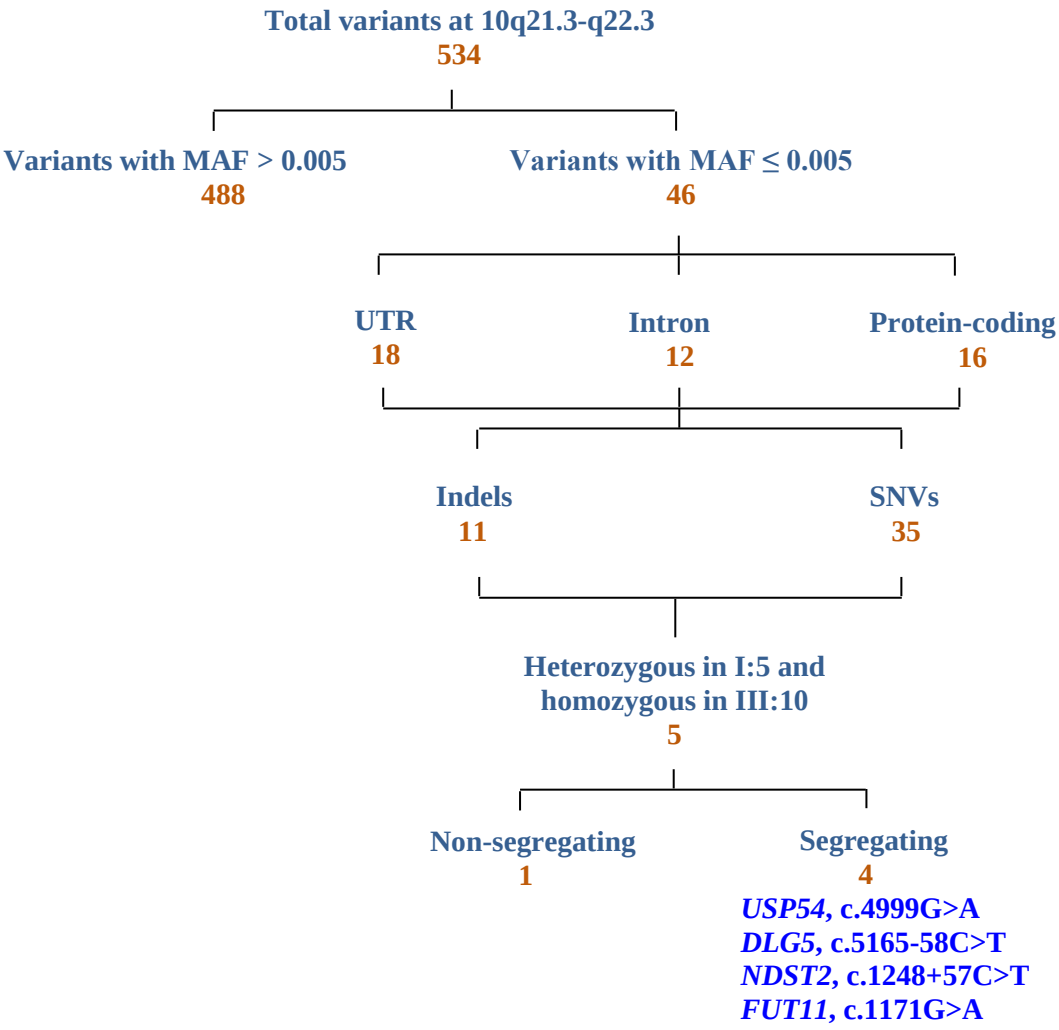


Figure 3.2: Targeted analysis of the 10q21.3-q22.3 region in whole exome sequencing experiment. The schematic represents the work flow for screening and filtering the variants in the 10q21.3-q22.3 target region to identify potentially causative variant/s. Each category with the number of variants identified is shown.

Table 3.1: Summary statistics for exome sequencing of processed reads

Alignment features	Individual I:5		Individual III:10	
Read length in bases (paired)	100		100	
Total number of reads[a]	71027384		67532114	
Total reads for alignment (percentage high quality reads)[b]	63948526	(90.0)	60883132	(90.15)
Total reads aligned (percentage reads aligned)[c]	63754642	(99.70)	60740112	(99.77)
Reads aligned to whole exome (percentage reads aligned)	31149010	(48.71)	29682490	(48.75)
10q21.3-q22.3 target length	319541		319541	
10q21.3-q22.3 target covered (percentage target covered)	287393	(89.94)	287545	(89.99)

[a] Number of raw reads, [b] Number of reads yielded post processing of raw reads. Reads with >70% bases with Phred score >20, [c] All reads aligned including exome and to other regions of the genome.

Table 3.2: Sequence coverage summary of the 10q21.3-q22.3 target region

Coverage attributes	Individual I:5		Individual III:10	
	10q21.3-q22.3	Whole exome	10q21.3-q22.3	Whole exome
%Total target covered with at least 5X read depth	85.21	81.24	84.46	80.69
%Total target covered with at least 10X read depth	81.38	77.11	80.23	76.33
%Total target covered with at least 15X read depth	77.57	73.42	76.45	72.40
%Total target covered with at least 20X read depth	73.61	69.72	72.29	68.40

Table 3.3: Rare genetic variants identified by exome sequencing analysis at the 10q21.3-q22.3 locus

Genomic position	Gene	Sequence variant	Hom and present in either or both (Y/-)	Het and present in either or both (Y/-)	Het in I:5 / Hom in III:10 (Y/-)	Effect on protein	Variant ID	Frequency in control databases				Segre-gation (Y/N)
								ExAC	1000G	ESP	dbSNP138	
69366667[a]	*CTNNA3*	NM_013266.3:c.240T>C	-	Y	-	p.Ala80=	rs564981915	0.0004	0.0002	-	0.0002	-
69991194[a]	*ATOH7*	NM_145178.3:c.245T>C	Y	-	-	p.Leu82Gln	-	-	-	-	-	-
70097094[c]	*HNRNPH3*	NM_012207.2:c.112+4C>T	-	-	Y	-	rs780458079	0.000008	-	-	-	N
70502424[c]	*CCAR1*	NM_018237.3:c.518+102T>C	Y	-	-	-	-	-	-	-	-	-
70661096[a]	*DDX50*	NM_024045.1:c.-55G>A	-	Y	-	-	-	-	-	-	-	-
70962242[a]	*SUPV3L1*	NM_003171.3:c.1566C>T	Y	-	-	p.Tyr391=	rs551672481	0.0003	-	-	-	-
71657272[a]	*COL13A1*	NM_080798.3:c.544-1181A>G	Y	-	-	-	rs1227746	0.0013	0.0046	0.0039	0.0046	-
71658512[a]	*COL13A1*	NM_080798.3:c.603G>A	Y	-	-	p.Thr201=	rs1227747	0.0013	0.0046	0.0040	0.0046	-
71898273[c]	*TYSND1*	NM_173555.3:c.*1407T>A	Y	-	-	-	rs550607849	-	0.0006	-	0.0006	-
72294208[a]	*PALD1*	NM_014431.2:c.1050G>A	-	Y	-	p.Leu350=	rs750397642	0.00007	-	-	-	-
72493659[a]	*ADAMTS14*	NM_080722.4:c.1227C>T	-	Y	-	p.Asp409=	rs144216765	0.0007	0.0006	0.0003	0.0006	-
72639404[a]	*SGPL1*	NM_003901.3:c.*2312A>C	-	Y	-	-	rs199642597	-	-	-	-	-
72643671[a]	*PCBD1*	NM_000281.3:c.*36_*41del	Y	-	-	-	-	-	-	-	-	-
73039616[c]	*UNC5B*	NM_170744.4:c.118C>A	-	Y	-	p.Pro40Thr	-	-	-	-	-	-
73407232[b]	*CDH23*	NM_001171935.1:c.1449+858delA	-	Y	-	-	rs56367143	-	-	-	-	-
74094665[b]	*DNAJB12*	NM_001002672.2:c.*903A>G	Y	-	-	-	rs1879471	-	0.0002	-	0.0002	-
74695287[a]	*PLA2G12B*	NM_001318124.1:c.*88G>A	Y	-	-	-	rs2461917	-	0.0002	-	0.0002	-
74767338[a]	*P4HA1*	NM_000917.3:c.*641_642insA	-	Y	-	-	-	-	-	-	-	-
74837576[a]	*P4HA1*	NM_000917.3:c.-32-2903delT	-	Y	-	-	rs11334323	-	-	-	-	-

Genomic position	Gene	Sequence variant	Hom and present in either or both (Y/-)	Het and present in either or both (Y/-)	Het in I:5 / Hom in III:10 (Y/-)	Effect on protein	Variant ID	Frequency in control databases				Segre-gation (Y/N)
								ExAC	1000G	ESP	dbSNP138	
75258443[c]	*USP54*	NM_152586.3:c.4999G>A	-	-	Y	p.Asp1606Asn	rs117429793	0.0063	0.0076	0.0019	0.0009	Y
75434363[a]	*AGAP5*	NM_001144000:c.2037C>T	-	Y	-	p.Cys685=	rs3998255	-	-	-	-	-
75434641[a]	*AGAP5*	NM_001144000:c.1777C>T	-	Y	-	p.Arg593Trp	rs3957125	-	-	-	-	-
75434973[a]	*AGAP5*	NM_001144000:c.1445A>G	Y	-	-	p.Tyr482Cys	rs3998268	-	-	-	-	-
75439958[a]	*AGAP5*	NM_001144000:c.498-94G>A	-	Y	-	-	rs71507079	-	-	-	-	-
75533410[c]	*FUT11*	NM_173540.2:c.1171G>A	-	-	Y	p.Val391Ile	rs199728665	0.0022	0.0026	0.0002	0.0026	Y
75559663[a]	*ZSWIM8*	NM_015037.3:c.4666-106_4666-105insTCTA	-	Y	-	-	rs66528433	-	-	-	-	-
75566358[c]	*NDST2*	NM_003635.3:c.1248+57C>T	-	-	Y	-	-	-	-	-	-	Y
75574112[a]	*CAMK2G*	NM_001222.3:c.*178C>A	Y	-	-	-	-	-	-	-	-	-
75878916[a]	*VCL*	NM_003373.3:c.*991_*992insTA	-	Y	-	-	rs397804906	-	-	-	-	-
75883135[a]	*AP3M1*	NM_012095.5:c.*432_*433insTT	-	Y	-	-	rs34725483	-	-	-	-	-
76074331[a]	*ADK*	NM_001123.3:c.144-94T>C	-	Y	-	-	rs187598971	-	0.0014	-	0.0014	-
76736159[a]	*KAT6B*	NM_001256468.1:c.1444+620A>G	-	Y	-	-	rs191600556	-	0.0018	-	0.0018	-
77157909[a]	*ZNF503*	NM_032772.5:c.*597_*598insA	-	Y	-	-	rs534047467	-	0.0055	-	0.0054	-
78868166[a]	*KCNMA1*	NM_0022247.3:c.1223+71_1223+73delAAA	-	Y	-	-	rs138226779	-	-	-	-	-
79556410[c]	*DLG5*	NM_004747.3:c.5165-58C>T	-	-	Y	-	rs575452137	-	0.0012	-	0.0012	Y
81316528[c]	*SFTPA2*	NM_001320814.1:c.*437A>C	-	Y	-	-	rs1914674	-	-	-	-	-
81316603[c]	*SFTPA2*	NM_001320814.1:c.*362G>A	Y	-	-	-	rs1059082	-	-	-	-	-
81904003[a]	*PLAC9*	NM_001012973.1:c.187C>T	-	Y	-	p.Leu63=	-	-	-	-	-	-
81925797[c]	*ANXA11*	NM_0.001157.2:c.858+42_858+43insGTGT	-	Y	-	-	rs753895566	0.00002	-	-	-	-

Genomic position	Gene	Sequence variant	Hom and present in either or both (Y/-)	Het and present in either or both (Y/-)	Het in I:5 / Hom in III:10 (Y/-)	Effect on protein	Variant ID	Frequency in control databases				Segre-gation (Y/N)
								ExAC	1000G	ESP	dbSNP138	
82277753[a]	*TSPAN14*	NM_030927.2:c.*21G>A	-	Y	-	-	rs370735521	0.00005	-	0.0002	-	-
86001085[a]	*LRIT1*	NM_015613.2:c.109C>A	-	Y	-	-	-	-	-	-	-	-
86132205[a]	*CCSER2*	NM_018999.3:c.1397G>A	-	Y	-	p.Trp466*	-	-	-	-	-	-
86185654[a]	*CCSER2*	NM_018999.3:c.1868+5T>G	Y	-	-	-	rs1336202	0.0003	-	-	-	-
88196174[a]	*WAPL*	NM_015045.3:c.*1125_*1126insGTT	-	Y	-	-	rs34285032	-	-	-	-	-
88425754[a]	*OPN4*	NM_0332282.3:c.*246A>G	Y	-	-	-	rs2736687	-	0.0014	-	-	-
88703471[a]	*MMRN2*	NM_02756.2:c.1070C>T	-	Y	-	p.Thr357Met	rs147017508	0.00005	-	0.0002	-	-

[a] Variants examined and eliminated by sequence analysis at >10X read depth in SAMtools. [b] Rare variants identified from UTR regions covered by Sanger sequencing. [c] Variants confirmed and examined by Sanger sequencing. Y = Yes, N = No.

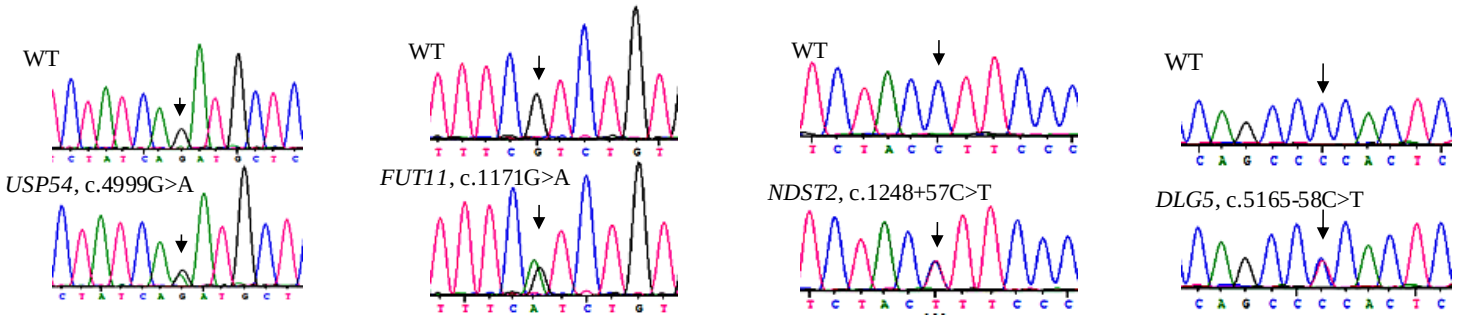


Figure 3.3: Electropherogram of segregating genetic variants in HWE150. Representaion of the four segregating genetic variants *USP54*, c.4999G>A; *FUT11*, c.1171G>A; *NDST2*, c.1248+57C>T and *DLG5*, c.5165-58C>T in an unaffected and an affected individual.

3.2.3. Bioinformatic analysis of the segregating variants

To analyze the segregating variants for their pathogenic potential using in-silico approach, the three variants were examined employing bioinformatic prediction tools. The two intronic variants in *DLG5*, c.5165-58C>T and *NDST2*, c. 1248+57C>T were analyzed using Mutation Taster, NNSplice, NetGene2 and Human Splicing Finder. Mutation Taster predicted both the variants to be polymorphic in nature, with no pathogenic effect imparted by the variant allele. NNSplice predicted no pathogenicity for the *DLG5* variant, and did not yield any splice site predictions for *NDST2*. NetGene2 yielded no prediction outcome for the wild-type and variants alleles of the two genes, suggesting these sites might not be involved in splicing. Human Splicing Finder yielded no probable impact on splicing by the *DLG5* variant, and the *NDST2* variant might alter an intronic ESS (Exonic Splicing Silencer) site and probably does not affect splicing. These two intronic variants, although rare, resided far away from regions of predicted splice sites and were not scored for pathogenic effect, or any impact on splice site regulation by the prediction tools (Table 3.4). Conversely, the variant *FUT11*, c.1171G>A although comparatively common in the ethnically-matched control population, was found to be pathogenic by bioinformatic predictions (Table 3.5). Subsequently, *FUT11* was examined in additional HWE patients. Segregation of *FUT11*, c.1171G>A in HWE150 is represented (Figure 3.4).

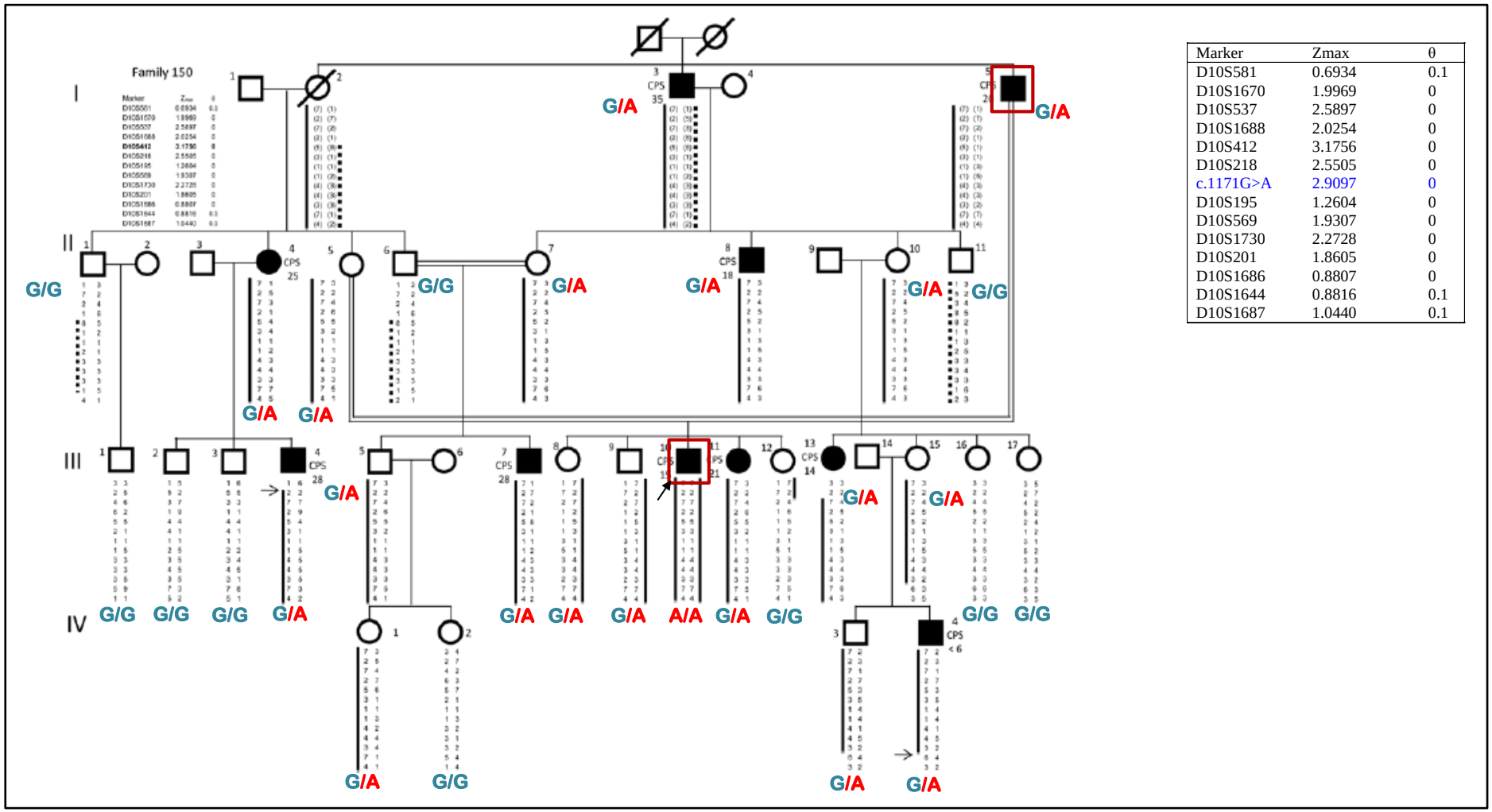
Table 3.4: Splice site predictions for intronic variants in *DLG5* and *NDST2*

Variants	Mutation Taster prediction (scores)[a]	NNSplice scores	Human Splicing Finder
DLG5-WT	Polymorphism (0.3488)	0.59	
DLG5,c.5165-58C>T	Polymorphism (0.3519)	0.53	No significant splicing motif alteration detected. Probably has no impact on splicing.
NDST2-WT	Polymorphism (0.77)	No predictions made	-
NDST2,c.1248+57C>T	Polymorphism (0.90)	No predictions made	Alteraion of a predicted intronic ESS[b] site. Probably no impact on splicing.

[a] Mutation Taster provides score for the wild-type predicted splice site based on the efficacy of prediction. The impact of variant allele is predicted based on the threshold score of the wild-type allele. [b] Exonic Splicing Silencer.

Table 3.5: Predictions for *FUT11*, c.1171G>A

Variant	SIFT	PolyPhen-2	PROVEAN	Mutation Taster	FATHMM
FUT11, c.1171G>A	Damaging	Possibly damaging	Neutral	Disease-causing	Tolerated



Marker	Zmax	θ
D10S581	0.6934	0.1
D10S1670	1.9969	0
D10S537	2.5897	0
D10S1688	2.0254	0
D10S412	3.1756	0
D10S218	2.5505	0
c.1171G>A	2.9097	0
D10S195	1.2604	0
D10S569	1.9307	0
D10S1730	2.2728	0
D10S201	1.8605	0
D10S1686	0.8807	0
D10S1644	0.8816	0.1
D10S1687	1.0440	0.1

Figure 3.4: Segregation of *FUT11*, c. 1171G>A in HWE150. Family pedigree of HWE150 representing the segregation of c.1171G>A with the disease haplotype. The maximum LOD score obtained for the markers and the variant is represented in the box.

3.2.4. Sequence analysis of *FUT11* in south Indian HWE patients reveals two heterozygous rare variants in the coding exons

To further examine the potential role of *FUT11* in HWE, complete transcript structure of the gene was sequenced by Sanger-based method in a set of 288 HWE cases from southern India. This was to find additional rare variants in the gene which may be functionally relevant and provide further evidence for *FUT11* in HWE.

Twelve single nucleotide variants (SNVs) were identified in the HWE cases (Table 3.6). Eight among the 12 variants were non-coding and the other four, coding. Among the 12, two were unreported SNVs and the remaining ten were represented in control databases (1000 genomes, ExAC databases). All the variants were examined for their frequencies in 288 ethnically matched controls. Following this examination, two genetic variants, c.1171G>A (p.Val391Ile) and c.1448A>G (p.His483R) were identified in *FUT11* (Figure 3.5, Tables 3.7 and 3.8). c.1171G>A was reported as a rare variant, present at a frequency of 0.008 among south Indian controls, and c.1448A>G was absent among the controls examined. Variant c.1171G>A (p.Val391Ile) identified in Family 150 was found among four additional, unrelated HWE families.

Table 3.6: *FUT11* variants among HWE patients

Exon/ Intron	Nomenclature	Amino acid change	Ref SNP ID	Allele frequency in Indian controls (n=576)	Frequency in control databases
5'UTR	NM_173540.2:c.-18C>A	-	rs750557083	0	0.0004
Exon 1	NM_173540.2:c.13C>T	p.Pro5Ser	rs565233785	0	0.0002
Intron	NM_173540.2:c.708+74C>A	-	≠	0.026	-
Intron	NM_173540.2:c.708+90C>A	-	≠	0.026	-
Intron	NM_173540.2:c.709-62C>A	-	rs5017445	0.13	-
Intron	NM_173540.2:c.709-46C>A	-	rs1886668800	0.038	0.002
Exon 2	NM_173540.2:c.1171G>A	p.Val391Ile	rs199728665	0.004	0.0026
Exon 3	NM_173540.2:c.1392C>T	p.Leu464=	rs143425365	0	0.0008
Exon 3	NM_17350.2:c.1448A>G	p.His483Arg	rs767940642	0	-
3'UTR	NM_17350.2:c.*527A>G	-	rs76033784	0.033	0.1438
3'UTR	NM_17350.2:c.*364C>G	-	rs544772818	0.002	0.0002
3'UTR	NM_17350.2:c.*135A>G	-	rs558392531	0.002	0.001

≠ Variants not reported in any of the control databases.

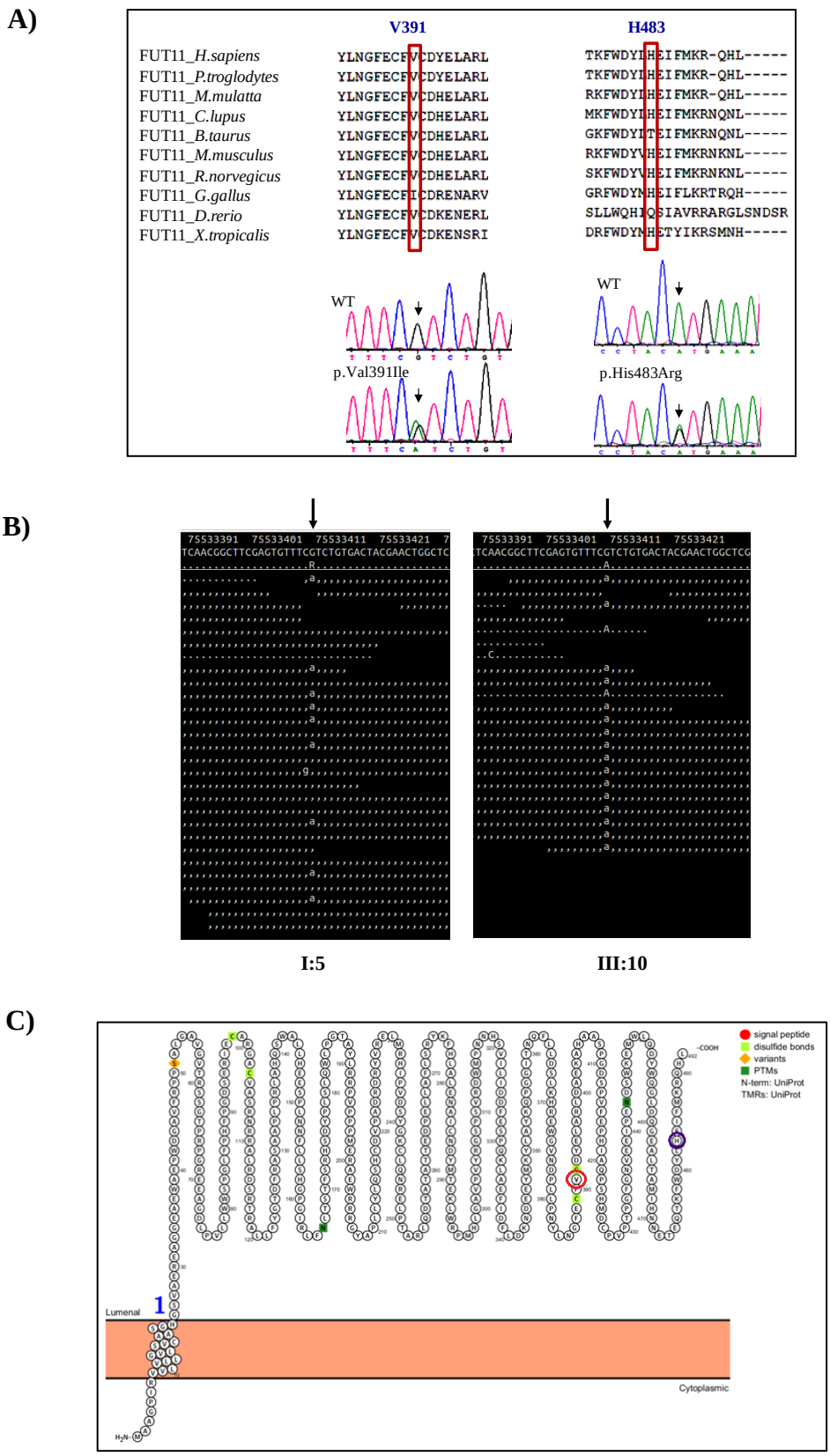


Figure 3.5: Sequence of rare variants in FUT11. (A) Representation of the amino acid conservation of the corresponding genetic variants and the electropherogram representation in an unaffected and affected individual. (B) SAMtools alignment snapshots of c.1171G>A variant in individuals I:5 and III:10. (C) FUT11 predicted structure and location of rare HWE variants marked with circles (Protter v1.0).

Table 3.7: Rare non-synonymous *FUT11* variants in 288 HWE probands

Nucleotide change	Amino acid change	Family	Type	dbSNP ID	MAF in control chromosomes (n=576)	Minor allele frequency (MAF) in control databses			
						dbSNP135	1000G	ESP	ExAC
c.1171G>A	p.Val391Ile	HWE 150 HWE 014 HWE 072 HWE 272 HWE 289	Familial Familial Familial Familial Familial	rs199728665	0.0087	0.0026	0.003	0.0003	0.0022
c.1448A>G	p.His483Arg	HWE 230	Familial	rs767940642	0	0	-	-	0.000008249

Table 3.8: Bioinformatic analyses of the two rare variants obtained in *FUT11*

Mutation	Conservation Score	SIFT Score	PolyPhen-2 Hum Var Score	PROVEAN Score	Mutation Taster Probability Value	Mutation Assessor FI Score	FATHMM Score
p.Val391Ile	-0.899	Damaging 0.042	Probably damging 0.971	Neutral -0.84	Disease causing 0.999	Low 1.64	Tolerated 1.79
p.His483Arg	-0.535	Damaging 0.033	Benign 0.112	Neutral -0.77	Disease causing 0.998	Medium 1.995	Tolerated 1.37

3.2.5. *FUT11* and its protein

FUT11 (fucosyltransferase 11) located at 10q22 (Figure 3.6A) comprises 4148 bases of genomic sequence length encoding the 5'UTR, three coding exons, introns and 3'UTR, according to GRCh37.p13 assembly (Figure 3.6B). The corresponding protein comprises 492 amino acids (Figure 3.6C). FUT11 is a predicted type-II transmembrane protein and a putative alpha-1-3/4 fucosyltransferase that had been identified through sequence homology-based queries (Baboval and Smith 2002). Alpha-1-3/4 fucosyltransferases (α-1-3/4 FUTs) catalyze the transfer of fucose from GDP-fucose to *N*-acetyl glucosamine (GlcNAc) of glycans in an α1-3 or α1-4 glycosidic linkage. FUT11 has significant sequence similarity only to α-1-3/4 FUTs and possess peptide motifs that are evolutionarily conserved among the known vertebrate α-1-3/4 FUTs. Studies in mice have although not been able to detect fucosyltransferase activity, FUT11 was observed to hydrolyze GDP-fucose (Baboval and Smith 2002). Substrates for this enzyme are not yet known. The three exons in *FUT11* are of 708bp, 626bp and 145bp respectively. The resultant protein consists of two major domains: a glycosyltransferase N-terminal domain (glyco_tran_10_N), 74 amino acids in length and a glycosyltransferase C-terminal domain (glyco_tran_10), 140 amino acids in length, both belonging to the glycosyltransferase 10 superfamily. Portion of the first exon encodes the glyco_tran_10_N domain and the second domain is encoded by 303bp of the first and 117bp of the second exon respectively. According to NCBI BLAST of conserved domains, the N-terminal domain is the likely binding-region for fucose-like substrate and the C-terminal domain likely binds to ADP/GDP.

3.2.6. Expression of *FUT11* transcripts in human brain regions

FUT11 transcripts in different regions of the human brain were analyzed with RT-PCR carried out using a forward and reverse primer pair located within the first and second exons of the gene, respectively. The amplified product corresponds to a size of 500 bases (Figure 3.7A). Nucleotide sequences of the gel purified products represented a sequence match with the *FUT11* transcript (NM_173540). This suggested an evidence for *FUT11* expression at the transcript level in human brain regions. The RNA expression profile across different tissues was determined from the HPA RNA-seq of normal tissues (Human Protein Atlas RNA-sequencing). The transcript is universally expressed across different tissues (Figure 3.7B).

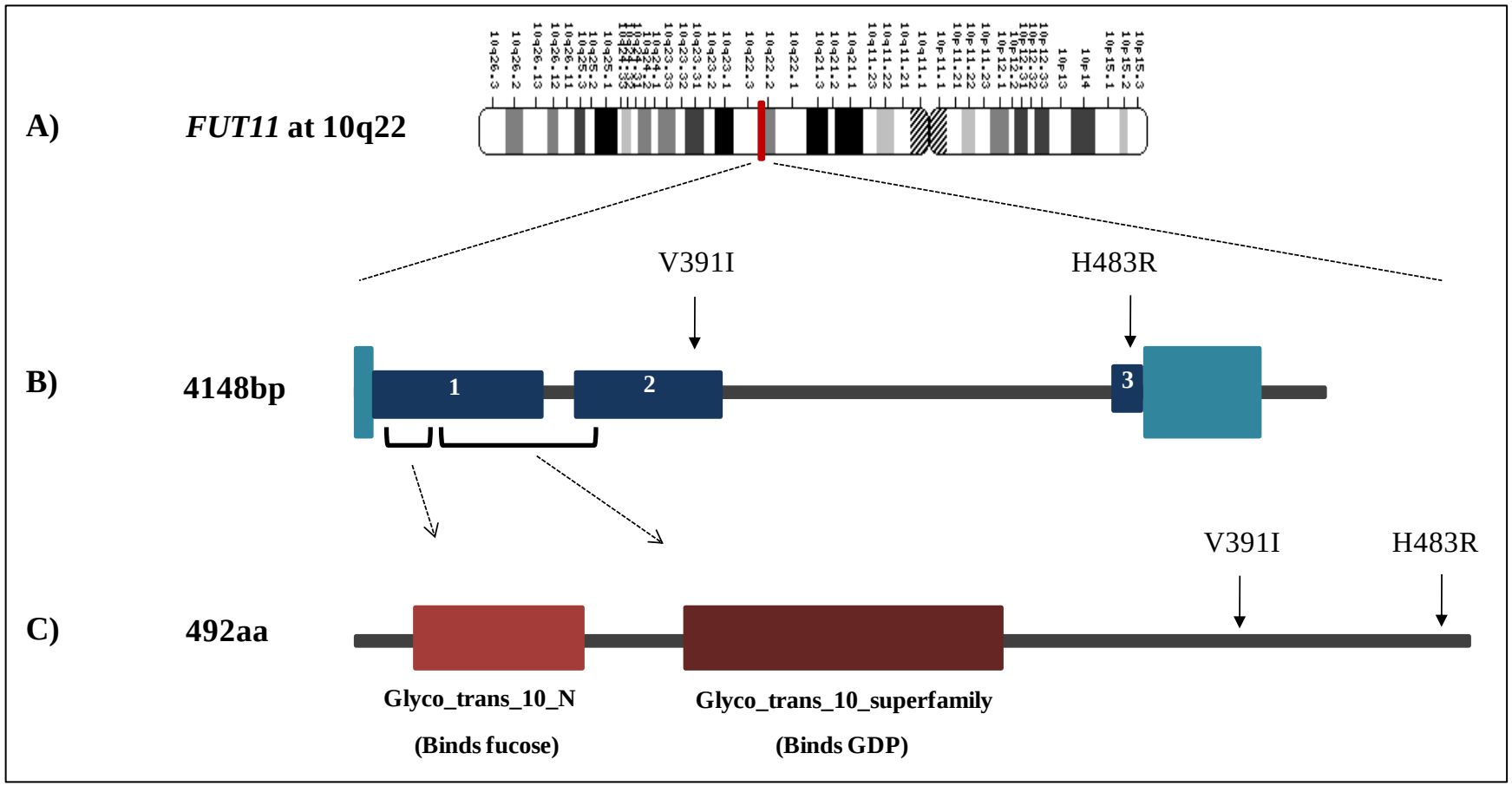


Figure 3.6: *FUT11* structure. (A) Chromosome 10 representing the location of *FUT11* at 10q22. (B) Representation of *FUT11* structure of 4148 bases comprising three coding exons (dark blue boxes) and 5'- and 3'- UTRs (light blue boxes). The arrows point the location of the two gene variants identified among HWE patients. (C) FUT11 protein of 492 amino acids showing the domain architecture and the location of the two variants in the C-terminal tail of the protein.

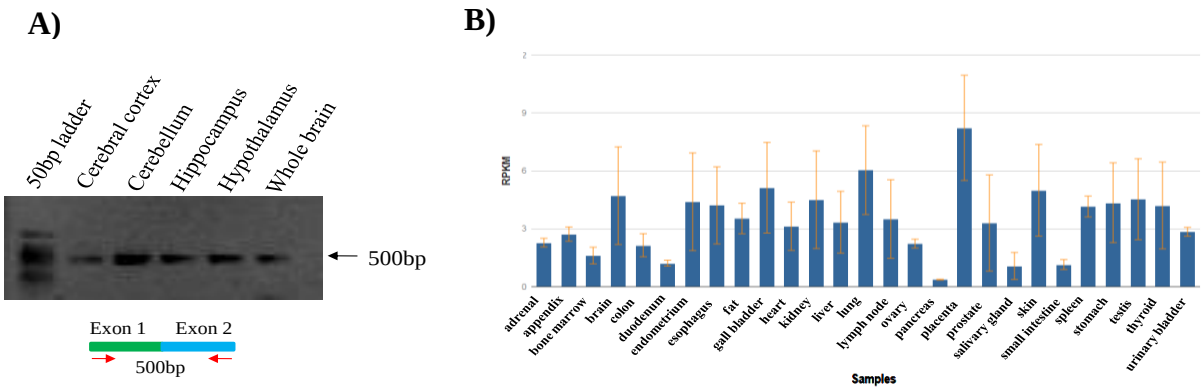


Figure 3.7: *FUT11* transcript expression. (A) Expression of *FUT11* transcript amplified from different regions of the human brain cDNA. (B) HPA RNA-seq profile of *FUT11* transcript across different tissues. RPKM: reads per kilobase million.

3.2.7. Expression analysis of FUT11 in cultured mammalian cells and human brain regions

Western blot analysis was performed on lysates from untransfected HEK293 cells and those transiently transfected with the wild-type FUT11 and mutant-carrying FUT11 plasmids. The predicted molecular weight of FUT11 is 55.8kDa (http://web.expasy.org/compute_pi/), but the protein is anticipated to undergo post-translational modifications that provide it a size of around 72kDa. In this study, the endogenous and overexpressed proteins migrate at approximately 72kDa size range. No significant difference in expression levels of the mutants with respect to the wild- type FUT11 proteins was observed (Figure 3.8A). The protein was found to express in different regions of the human brain: amygdala, basal ganglia, cerebellum, hippocampus, hypothalamus, frontal lobe, parietal lobe temporal lobe and occipital lobe at a size of 72kDa (Figure 3.8B).

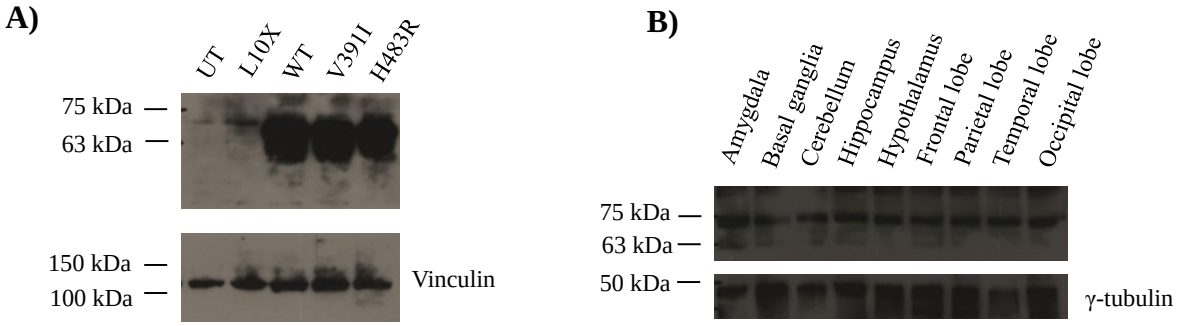


Figure 3.8: FUT11 expression in cells and brain regions. (A) Over-expression of FUT11 wild-type and mutant proteins, and untransfected FUT11 in HEK293 cells. (B) Expression of FUT11 in different regions of the human brain. The ladder is represented to the left of the images.

3.2.8. FUT11 localizes to endoplasmic reticulum (ER) and golgi in cultured mammalian cells

Immunolocalization studies of FUT11 in cultured mammalian cells found the protein to localize in ER and golgi under endogenous conditions. In HEK293 cells, FUT11 showed localization to ER and golgi. In C6 glioma cells, the protein showed exclusive localization in ER and in SH-SY5Y cells, FUT11 did not reveal any specific localization pattern although some dot-like structures were visible which did not correspond to ER or golgi (Figure 3.9A). The protein co-localization was confirmed with golgi and ER markers (Figure 3.9B).

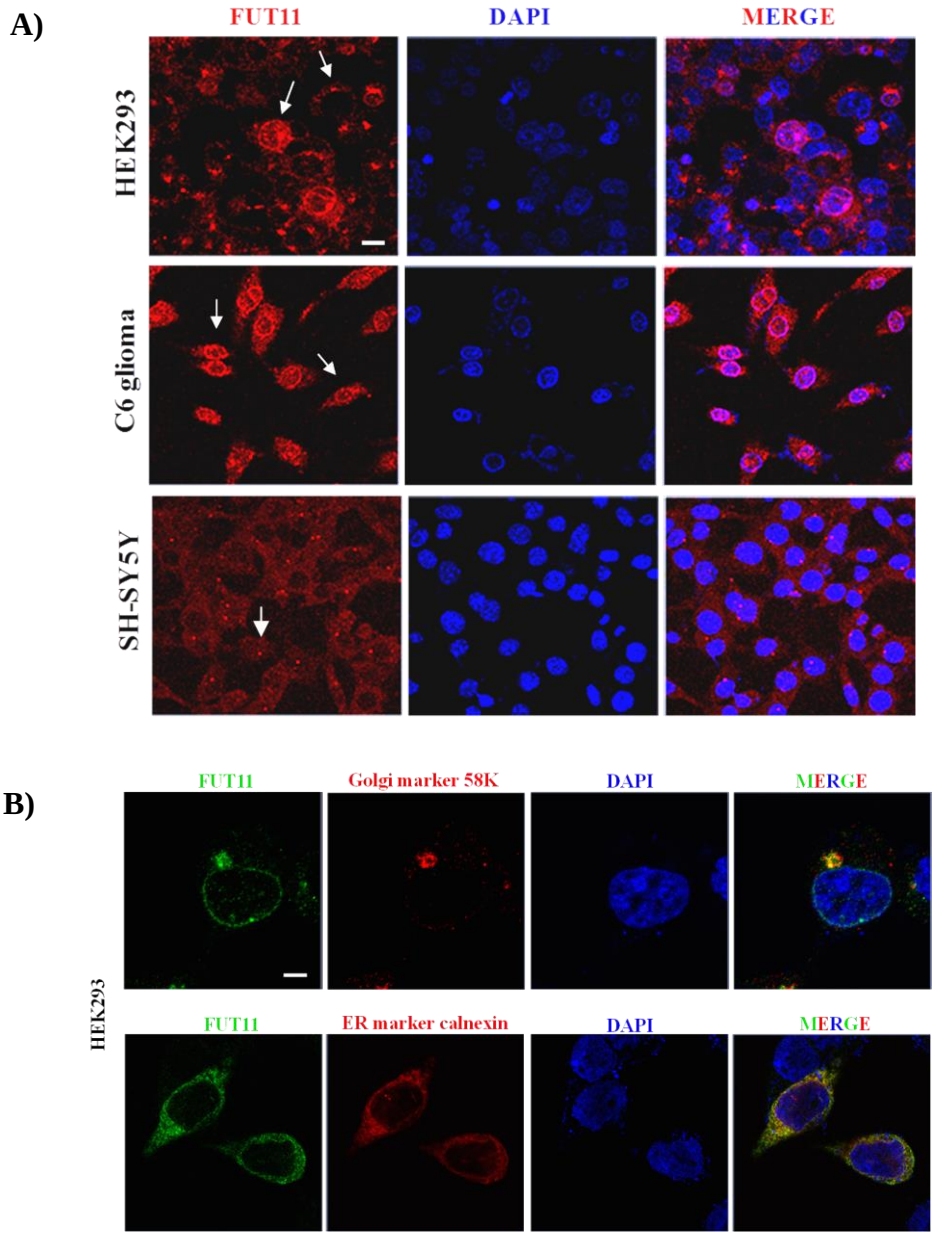


Figure 3.9: FUT11 immunolocalization in cells. (A) Endogenous FUT11 expression in HEK293, C6 glioma and SH-SY5Y cells. Arrows point towards FUT11 localization in ER and golgi. Scale bar=20µm. (B) FUT11 co-localization with ER and golgi markers. Scale=10µm.

3.2.9. Over-expression of FUT11 wildtype and mutants suggests no differential localization in cells

To determine the expression of FUT11 protein in a transient transfection system, HEK293 cells transfected with the wildtype and mutant plasmids were fixed and immunostained. At the antibody dilution of 1:1000 used for analyzing the overexpressed protein, the endogenous protein could not be determined. The over-expressed wildtype protein showed a distinct endoplasmic reticular localization. No specific golgi localization of the over-expressed protein could be detected. The mutant proteins showed similar pattern upon over-expression with no apparent protein localization defects with respect to the wild-type protein (Figure 3.10).

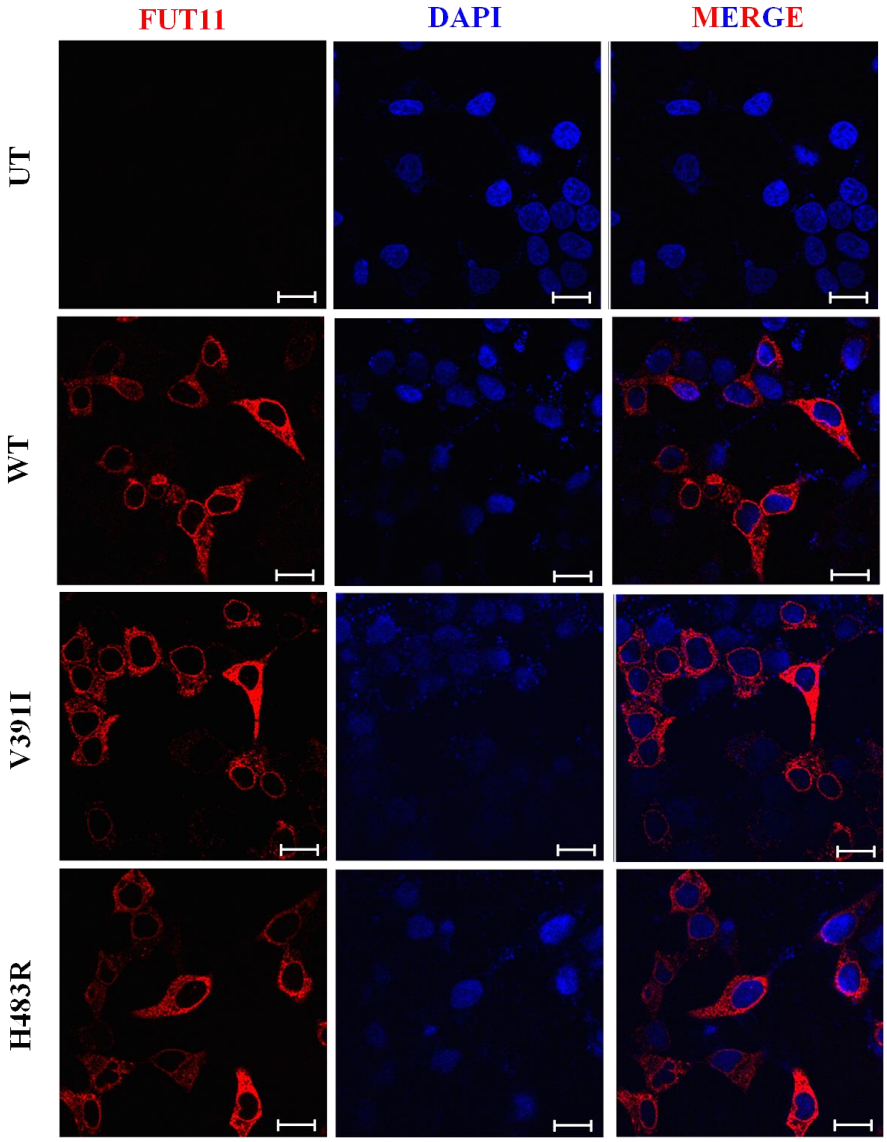


Figure 3.10: FUT11 overexpression in HEK293 cells. Representation of FUT11-wildtype and mutant protein over-expression in HEK293 cells. Scale bar=20µm.

3.3. Discussion

In this chapter, I have presented the whole exome-based genetic analysis of a family, HWE150 wherein the disorder has been previously shown to map to chromosome 10q21.3-q22.3 (Ratnapriya et al 2009b).

I undertook whole-exome based sequencing and targeted analysis of the 10q21.3-q22.3 locus which revealed *FUT11* as the most probable candidate gene for epilepsy in the family. HWE150 is a multi-generation and multi-affected family that segregates HWE in an autosomal dominant manner with several of its members affected with the disorder. Targeted analysis at the 10q21.3-q22.3 locus found three rare segregating variants in the family: *FUT11* (fucosyltransferase 11), c.1171G>A; *NDST2* (N-deacetylase and N-sulfotransferase 2), c.1248+57C>T; and *DLG5* (Disks large homolog 5), c.5165-58C>T that occurred at frequencies of 0.008, 0 and 0.003, respectively, among 288 ethnically matched controls. Bioinformatic analysis did not predict any pathogenic effect on splice site regulation for the two intronic variants. Although comparatively frequent in the control population, the non-synonymous variant in *FUT11* was predicted to be a potentially pathogenic genetic variant. On the basis of this analysis, prediction tools and online interpretation resources, FUT11 was prioritized as the probable gene for HWE in this family. It is a complex task to distinguish disease-causing genetic variants from an abundance of other candidate variants using exome/ genome sequencing for Mendelian diseases (Eilbeck et al 2017). The interpretation and prioritization of such variants can be carried out on the basis of clinical presentation and genotype-phenotype association, and functional assessment and burden testing to establish the plausibility of the gene to cause a disease (Sosnay and Cutting 2014, Eilbeck et al 2017). Bioinformatic analysis provided further evidence for *FUT11* as a candidate for HWE in the family. *FUT11* transcript structure was examined in a cohort of 288 HWE patients that identified two additional rare variants in the gene increasing the supporting evidence for this gene in the disorder.

NDST2 is an 883 amino acids protein and one among the four NDST molecules. NDST2 is an essential bifunctional enzyme that catalyzes the N-deacetylation/N-sulfation of glucosamine (GlcNAc) of the glycosaminoglycan, an initial step in processing the proteoglycan heparan sulphate or heparin (Humphries et al 1999). NDST2 integrates into the glycosaminoglycan metabolism pathway through heparan sulphate metabolism. It is ubiquitously expressed in human tissues. Heparan sulphate proteoglycans aggregate with tau

proteins and are probably involved in Alzheimer's disease (Zhang et al 2014). There are no evidences in literature or mutations known in this gene for any human disorders including brain disorders or epilepsy. In this study, bioinformatic analyses for *NDST2*, c.1248+57C>T using different prediction tools did not support any pathogenic effect of this variant that lies far from the region of predicted splice site. No available evidence in literature to support the role of this protein in the brain or its involvement in brain disorders, and in-silico predictions for this variant allows us to consider it as an unimportant variant for the family.

DLG5 is a 1919 amino acids protein, and one among the seven known DLG proteins that are membrane-associated guanylate kinase homologs (MAGUKs). The protein is expressed ubiquitously in all tissues, and weakly in the brain. DLG5 maintains cell shape and cell polarity (Stoll et al 2004), and is involved in spine formation, synaptogenesis and synaptic transmission in cortical neurons (Wang et al 2014). $Dlg5^{-/-}$ mice present with kidney cysts and hydrocephalus (Nechiporuk et al 2007). Little is known about the function of DLG5. Evidences indicate its role in the developing brain, but no mutations in this gene are reported in human brain disorders. The *DLG5*, c.5165-58C>T intronic variant identified in HWE150 is based away from the region of known splice site. In-silico predictions suggested this variant to be non-pathogenic, with no effect on the resultant protein.

FUT11 is a predicted α-(1,3/4)-fucosyltransferase that catalyzes the transfer of fucose moieties on to substrate molecules, a process known as fucosylation. There are 13 known fucosyltransferases (FUTs) (Table 3.9). Eleven of them are alpha-FUTs, which are terminal modifiers of oligosaccharides. Ten of these localize to golgi. Cellular localization of FUT11 is unknown. Two O-FUTs that add fucose directly to the polypeptide chain (Wang et al 2001) localize to endoplasmic reticulum. *FUT11* was identified as one of the two novel genes in a *Drosophila* genome search to identify enzymes involved in the metabolism of fucosylated glycans, and their counterparts in humans were predicted as putative fucosyltransferases (Roos et al 2002). In mice, FUT11 transcripts are universally expressed (Baboval and Smith 2002). Although predicted to be a type-II transmembrane protein with its N-terminus in the cell cytoplasm, mouse FUT11 overexpression in COS7 cells showed diffused cytoplasmic staining (Baboval and Smith 2002). FUT11 has enzymatic activity (Mollicone et al 2009) but not as much with acceptors used by other previously characterized mammalian FUTs. This is consistent with the phylogenetic analysis that identified FUT11 to belong to an evolutionarily distinct group of FUTs (Roos et al 2002). FUT11 transcript expression was found altered in superior temporal gyrus of elderly patients with

schizophrenia (Mueller et al 2017). Fucosylation is an important aspect of brain proteins where it is a determinant of neuronal morphology and regulator of synaptic proteins (Murrey et al 2006); although it remains unknown why there is high expression of core *fucosylated* N-glycans in *brain* tissues (Gu et al 2015). Fucosylation of brain proteins has been associated with retention of learned behaviour in rats (Wetzel et al 1980). Defucosylation of synapsin protein leads to stunted neurites and delayed synapse formation in mice (Murrey et al 2006).

Table 3.9: Types of fucosyltransferases

Types of fucosyltransferases (FUTs)	FUTs	Localization	Function
α(1,2)-FUTs	FUT1/2	Golgi	H blood group antigen and related structures
α(1,3/4)-FUTs	FUT3	Golgi	$Lewis^x$ and sialyl $Lewis^x$ antigens
α(1,3)-FUTs	FUT4-7, FUT9	Golgi	$Lewis^x$ and sialyl $Lewis^x$ antigens
α(1,6)-FUTs	FUT8	Golgi	Adds fucose to asparigine-linked GlcNAc moieties
Predicted α(1,3)-FUTs	FUT10	Golgi	Orphan FUTs
Predicted α(1,3)-FUTs	FUT11	?	Orphan FUTs
O-FUTs	POFUT1/2	ER	Majorly known in Notch pathway

Modified from Becker and Lowe 2003

Evidence in literature suggests the importance of fucosylation in brain development and function. Bioinformatics analysis of *FUT11*, c.1171G>A in HWE150 supported this variant as potentially pathogenic, which might impact the function of the protein. These suggestions strengthen the evidence to prioritize *FUT11* as the probable candidate gene for epilepsy in this family. Although the possible contribution of the two intronic variants in *NDST2* and *DLG5* cannot be completely disregarded, evidence at this point provides support for *FUT11* as a better candidate for epilepsy.

I found FUT11 to localize in both, endoplasmic reticulum (ER) and golgi in cultured mammalian cells. C6 glioma cells expressed FUT11 specifically in the ER, and HEK293 cells expressed the protein in both ER and golgi. This makes FUT11 the only known α-(1,3/4)-FUT that localizes to ER. Fucosylation being an important post-translational modification in almost all proteins and an important aspect of brain function, the probable role of FUT11 in the human brain and epilepsy cannot be dismissed. Previous phylogenetic studies and the evidence for unusual cellular localization of this protein from the present

study, makes it an interesting molecule with a perceived atypical function. The two genetic variants identified in this study lies in the C-terminal tail of FUT11 that floats within the ER/golgi lumen, the site where glycosylation occurs. Overexpression of the mutant-FUT11 proteins in cultured HEK293 cells did not show any obvious abnormality with respect to the wild-type protein. However, dysfunctional proteins do not necessarily exhibit disrupted localization in cells. Val391Ile lies within a predicted disulfide bond between cysteine at amino acid positions 389 and 392. The variation from Val to Ile adds a bulky methyl group that could disrupt the disulfide bond and disturb the enzymatic activity of the protein. His483 lies further towards the C-terminal end and could influence the stability of this protein and its function.

It is not unusual to find more than one segregating variants in such sequencing studies. Identifying three segregating variants in HWE150 makes it difficult to pinpoint the causative gene in this family. The allele frequency of *FUT11* variant was found to be higher among controls as compared to that for the *NDST2* and *DLG5* variants. Bioinformatic predictions suggested the *FUT11* variant to be functionally damaging. Functionally disruptive gene variants represented at higher frequencies among controls have been identified for Fanconi anaemia (Rogers et al 2014) and cardiomyopathies (Dhandapany et al 2009). The penetrance value of the causative allele in HWE150 is estimated to be 60%, indicating that the allele will not manifest phenotypically in all individuals where it is present. While there is prima facie support for *FUT11* as a candidate gene in the family, it will not be prudent to exclude potential contribution of the *NDST2* and *DLG5* variants. A minigene assay may be carried out to examine whether splice site regulation is affected in these two variants. If results of these assays suggest splice site misregulation, further studies to examine the respective gene/s in HWE patients may be initiated. Negative outcome of these experiments however, would help eliminate their contribution and strengthen the evidence for *FUT11*. This would need to be followed up by a fucosylation assay to determine any differential enzymatic activity of the wild-type and mutant proteins.

In summary, this chapter provides genetic and bioinformatic evidences for *FUT11* as a potential candidate gene for HWE. The study suggests a relatively common genetic variant in this gene underlying HWE in the family. Glycosylation being an important aspect of brain physiology, detailed functional analysis of FUT11 shall help understanding its role in the brain and in epileptogenesis.

Chapter 4

Towards identification of causative genetic locus for hot water epilepsy family, HWE307

Summary

In this chapter, I describe genetic analysis of a family, HWE307, with several of its members affected with hot water epilepsy employing a combined approach of genome-wide linkage mapping and whole exome sequencing. This study led to the identification of three probable genetic loci for HWE307. In the linkage study conducted involving 15 members of the family, two-point LOD scores obtained were: 1.84 (θ=0, 90% penetrance value and 1% phenocopy rate) for D10S1664, D10S1661 at 10p13 and D10S600 at 10p12.1; and 1.58 for D5S2034 at 5q35.1. Eleven additional markers at the 10p13 region and seventeen additional markers at the 5q35.1 region were analyzed to refine the potential haplotypes. This identified two linked regions at 10p14-p12.1 and 5q31.3-q35.3 in the family. The critical genomic regions encompassed 18.7Mb of the genome sequence at 10p14-p12.1 defined by D10S547 and D10S1732 and, 36Mb at 5q31.3-q35.3 defined by D5S2017 and D5S2030. Additionally, manual haplotyping suggested segregation of two markers D7S502 and D7S669 with the phenotype. This region spanned 42Mb of sequences between D7S519 and D7S630 at 7p12.3-7q21.13. I examined gene transcripts at these three genomic intervals in the data obtained from the whole exome sequencing of four affected members of HWE307. All sequence variants: known, novel (unreported) and the ones with minor allele frequencies (MAF) ≤ 0.005 were examined. Among a number of variants obtained, five rare variants were observed: c.5396G>A in *ARHGAP21* and c.2T>C in *C10orf67* at 10p14-p12.1; c.578C>T in *C1QTNF2* at 5q31.3-q35.3; c.383C>T in *WBSCR27* and c.50C>T in *DTX2* at 7p12.3-7q21.13. These five variants co-segregate with the epilepsy phenotype in the family. Further genetic analysis of these gene variants at the three genetic loci, is proposed.

This chapter describes genetic analysis of HWE307 with the goal to identify the locus for hot water epilepsy in the family. The study led to suggestion of three potential loci for HWE at chromosomes 10p14-p12.1, 5q31.3-q35.3, and 7p12.3-7q21.13. A search for epilepsy-

causing mutation/s in these intervals identified five potential rare variants. Further investigation of these variants shall be helpful in unraveling the causative gene in the family.

4.1. Subjects and methods

4.1.1. Family ascertainment and clinical evaluation

A three-generation family, HWE307 (Figure 4.1) was ascertained through the proband, III:5, a 10-year old boy with hot water epilepsy, at Department of Neurology, National Institute of Mental Health and Neurosciences (NIMHANS), Bangalore. III:5 had a history of seizures since the age of 8. The family had a total of 15 individuals available for study, of which 7 were affected, 7 unaffected and 1, an asymptomatic carrier (II:1 was unavailable for study). HWE transmitted in an autosomal dominant manner with incomplete penetrance in HWE307. Family information regarding the age at onset of seizures, type of seizures, timings of seizures in relation to hot water bathing, history of febrile seizures, presence of non-reflex seizures in the family, development of non-reflex seizures in the HWE patients and family histories of epilepsy were obtained from patients and their surrogate respondents who had observed several of the seizure episodes. Diagnoses were made by qualified neurologist by evaluating clinical histories of seizures from the patients supported with the information obtained from an eye witness of the seizures. Ictal EEGs were not available as they are difficult to obtain in HWE. A ten millilitre blood sample was obtained by venous puncture after obtaining written informed consent from all the participating members of the family. Genomic DNA was isolated from the peripheral white blood cells using a standard phenol-chloroform method (Sambrook and Russell, 2001). The study had approval of the institutional bioethics and biosafety committees.

4.1.2. Whole genome-wide linkage mapping

A genome-wide linkage mapping was carried out using 382 fluorescence-labeled microsatellite markers of the ABI PRISM MD-10 Linkage Mapping Set v2.5 (Applied Biosystems, USA), spread over the autosomes at an average resolution of 9.5±2.8cM. The marker order and the inter-marker distances were determined from the Généthon (Dib et al 1996) and deCODE (Kong et al 2002) linkage maps. Individual DNA samples were amplified for the markers using PCR on a GeneAmp9700 machine (Applied Biosystems). The amplicons generated by the labeled primers (FAM: blue, VIC: green, NED: yellow) were pooled according to a pre-categorized panel of markers for each of the family members examined. Polymerase Chain Reactions (PCRs) were performed with 50ng genomic DNA,

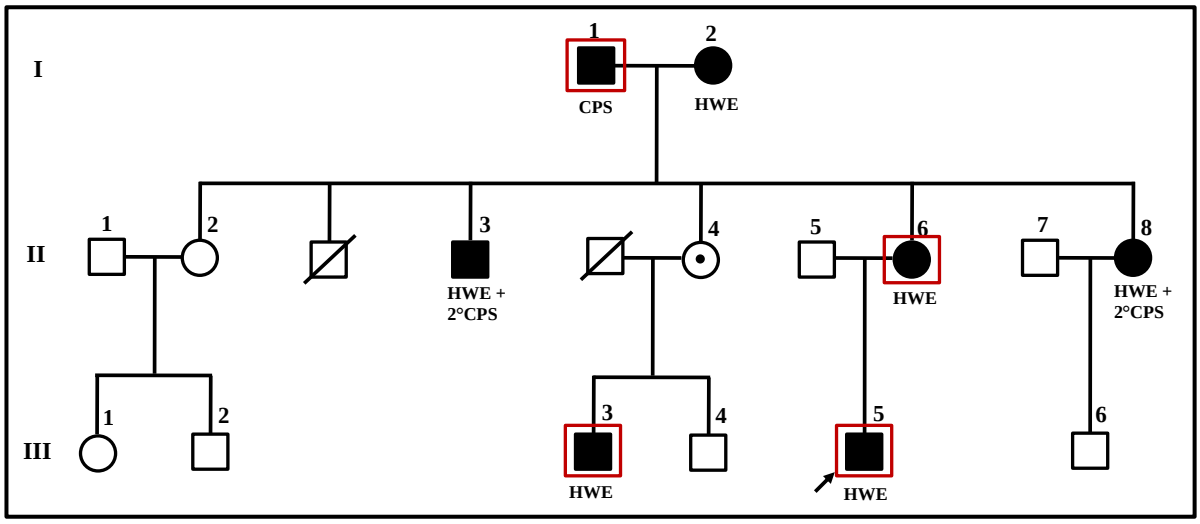


Figure 4.1: Family pedigree of HWE307. The three generations of family are marked as I, II and III on the left hand side of the pedigree. Males are denoted by squares, females by circles, the filled symbols denote affected individuals and the unfilled ones denote unaffecteds. Symbol with a dot represents an asymptomatic carrier. The proband III:5 is marked by an arrow pointing towards the symbol. CPS: complex partial seizure, HWE: hot water epilepsy, 2°CPS: secondary complex partial seizure. Individuals *boxed in red* were exome sequenced.

5µM of each primer, 9µl of True Allele™ PCR premix (Applied Biosystems) in a 15µl reaction volume. The pooled products were mixed with a cocktail of Hi-Di Formamide and GeneScan™ LIZ500 (Applied Biosystems) as an internal size standard, assembled into a 96-well microtitre plate, thermally denatured at 96°C, snap chilled and electrophoresed on an ABI3730 Genetic Analyzer (Applied Biosystems). DNA from a reference individual CEPH (1347-02) of known genotype was used as a control in each electrophoresis run. The allele sizing and genotype assignment was done using Genescan Analysis v3.7 and ABI PRISM GeneMapper™ v3.7, respectively.

LOD score analysis: Parametric two-point LOD scores were calculated using MLINK v.5.2 of LINKAGE (Lathrop and Lalouel 1984) considering an autosomal dominant mode of inheritance for HWE307, for penetrance values 60-90% at 1% phenocopy rate and a disease allele frequency of 0.0001. Recombination rate (θ) was considered to be same in males and females and the allele frequencies for individual markers were determined from the family data available. Multipoint LOD score analysis was carried out using GENEHUNTER v2.1 (Kruglyak et al 1996) of easyLINKAGE v5.02 (Linder and Hoffmann 2005). Additional genetic mapping markers spaced on an average of 2cM apart, were examined at the three

regions with LOD score greater than 1.0 to gather more information regarding the haplotype sharing and recombination boundaries. Eleven additional markers at 10p13 locus (D10S602, D10S591, D10S1779, D10S570, D10S1664, D10S191, D10S1661, D10S211, D10S600, D10S1732, D10S193) (Appendix A4.1); seventeen markers at 5q35.1 locus (D5S2098, D5S1984, D5S500, D5S2017, D5S2090, D5S2014, D5S2112, D5S2066, D5S415, D5S671, D5S625, D5S429, D5S2069, D5S498, D5S2034, D5S2030, D5S2006) (Appendix A4.2); and thirteen markers at 3q28 locus (D3S3689, D3S3668, D3S1282, D3S1556, D3S3730, D3S3592, D3S3686, D3S3530, D3S1314, D3S2748, D3S1305, D3S1265, D3S3550) (Appendix A4.3) were examined using the same protocol as mentioned above.

Haplotype analysis: Haplotypes for markers across all the chromosomes were generated manually according to the marker order lists of Généthon and deCODE, allowing for minimum number of inter-marker recombination events, and checked with MaxProb GENEHUNTER v2.1 generated haplotypes. Additionally, haplotypes at the 10p13, 5q35.1 and 3q28 regions were analyzed manually for the incorporated fine mapping markers.

Manual haplotype analysis of the whole genome was performed to eliminate a chance of missing out a region of shared markers that might have gone undetected in LOD score analysis. These haplotypes were checked with MaxProb GENEHUNTER v2.1 generated haplotypes. The analysis was carried out considering the following three possibilities (i) haplotypes inherited and shared in the family from affected individual I:1, (ii) haplotypes inherited and shared in the family from affected individual I:2, (iii) haplotypes inherited and shared in the family from both the affected individuals I:1 and I:2.

Regions of the genome with LOD scores above 1.0 and of shared haplotype were considered for targeted analysis in the whole exome data. Chromosome intervals 10p14-p12.1, 5q31.3-q35.3 and 7p12.3-7q21.13 were pursued for analysis in the whole exome dataset.

4.1.3. Positional candidate gene sequencing by Sanger-based method

As a first approach towards searching for causative genes in HWE307, the regions 10p14-p12.1 and 5q31.3-q35.3, respectively, were prioritized for candidate gene sequencing. Twenty nine genes were selected through literature study, based on their expression in human brain and/or involvement in epilepsy or other neurological disorders (Table 4.1). The exonic-, flanking intronic- and untranslated regions of 19 genes at 10p14-p12.1 and 10 genes at 5q31.3-q35.3 were examined to identify potential causative mutation, using DNA from the proband III:5.

Table 4.1: Candidate genes examined at the 10p14-p12.1 and 5q31.3-q35.3 intervals

10p14-p12.1		
Gene symbol	**Gene name**	**Physical location, strand**
CELF2	CUGBP, Elav-like family member 2	10838851-11378674, +ve
CAMK1D	Calcium/calmodulin-dependent protein kinase 1D	12391583-12871735, +ve
OPTN	Optineurin	13142082-13180276, +ve
FRMD4A	FERM domain containing 4A	13685706-14372866, -ve
CDNF	Cerebral dopamine neurotrophic factor	14861251-14879983, -ve
HSPA14	Heat shock 70kDa protein 14	14880261-14913740, +ve
NMT2	Glycylpeptide N-tetradecanoyltransferase 2	15147771-15210695, -ve
ITGA8	Integrin, alpha 8	15559088-15761770, -ve
RSU1	Ras suppressor protein 1	16632615-16859453, -ve
CUBN	Cubilin (intrinsic factor-cobalamin receptor)	16865965-17171816, -ve
SLC39A12	Solute carrier family 39 (zinc transporter), member 12	18240768-18332221, +ve
ARL5B	ADP-ribosylation factor-like 5B	18948313-18966940, +ve
PLXDC2	Plexin domain containing 2	20105372-20569115, +ve
MLLT10	Myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 10	21823101- 22032559, +ve
PIP4K2A	Phosphatidylinositol-5-phosphate 4-kinase, type II, alpha	22823766-23003503, -ve
GAD2	Glutamate decarboxylase 2 (pancreatic islets and brain,65kDa)	26505236- 26593491, +ve
YME1L1	YME1-like 1 (S. cerevisiae)	27399383-27443321, -ve
RAB18	RAB18, member RAS oncogene family	27793249-27829099, +ve
MPP7	Membrane protein, palmitoylated 7 (MAGUK p55 subfamily member 7)	28339922-28571067, -ve
5q31.3-q35.3		
Gene symbol	**Gene name**	**Physical location, strand**
GRIA1	Glutamate receptor, ionotropic, AMPA1	152870084-153193429, +ve
SAP30L	Sin3A Associated Protein P30-like	153825517-153840613, +ve
SGCD	Sarcoglycan, delta (35kDa dystrophin-associated glycoprotein)	155753767-156194799, +ve
ADAM19	A Disintegrin And Metalloproteinase Domain 19	156904312-157002783, -ve
PTTG1	Pituitary tumor-tansforming 1	159848865-159855746, +ve
GABRB2	Gamma aminobutyric acid (GABA) A receptor, beta 2	160715436-160975130, -ve
GABRA6	Gamma aminobutyric acid (GABA) A receptor, alpha 6	161112658-161129598, +ve
GABRA1	Gamma aminobutyric acid (GABA) A receptor, alpha 1	161274197-161326965, +ve
GABRG2	Gamma aminobutyric acid (GABA) A receptor, gamma 2	161494648-161582545, +ve
KCNIP1	Kv channel interacting protein 1	169780881-170163636, +ve

The gene sequences were obtained from Genebank sequence database NCBI, NIH, USA. Oligonucleotide primers for the respective gene regions (Appendix A4.6) were designed using Primer3 v0.4.0 (http://bioinfo.ut.ee/primer3-0.4.0/primer3/). The PCR amplified products were purified using Millipore Multiscreen $PCR_{\mu 96}$ filter plates and cycle sequencing was performed using ABI PRISM BigDye^{TM} Terminator v3.1 cycle sequencing reagents. Following cycle sequencing, the samples were ethanol precipitated, resuspended in formamide, denatured and examined on an ABI PRISM 3730 DNA Analyzer. The comparative sequence analysis was performed using SeqMan II 5.01 (DNASTAR Inc., USA).

4.1.4. Whole exome sequencing

To examine a large number of genes in a single experiment, whole exome sequencing was carried out on four affected members, I:1, II:6, III:3 and III:5 of the family. Exome capture was performed according to standard sample preparation protocol using NEXTERA Rapid Capture Expanded Exome kit (Illumina Inc., USA). The kit covers 62Mb (1.7%) of genomic sequence comprising coding exons, untranslated regions (UTRs) and microRNAs with their respective probes constructed against the human GRCh37/hg19 reference genome. DNA library was prepared using 50ng of genomic DNA that was fragmented into 300bp size lengths and tagged with adapter sequences using Nextera transposomes, in a single step process called tagmentation. Tagmented DNA was purified from the transposomes using DNA-binding beads and filter purification system and assessed on an Agilent Technologies 2100 Bioanalyzer using a High Sensitivity DNA chip (Agilent Technologies, USA). The purified DNA tagments were indexed, added with common adapters for cluster generation and enriched through a PCR step. The DNA library was purified with AMPure XP beads that provide a size selection step to remove short library fragments. Quality control check was performed using an Agilent Technologies Bioanalyzer 2100 using DNA 1000 chip. The libraries were normalized to the same dilution, pooled and hybridized twice with biotinylated probes (temp, time: 58°C, 3hrs and 58°C, 16hrs). The target DNA regions were captured with streptavidin magnetic beads and heat washed to remove non-specific binding to the beads. This enriched DNA was again hybridized with capture probes to ensure high specificity of the target regions and purified using the same protocol. These libraries were enriched through a PCR step, bead purified and quantified for optimum cluster density. Cluster was generated on cBOT and sequencing was performed on an Illumina HiSeq 2500 platform using a paired-end 2x 100bp protocol (Illumina Inc.).

4.1.5. Whole exome sequencing analysis

The raw intensity output files from the Illumina Hiseq 2500 were demultiplexed and converted into raw reads using bcl2fastq (Illumina Inc.). These reads were trimmed in two steps. First, any adapter or primer contaminated reads were removed and then, the reads were trimmed on the basis of Phred score of bases. Bases with score ≤ 20 at the 3'ends of reads were removed. Only reads with at least 70% bases having a Phred score more than 20 were considered. This was performed using *in-house* scripts and Cutadapt. These processed reads were aligned to the NCBI reference human genome GRCh37/hg19 using Burrows-Wheeler Aligner (Li and Durbin 2009). Post-alignment, the files were processed for variant calling and to reduce false positives. They were chromosomally sorted and PCR duplicates were removed using SAMtools (Li et al 2009), locally realigned around InDels and base quality scores were recalibrated using Genome Analysis Toolkit (GATK) (DePristo et al 2011) and coverage was calculated using *in-house* scripts. Variants were called using the HaplotypeCaller module of GATK and annotated against public databases dbSNP144, 1000 Genomes phase3 and Exome Aggregation Consortium (ExAC) 3.0 using *in-house* scripts and Annovar. The SNVs and InDels obtained for the HWE307 exome sequenced individuals, I:1, II:6, III:3 and III:5, were segregated according to each individual and the ones which were common in two, three or all four individuals. Functional annotation of the variants was performed using SnpEff (Cingolani et al 2012). Bioinformatic prediction of variant impact on protein function was determined by SIFT (Kumar et al 2009) and PolyPhen-2 (Adzhubei et al 2010). The variants examined were covered by at least 5x of read depth and called from both the DNA strands

.4.1.6. Analysis of three genomic regions in HWE307 members

The whole exome sequence analysis for HWE307 affected members: I:1, II:6, III:3 and III:5, was carried out with information obtained from LOD score and manual haplotype analyses. The regions 10p14-p12.1 defined by D10S547 and D10S1732, 5q31.3-q35.3 defined by D5S2017 and D5S2030 and 7p12.3-7q21.13 defined by D7S519 and D7S630 were taken forward for examination in the whole exome dataset. All features obtained post data processing comprised variants that were categorized as, (i) total variants present in the four individuals, (ii) variants common in all four individuals and (iii) variants common in at least three or two of the individuals. Variants belonging to this third category were unique, and none of them overlapped with those in category ii. The variants were filtered for (i) coding and 5'UTR, and (ii) novel (unreported) and rare variants which were reported in

databases with a MAF ≤ 0.005. All variants that passed these filters were analyzed for the three regions: 10p14-p12.1, 5q31.3-q35.3 and 7p12.3-7q21.13. Exclusive variants within these regions that might have gotten missed in either individual were also considered for Sanger-based examination. Variants, inclusive of those at a read depth lower than 5X were analyzed manually, by aligning them in SAMtools against the reference genomic DNA GRCh37/hg19 (UCSC). These variants were prioritized based on allele heterozygosity and haplotype sharing for the three regions: presence in all four individuals for the regions 10p14-p12.1 and 5q31.3-q35.3, and presence in three individuals (II:6, III:3 and III:5) for the region 7p12.3-7q21.13. The variants obtained from this analysis were examined in an *in-house* exome sequenced set of 52 control individuals. The filtered variants were validated by bi-directional Sanger sequencing and examined for segregation in the family.

4.1.7. Sanger-based validation of variants

Sanger sequencing was carried out to validate the variants obtained from the whole exome dataset. Primers were designed spanning the variant carrying regions (Appendix A4.7) and confirmed by bi-directional sequencing of the four exome sequenced individuals, as described previously. Sanger sequencing was further employed to determine the co-segregation of true variants (variants obtained from exome sequencing and confirmed by Sanger) with the phenotype in the family. Co-segregating variants were examined in an ethnically matched control set of at least 192 normal individuals.

4.1.8. Bioinformatic analysis

The nucleotide and protein sequences for different species were obtained for *C10orf67* at 10p12.2, *ARHGAP21* at 10p12.1, *C1QTNF2* at 5q33.3, *WBSCR27* at 7q11.23 and *DTX2* at 7q11.23 from NCBI Map Viewer database, 37.3 (http://www.ncbi.nlm.nih.gov/mapview/). Multiple sequence alignment for conservation analysis was performed with NCBI HomoloGene (http://www.ncbi.nlm.nih.gov/homologene/). Functional regions and domains of the protein were analyzed using protein BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

4.2. Results

4.2.1. Whole genome-wide linkage analysis

In the parametric two-point LOD score analysis of Family 307, I obtained LOD scores above 1.0 for three markers: 1.58 for D10S1653 at 10p13, 1.23 for D5S400 at 5q35.1 and 1.14 for

D3S1601 at 3q28, for recombination frequency θ=0, 90% penetrance value and 1% phenocopy rate.

LOD score and haplotype analysis at 10p14-p12.1: A LOD score of 1.58 for D10S1653 at 10p13 was obtained. Following addition of markers for the fine mapping analysis, I obtained a highest LOD score of 1.84 at θ=0 for the two markers D10S1661 and D10S600 (Table 4.2). Parametric multipoint LOD score analysis in the region provided a score of 1.84 for the markers D10S1664-D10S191-D10S1653-D10S1661 (Appendix A4.4). Haplotype analysis indicated a 9- marker region absent in I:2 and transmitted from I:1. This region of 18.7Mb of genome sequence at 10p14-p12.1 was present among all the affected members and the asymptomatic carrier II:4 (Figure 4.2). The recombination events were marked between D10S547 and D10S570 (centromere-distal) in II:6 and III:5 and between D10S600 and D10S1732 (centromere-proximal) in III:3.

Table 4.2: Representative LOD scores for the interval 10p14-p12.1 in HWE307

	LOD scores (Z) at recombination values (θ)				
Markers[a]	**θ = 0**	**θ = 0.1**	**θ = 0.2**	**θ = 0.3**	**θ = 0.4**
D10S547[c,e]	-1.90	-0.11	0.00	0.01	0.00
D10S570[b,d,e]	1.02	0.77	0.51	0.25	0.06
D10S1664[b,d]	1.84	1.45	1.02	0.55	0.15
D10S191[b,d]	1.29	0.93	0.56	0.24	0.05
D10S1653[b,c]	1.58	1.23	0.84	0.44	0.11
D10S1661[b,d]	1.84	1.45	1.02	0.56	0.16
D10S548[b,c]	-0.18	-0.01	0.03	0.02	0.00
D10S211[b,d]	0.00	0.00	0.00	0.00	0.00
D10S197[b,c]	0.65	0.49	0.33	0.17	0.05
D10S600[b,d,e]	1.84	1.44	1.02	0.56	0.15
D10S1732[d,e]	-0.11	0.57	0.48	0.27	0.07

[a] Marker order is as per the human genome physical map (Human Genome Map Viewer Build 37.3 database, NCBI, NIH, USA), [b] Markers at the linked region, [c] Panel markers, [d] Fine mapping markers, [e] Recombination boundaries.

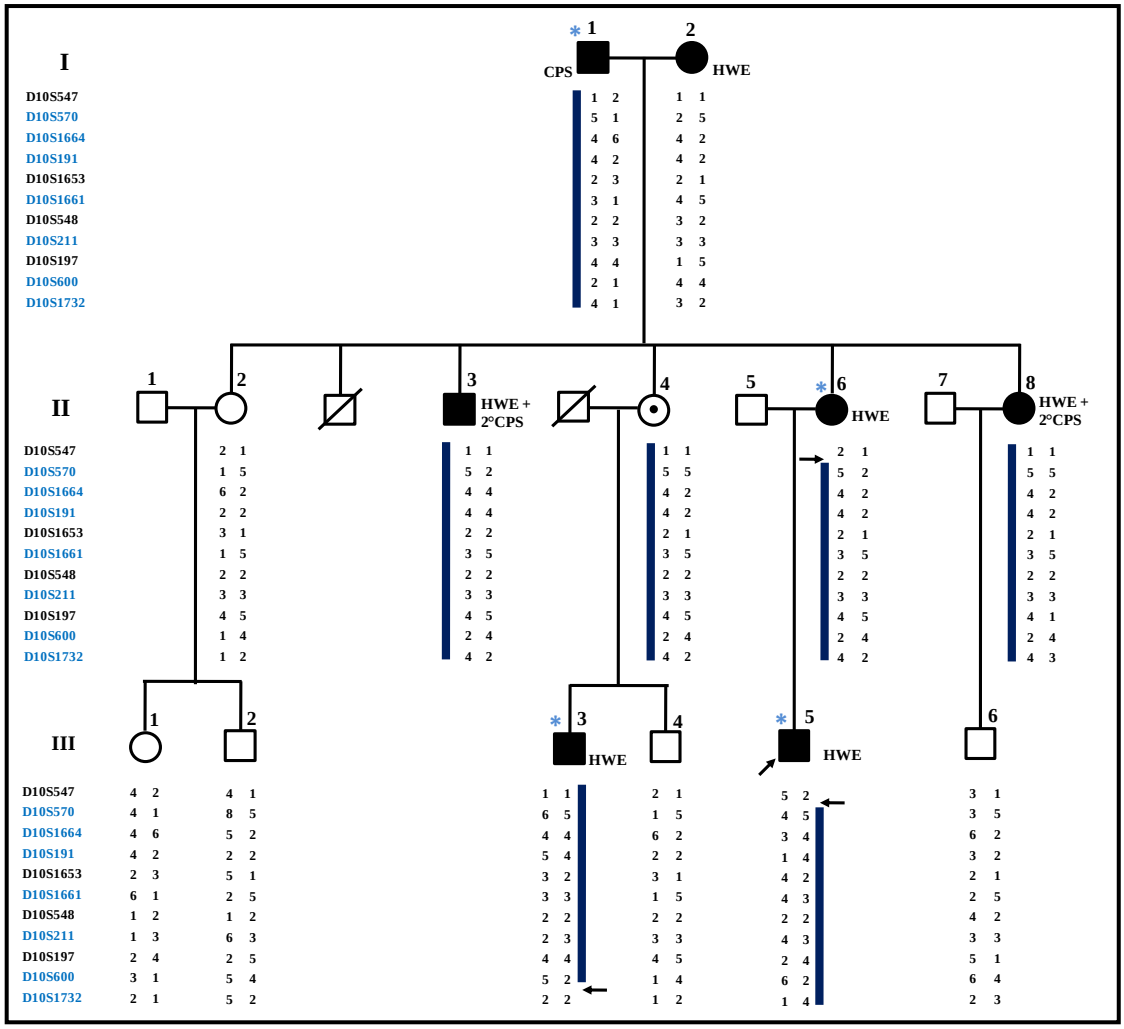


Figure 4.2: HWE307 depicting haplotype at 10p14-p12.1: Representative markers for the region are shown in a telomere-to-centromere order towards the left. Panel markers are represented in black and fine mapping markers in blue. Haplotypes for all the individuals are shown below the symbols. The blue bar represents the shared haplotype. The black arrows represent the recombination boundaries. The four whole exome sequenced individuals are marked with *.

LOD score and haplotype analysis at 5q31.3-q35.3: A LOD score of 1.23 for D5S400 at 5q35.1 was obtained. Fine mapping analysis provided a LOD score of 1.58 at θ=0 for D5S2034 (Table 4.3). The parametric multipoint LOD score obtained was 1.56 at the markers D5S410-D5S2112-D5S422 (Appendix A4.5). Haplotype analysis found a set of 15-marker region spanning 36Mb of genome sequence. This region was absent in I:2, segregated from I:1, shared by all the affected members and the apparently asymptomatic carrier II:4 (Figure 4.3). The recombination events occurred between D5S2017 and D5S436 (centromere-proximal) in II:8 and III:5 and between D5S2034 and D5S2030 (centromere-distal) in III:5. Portions of this region were shared by two unaffected individuals III:4 and

III:6. Four markers in the upper portion of the haplotype and ten in the lower half, were shared by unaffecteds III:4 and III:6 respectively, except for a single marker D5S2112, shared only by the affecteds and the asymptomatic carrier. This small region of 9.37Mb spanning 5q33.1-q34 was hence critical for further consideration.

Table 4.3: Representative LOD scores for the interval 5q31.3-q35.3 in HWE307

	LOD scores (Z) at recombination values (θ)				
Markers[a]	**θ = 0**	**θ = 0.1**	**θ = 0.2**	**θ = 0.3**	**θ = 0.4**
D5S2017[d,e]	-1.20	-0.57	-0.24	-0.08	-0.01
D5S436[b,c,e]	0.03	0.13	0.11	0.07	0.02
D5S2090[b,d]	0.84	0.79	0.58	0.31	0.07
D5S2014[b,d]	0.49	0.51	0.38	0.19	0.04
D5S410[b,c]	0.24	0.32	0.25	0.14	0.04
D5S2112[b,d]	1.28	0.99	0.65	0.33	0.08
D5S422[b,c]	0.33	0.36	0.28	0.16	0.05
D5S2066[b,d]	0.52	0.53	0.39	0.21	0.06
D5S415[b,d]	1.03	0.78	0.52	0.26	0.06
D5S671[b,d]	0.85	0.78	0.57	0.31	0.08
D5S400[b,c]	1.23	0.93	0.63	0.32	0.08
D5S625[b,d]	-0.63	-0.22	-0.06	0.00	0.00
D5S429[b,d]	1.28	0.98	0.66	0.34	0.08
D5S2069[b,d]	0.80	0.74	0.54	0.29	0.08
D5S498[b,d]	0.84	0.77	0.56	0.31	0.08
D5S2034[b,d,e]	1.58	1.23	0.85	0.45	0.12
D5S2030[d,e]	-3.00	-0.16	0.01	0.03	0.01

[a] Marker order is as per the human genome physical map (Human Genome Map Viewer Build 37.3 database, NCBI, NIH, USA), [b] Markers at the linked region, [c] Panel markers, [d] Fine mapping markers, [e] Recombination boundaries.

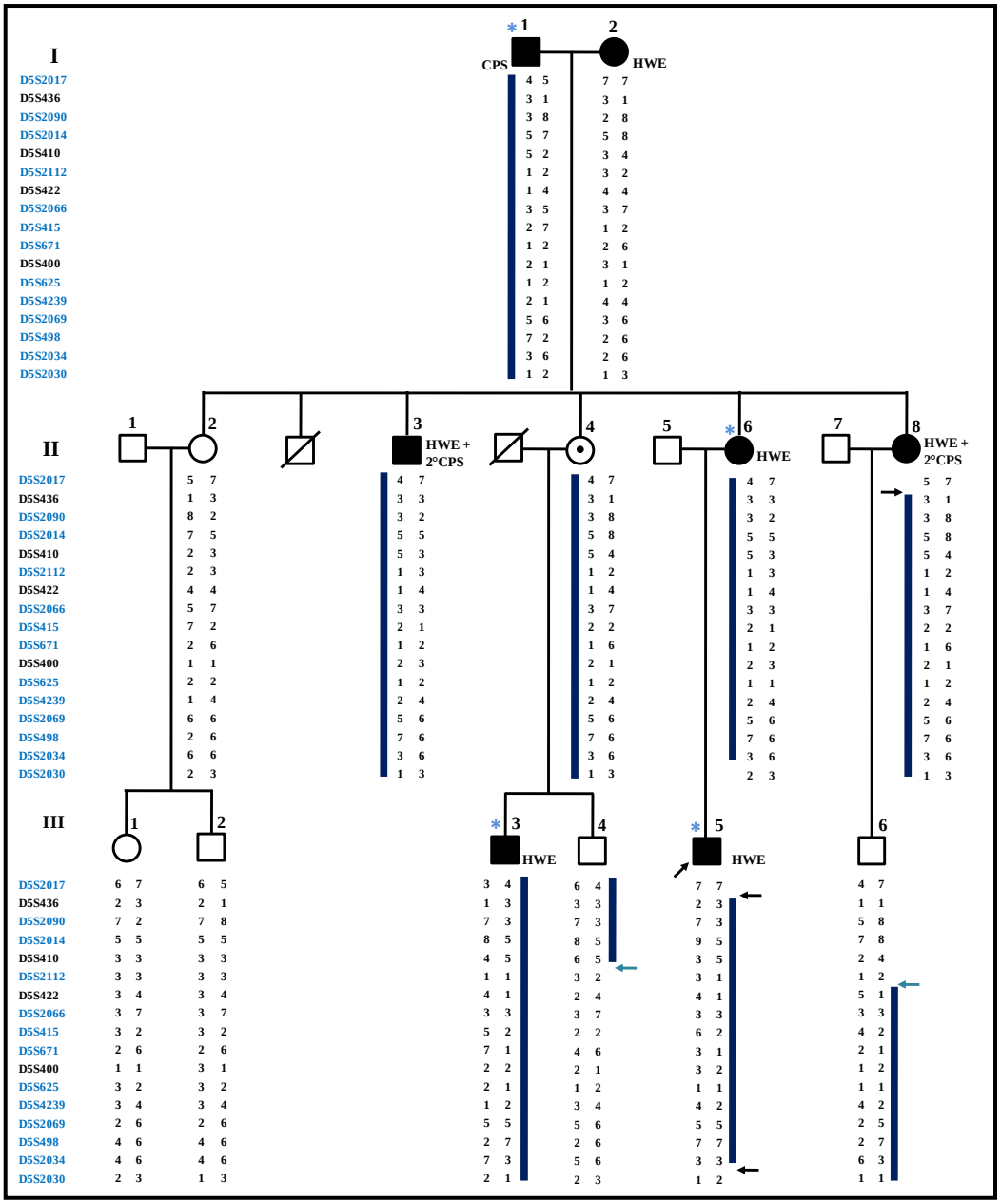


Figure 4.3: HWE307 depicting haplotype at 5q31.3-q35.3: Representative markers for the region are shown in a centromere-to-telomere order towards the left. Panel markers are represented in black and fine mapping markers in blue. Haplotypes for all the individuals are shown below the symbols. The blue bar represents the shared haplotype. The black arrows represent the recombination boundaries. The blue arrows represent the single marker region not shared by the two unaffecteds. The four whole exome sequenced individuals are marked with *.

LOD score and haplotype analysis at 3q28: D3S1601 at 3q28 that yielded a LOD score of 1.14 was not informative in affected individuals II:8 and III:5. Fine mapping analysis of the 15Mb region at 3q28 provided a LOD score of 1.3 for markers D3S3530 and D3S1314. These were however, uninformative genotypes with low heterozygosity. Haplotype analysis

of this 15Mb region did not show any segregation with the disease phenotype. Consequently, the 3q28 region was not considered for further examination.

4.2.2. Manual haplotype analysis of the whole genome markers

Manual haplotype analysis found 23 additional markers across 7 chromosomes that could be of relevance. Among these, 21 markers were present in several unaffected members and hence considered unimportant for further consideration. I found a two-marker haplotype comprising D7S502 and D7S669, absent in I:1, inherited from I:2 and shared by all the affected individuals, one unaffected individual II:2 and asymptomatic carrier II:4 (Figure 4.4). A LOD score of 0.84 and 0.72 for D7S502 and D7S669, respectively had been obtained at θ=0, 90% penetrance value and 1% phenocopy rate. The region at 7p12.3-7q21.13 was defined by recombination boundaries between markers D7S519 and D7S502 (on the p arm) in individual I:2 and between markers D7S669 and D7S630 (on the q arm) in individuals II:4 and III:3. No other marker in the genome was found to be shared by all the affected individuals in HWE307.

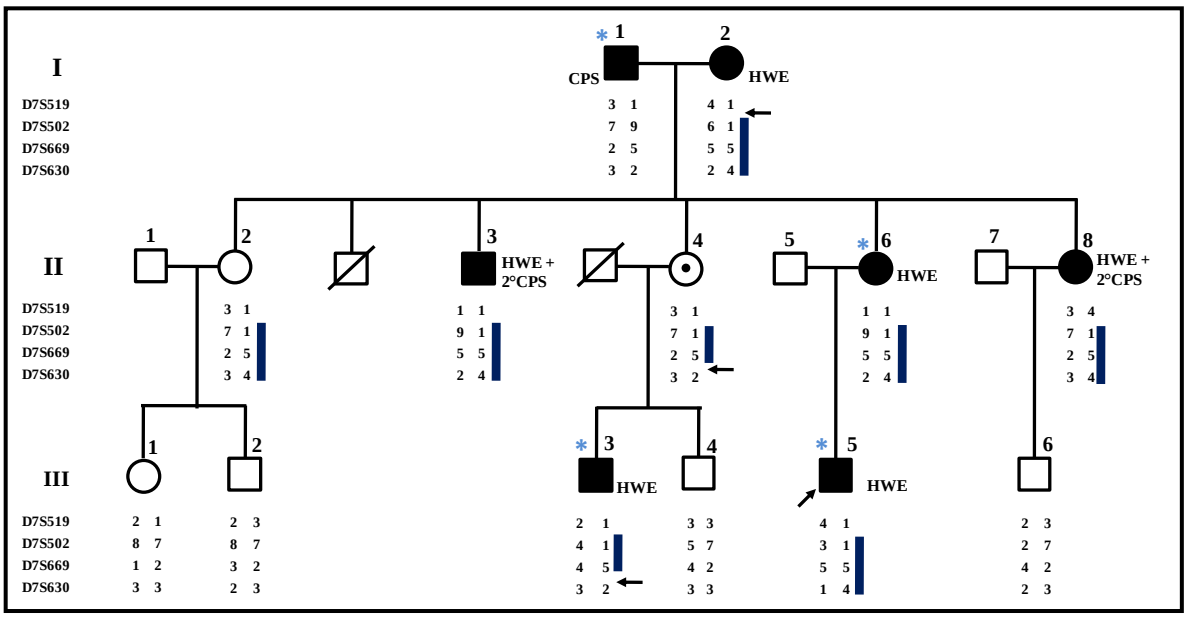


Figure 4.4: HWE307 depicting haplotype at 7p12.3-7q21.13: Representative markers for the region are shown towards the left in a telomere-to-telomere orientation across the centromere. Haplotypes for all the individuals are shown below the symbols. The blue bar represents the shared haplotype. The black arrows represent the recombination boundaries. The four whole exome sequenced individuals are marked with *.

4.2.3. Sequencing analysis

As a first step towards gene/variant identification, twenty nine positional candidate genes were sequenced. No segregating causative variants were found in them. Following which,

whole exome sequencing analysis was undertaken in four members of the family. Whole exome sequencing of the four individuals provided 99% coverage for the target regions with high quality reads (Tables 4.8 and 4.9). Data obtained from whole exome sequencing found a total of 618,150 variants in I:1; 597,851 in II:6; 819,416 in III:3 and 550,818 in III:5 (Table 4.4). The variant count was inclusive of known and novel SNVs and InDels. Post filtering them for (i) common in four, common in three and common in two of the exome sequenced individuals (Tables 4.5, 4.6, 4.7), (ii) coding and 5'UTR, (iii) novel and rare variants, (iv) the three regions of interest at 10p14-p12.1, 5q31.3-q35.3 and 7p12.3-7q21.13, and (v) manual analysis in SAMtools; a total of 18 relevant variants were found in the three regions of interest (Figure 4.5). Variants examined were eliminated on the basis of being (i) homozygous in one or more affected individuals, (ii) present in only two or three of the affected individuals, (iii) false reads, after Sanger confirmation and variants annotated at repetitive sequence regions, and (iv) frequency of more than 0.005 in an *in-house* control set of 52 individuals. Nine among the 18 variants were confirmed for both the strands of DNA in the four exome sequenced individuals and examined for segregation in the family (Table 4.10, 4.11). Among these nine variants, three each lied within the critical regions at 10p14-p12.1, 5q31.3-q35.3 and 7p12.3-7q21.13.

Table 4.4: Total number of variants in the four exome-sequenced patients

Sample ID	SNV Count			In-Del Count			Total variations
	Novel	Known	Total	Novel	Known	Total	
I:1	15,805	536,187	551,992	5,886	59,755	65,641	618,150
II:6	15,320	517,710	533,030	5,640	82,754	64,362	597,851
III:3	19,708	710,709	730,417	7,660	80,734	88,394	819,416
III:5	14,657	476,418	491,075	5,228	54,084	59,312	550,818

Table 4.5: Number of variants shared among all the four patients

Sample ID	SNV Count			In-Del Count			Total variations
	Novel	Known	Total	Novel	Known	Total	
Shared among I:1, II:6, III:3 and III:5	3,114	82,465	85,579	561	9,469	10,030	95,609

Table 4.6: Number of variants shared among three patients

Sample ID	SNV Count			In-Del Count			Total variations
	Novel	Known	Total	Novel	Known	Total	
I:1, II:6, III:3	1,016	25,625	26,641	257	2,848	3,105	29,747
I:1, II:6, III:5	774	21,983	22,757	200	2,298	2,498	25,255
I:1, III:3, III:5	766	16,816	17,582	194	1,887	2,081	19,663
II:6, III:3, III:5	868	19,357	20,225	215	2,225	2,440	22,665

Table 4.7: Number of variants shared among two patients

Sample ID	SNV Count			In-Del Count			Total variations
	Novel	Known	Total	Novel	Known	Total	
I:1, II:6	1,019	41,833	42,618	345	4,115	4,420	47,053
I:1, III:3	1,392	60,237	61,629	461	6,120	6,581	68,221
I:1, III:5	753	31,640	32,393	257	3,130	3,387	35,791
II:6, III:3	1,318	54,897	56,215	445	5,693	6,138	62,359
II:6, III:5	829	35,940	36,769	301	3,706	4,007	40,776
III:3, III:5	1,124	46,535	47,659	349	4,745	5,094	52,776

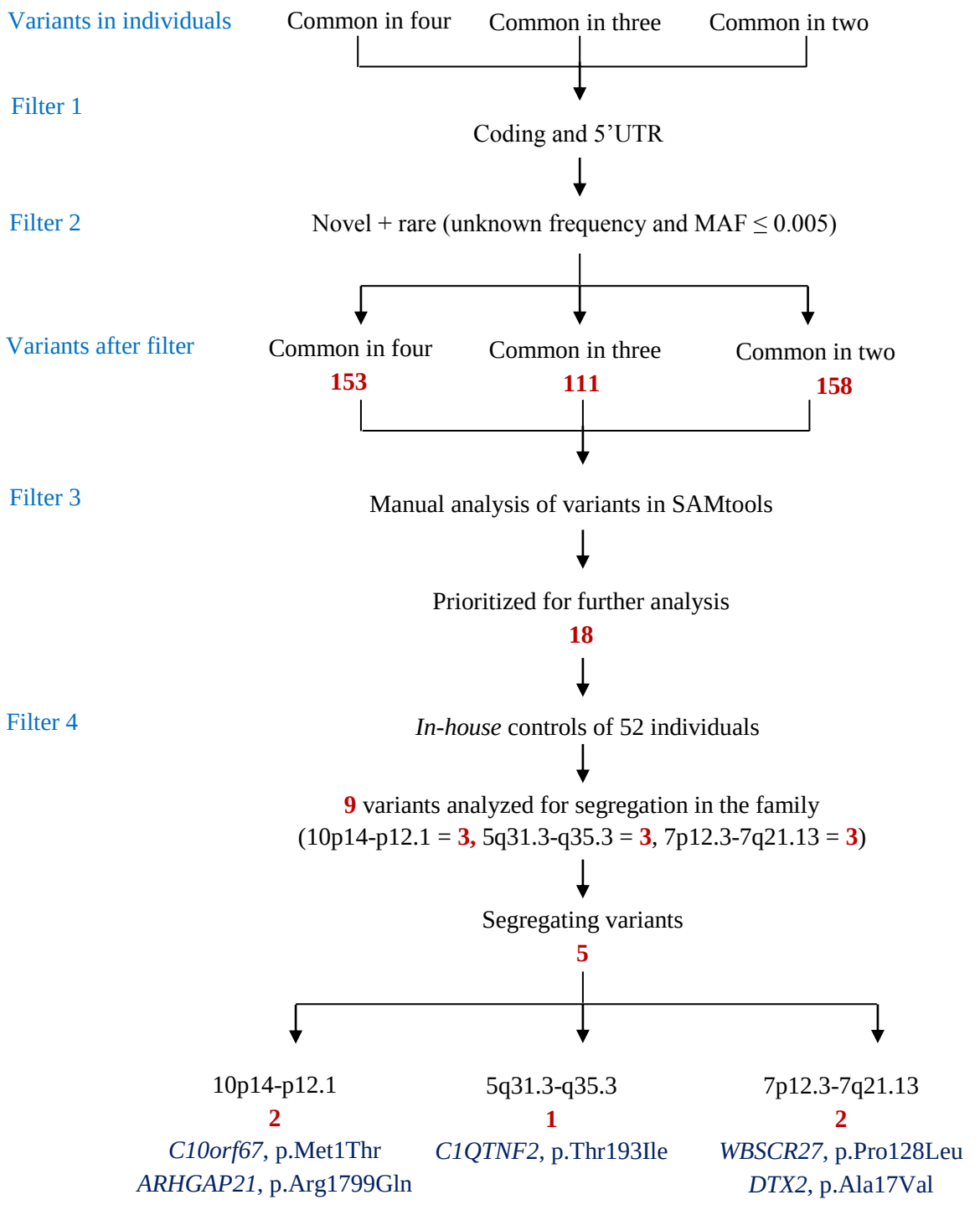


Figure 4.5: Analysis protocol employed for whole exome data analysis: Depiction of the protocol employed for whole exome analysis at the three regions 10p14-p12.1, 5q31.3-q35.3 and 7p12.3-7q21.13.

Table 4.8: Alignment statistics summary of processed reads for whole exome sequencing of HWE307

Alignment features	Individual I:1	Individual II:6	Individual III:3	Individual III:5
Total number of reads[a]	138,188,426	135,656,920	171,398,540	128,478,408
Total number of reads for alignment[b]	121,960,705	120,273,692	150,475,167	113,246,656
Total number of reads aligned[c]	112,829,531	111,429,486	140,023,138	104,245,687
% of reads aligned	92.51	92.65	93.05	92.05
Target length[d]	62,286,318	62,286,318	62,286,318	62,286,318
Target covered[e]	61,779,290	61,685,392	61,908,753	61,662,991
% of target covered	99.18	99.03	99.39	99.00
% of target covered with atleast 5X read depth	95.85	95.44	96.51	95.20
% of target covered with atleast 10X read depth	92.67	92.25	93.91	91.69
% of target covered with atleast 15X read depth	89.55	89.16	91.45	88.20
% of target covered with atleast 20X read depth	86.15	85.86	88.90	84.41
Average read depth (against covered coordinates)[f]	66.19	66.37	81.41	62.29
Average read depth (against bed coordinates)[g]	65.65	65.73	80.92	61.67

Alignment attributes of processed reads for the four exome sequenced samples is represented. [a] Number of raw reads, [b] Number of reads yielded post processing of raw reads. Reads with >70% bases with Phred score >20, [c] Reads aligned to the whole exome, [d] Whole exome sequence length, [e] Sequence length of exome covered by processed reads, [f] Average depth of reads covering all regions of the exome, [g] Average depth of reads covering co-ordinates for Illumina probes. Alignment statistics is with respect to NCBI reference human genome GRCh37/hg19.

Table 4.9: Processed read statistics summary for individual samples

Individuals	Read direction	Total no of reads	% of HQ Reads[a]	Read length			Total no of bases	Total no of HQ bases[b]	% of HQ bases	Non-ATGC character in reads		Adenine (A)/Thymine (T)/Guanine (G)/Cytosine (C) content (%)			
				Min	Max	Avg				Count	%	A	T	G	C
I:1	Forward[c]	65,357,272	100.00%	51	100	98.93	6,465,715,221	6,366,746,813	98.47	32,648	0.05	26.88	27.28	22.70	23.13
	Reverse[c]	65,357,272	100.00%	51	100	98.92	6,464,941,283	6,358,022,915	98.35	117,118	0.18	26.75	27.36	22.64	23.25
	Unpaired[c]	3,381,422	100.00%	51	100	98.49	333,029,548	308,031,562	92.49	13,463	0.40	24.70	25.08	24.05	26.14
II:6	Forward	64,425,082	100.00%	51	100	98.94	6,374,413,098	6,272,862,301	98.41	31,630	0.05	27.02	27.44	22.55	22.99
	Reverse	64,425,082	100.00%	51	100	98.94	6,374,384,124	6,275,124,487	98.44	114,816	0.18	26.89	27.49	22.51	23.11
	Unpaired	3,070,335	100.00%	51	100	98.48	302,364,958	279,923,306	92.58	12,816	0.42	24.84	25.26	23.93	25.94
III:3	Forward	81,578,157	100.00%	51	100	98.98	8,074,563,665	7,938,285,961	98.31	39,076	0.05	27.02	27.42	22.57	22.99
	Reverse	81,578,157	100.00%	51	100	98.99	8,075,444,765	7,955,585,151	98.52	145,392	0.18	26.89	27.45	22.53	23.13
	Unpaired	3,721,888	100.00%	51	100	98.52	366,665,416	340,360,992	92.83	15,873	0.43	24.89	25.19	24.10	25.81
III:5	Forward	60,651,072	100.00%	51	100	98.89	5,997,520,489	5,908,185,673	98.51	30,867	0.05	26.99	27.36	22.62	23.03
	Reverse	60,651,072	100.00%	51	100	98.87	5,996,414,030	5,895,709,176	98.32	108,416	0.18	26.80	27.49	22.51	23.19
	Unpaired	3,242,915	100.00%	51	100	98.43	319,211,869	294,989,034	92.41	12,786	0.39	24.73	25.15	23.99	26.10

Read and base features of individual samples is represented. [a] Reads with >70% bases with Phred score > 20, [b] Bases with Phred score > 20, [c] All reads aligned to the exome and considered for analysis.

Table 4.10: Variants detected at the 10p14-p12.1, 5q31.3-q35.3 and 7p12.3-7q21.13 regions from whole exome analysis

10p14-p12.1										
Nucleotide position	**Gene**	**Sequence variant**	**Het/ Hom**	**Location**	**Effect on protein**	**Variant ID**	**Frequency in control databases**			**Elimination[criteria] (Y/N)**
							dbSNP144	**ExAC**	**1000G**	
16563724	*C1QL3*	NM_001010908:c.-660_-662delCGC	Het	5'UTR	-	-	-	-	-	No
23633705	*C10orf67*	NM_153714:c.2T>C	Het	Exon 1	Start lost, p.(Met1Thr)	rs377367761	-	-	-	No
24873822	*ARHGAP21*	NM_020824:c.5396G>A	Het	Exon 26	p.Arg1799Gln	rs553473186	-	0.00003313	-	No
5q31.3-q35.3										
142814524	*NR3C1*	NM_001018077:c.-440delA	Hom	5'UTR	-	rs113797066	-	-	-	Yes[a]
147162228	*JAKMIP2*	NM_014790:c.-357_-358ins(TGA)5	Het	5'UTR	-	-	-	-	-	Yes[b]
150827303	*SLC36A1*	NM_078483:c.-76_-77insGAG	Het	5'UTR	-	rs10635760	-	-	-	Yes[a]
154202036	*FAXDC2*	NM_032385:c.674A>T	Het	Exon 7	p.His225Leu	rs199868107	0.00159744	0.0009	0.00159744	No
154242833	*CNOT8*	NM_004779:c.-6T>G	Het	5'UTR	Start gained, p.(Met1ext-2)	rs571066594	0.000599042	0.00008241	0.000599042	No
159776590	*C1QTNF2*	NM_031908:c.578C>T	Het	Exon 3	p.Thr193Ile	rs201362816	0.00239617	0.0012	00239617	No
159826788	*ZBED8*	NM_022090:c.-191_-192delTT	Het	5'UTR	-	rs141306620	-	-	-	Yes[a]
172756459	*STC2*	NM_003714:c.-1263C>A	Het	5'UTR	-	rs547309008	0.000998403	-	0.000998403	Yes[a]
172756460	*STC2*	NM_003714:c.-1264T>A	Het	5'UTR	-	rs560649548	0.000998403	-	0.000998403	Yes[a]

[a] Common in *in-house* controls (MAF>0.005), [b] False after Sanger sequencing and sequence repeat stretches.
All variants were coded according to Human Genome Variation Society (HGVS) version 15.11.
Frequency in databases wherever unavailable is denoted by '-'.

7p12.3-7q21.13										
Nucleotide position	**Gene**	**Sequence variant**	**Het/ Hom**	**Location**	**Effect on protein**	**Variant ID**	**Frequency in control databases**			**Elimination[criteria] (Y/N)**
							dbSNP144	**ExAC**	**1000G**	
50628742	*DDC*	NM_000790:c.-59G>A	Het	5'UTR	-	-	-	-	-	No
72413566	*POM121*	NM_001257190:c.2239C>G	Het	Exon 14	p.Pro747Ala	rs148845443	0	0.0006	0	Yes[a]
72413573	*POM121*	NM_001257190:c.2246T>C	Het	Exon 14	p.Met749Thr	rs373441970	0.00159744	0.0004	0.00159744	Yes[a]
73254749	*WBSCR27*	NM_152559:c.383C>T	Het	Exon 4	p.Pro128Leu	rs530654296	0.000798722	0.0019	0.000798722	No
76109876	*DTX2*	NM_001102594:c.50C>T	Het	Exon 3	p.Ala17Val	rs775816588	0	0.00002619	0	No
82784834	*PCLO*	NM_033026:c.1122_1123ins27	Het	Exon 2	Codon insertion, p.Gln374_Gln375ins9	-	-	-	-	Yes[a]

[a] Common in *in-house* controls (MAF>0.005), [b] False after Sanger sequencing and sequence repeat stretches.
All variants were coded according to Human Genome Variation Society (HGVS) version 15.11.
Frequency in databases wherever unavailable is denoted by '-'.

Table 4.11: List of nine rare, heterozygous variants common in four samples pursued for segregation analysis in HWE307

Loci	Gene	Sequence variant	Effect on protein	Segregation in HWE307	In-silico predictions				
					SIFT[a]		PolyPhen-2[b]		Mutation Taster[c]
					Score	Prediction	Score	Prediction	Prediction
10p14-p12.1	*C1QL3*	NM_001010908:c.-660_-662delCGC	-	non-segregating	0	0	0	0	0
	C10orf67	NM_153714:c.2T>C	Start lost, p.Met1Thr	co-segregating	0	damaging	0	benign	neutral
	ARHGAP21	NM_020824:c.5396G>A	p.Arg1799Gln	co-segregating	0.003	damaging	0.127	benign	neutral
5q31.3-q35.3	*FAXDC2*	NM_032385:c.674A>T	p.His225Leu	non-segregating	0.001	damaging	0.574	probably damaging	damaging
	CNOT8	NM_004779:c.-6T>G	Start gained, p.(Met1ext-2)	non-segregating	0	0	0	0	0
	C1QTNF2	NM_031908:c.578C>T	p.Thr193Ile	co-segregating	0.397	tolerated	0.01	benign	damaging
7p12.3-7q21.13	*DDC*	NM_000790:c.-59G>A	-	non-segregating	0	0	0	0	0
	WBSCR27	NM_152559:c.383C>T	p.Pro128Leu	co-segregating	0.1	tolerated	0.36	benign	damaging
	DTX2	NM_001102594:c.50C>T	p.Ala17Val	co-segregating	1	tolerated	0.007	benign	damaging

[a] Sorting Intolerant From Tolerant [SIFT] (Kumar et al 2009): SIFT predicts for amino acid substitutions with scores ranging from 0-1. Variant scores ranging from 0.0-0.05 are considered deleterious, with scores closer to 0.0 being more confidently predicted as deleterious. Scores between 0.05-1.0 are considered tolerated, with scores closer to 1.0 having more confidence for tolerated prediction.

[b] Polymorphism Phenotyping v2 [PolyPhen-2] (Adzhubei et al 2010): PolyPhen predicts for amino acid substitutions with scores ranging from 0-1 as being benign to damaging.

[c] Mutation Taster (Schwarz et al 2010): Can predict for amino acid substitutions and small insertions/deletions. Scores are provided for amino acid substitutions ranging from 0-215 but not for insertions/deletions. The scores however do not influence the prediction.

4.2.4. Whole exome sequencing and Sanger-based validation reveals five rare variants segregating in HWE307

The nine variants from the exome analysis were examined for segregation with the phenotype in the members of HWE307. I found five segregating variants: c.5396G>A in *ARHGAP21* and c.2T>C in *C10orf67* at 10p14-p12.1; c.578C>T in *C1QTNF2* at 5q31.3-q35.3; c.383C>T in *WBSCR27* and c.50C>T in *DTX2* at 7p12.3-7q21.13. Some features of these five variants are presented here.

(i) c.5396G>A in *ARHGAP21* at 10p12.1: The G to A variant in the cDNA coding position 5396bp lies in the exon26 of Rho GTPase-activating protein 21 (*ARHGAP21*, Gene ID: 57584). This change results in a non-synonymous amino acid alteration from arginine at 1799th position to glutamine in the C-terminal tail of the protein. The residue is relatively conserved across various species, except in *B.taurus, R.norvegicus* and *M.musculus* where it is replaced by another basic amino acid, lysine (Figure 4.6 i). This variant is predicted to translate into a full-length protein containing the amino acid change that might be functionally deleterious. The variant was found at a frequency of 1 in 384 among the control chromosomes examined.

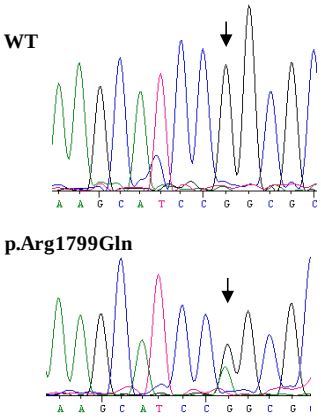
(ii) c.2T>C, *C10orf67* at 10p12.2: The T to C variant resides at the second residue of the cDNA of *C10orf67* (uncharacterized protein, Gene ID: 256815). This results in a change in the translation initiation amino acid methionine, converting it to threonine. The variant causes a loss of the translation start site that might lead to absence of the corresponding protein. Apart from *H.sapiens*, the only available homologue with a validated sequence for this gene/protein is in *P.troglodytes* (Figure 4.6 ii). This variant was absent in 384 control chromosomes.

(iii) c.578C>T, *C1QTNF2* at 5q33.3: The C to T change at the 578th bp in the cDNA lies in the third exon of the Complement C1q tumor necrosis factor-related 2 gene (*C1QTNF2*, Gene ID: 114898). This region codes for a portion of the C1q superfamily domain of the protein and the variant lies within this domain. The variant results in a non-synonymous amino acid change of threonine at 193rd amino acid to isoleucine at that position. The residue is not well conserved except in higher order mammals, *H.sapiens*, *P.troglodytes* and *M.mulatta* (Figure 4.6 iii). The variant would produce a full-length protein with the mutation that might or might not have an effect on the protein's function. The variant was found to be fairly common; present at a frequency of 3/384.

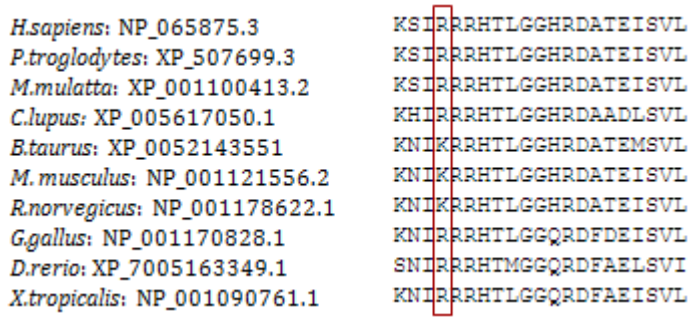
(A)

(B)

(i) *ARHGAP21*, c.5396G>A



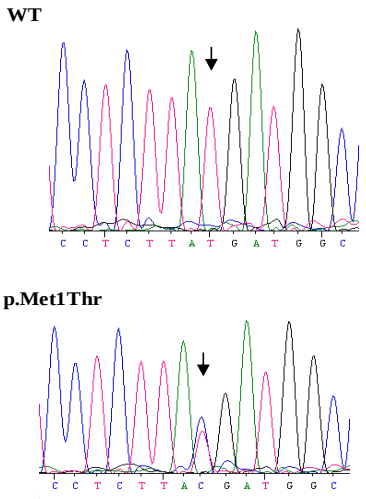
p.Arg1799Gln

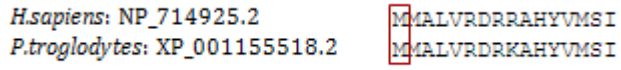


Conservation of Arg1799 residue in ARHGAP21 aross different species.

(ii) *C10orf67*, c.2T>C

p.Met1Thr

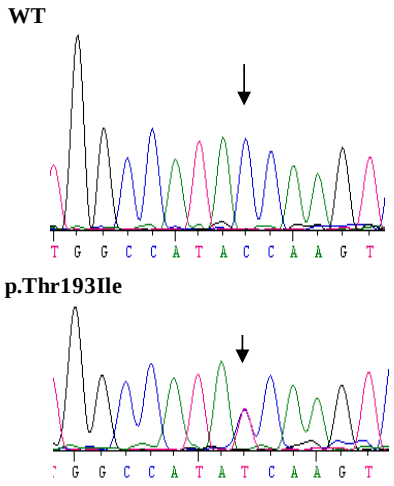


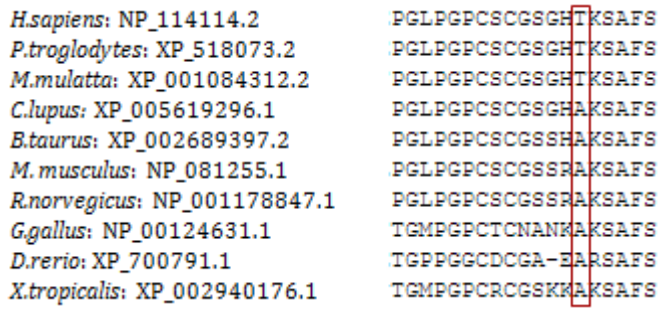


Conservation of Met1 residue in C10orf67 in humans and its only available homologue in chimpanzee, known as C10H10orf67.

(iii) *C1QTNF2*, c.578C>T

p.Thr193Ile





Conservation of Thr193 residue in C1QTNF2 across various species. The residue is conserved in 3 out of 10 species.

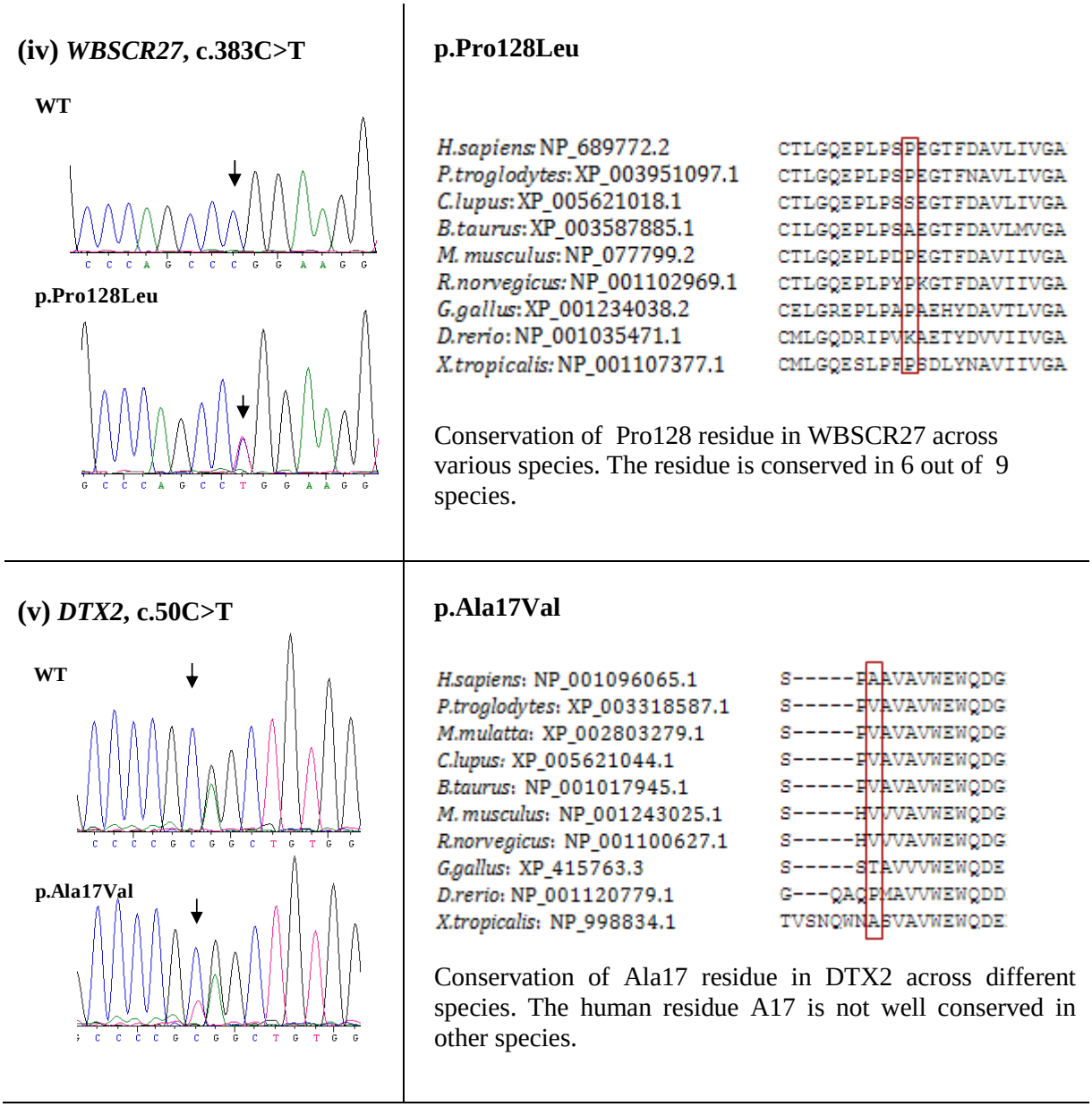


Figure 4.6: Rare, co-segregating variants in HWE307. (A) The left panel shows the representative sequences of a normal (WT) and an affected individual in the family. Arrows point to the nucleotides showing the variant. (B) The right panel denotes the conservation of the corresponding amino acids across different species.

(iv) c.383C>T, *WBSCR27* at 7q11.23: The C to T variant at the cDNA position 383bp lies in the fourth exon of the Williams-Beuren syndrome chromosomal region 27 gene (*WBSCR27*, Gene ID: 155368). This region in the gene codes for a methyltransferase domain, which has not been functionally evaluated. The change produces a non-synonymous amino acid alteration from proline at the 128th position to leucine. The residue is poorly conserved across different species (Figure 4.6 iv). The variant is predicted to

produce a mutated full-length protein which might have an altered methyltransferase activity. This variant was absent in 384 control chromosomes.

(v) c.50C>T, *DTX2* at 7q11.23: The 50th bp cDNA change from C to T lies in the third exon of the Probable E3 ubiquitin-protein ligase (DTX2, Gene ID: 113878). This variant results in a non-synonymous amino acid change from alanine at the 17th position to valine in the N-terminus of the protein. The residue lies two amino acids ahead of the ubiquitin-mediated proteolysis domain. It is not well conserved across species. Apart from *H.sapiens*, the amino acid is conserved only in *X.tropicalis*. All other species except *G.gallus* and *D.rerio* contains valine at the same position (Figure 4.6 v). Amino acid residues that are not well conserved across different species might not be important for the function of the protein. The variant c.50C>T would produce a mutated protein, the function of which might not be influenced by this alteration. This variant was present in 2 control chromosomes among 384.

4.3. Discussion

In this chapter, I present genome wide linkage mapping and whole exome analysis of a multi-generation, multi-affected family of hot water epilepsy, HWE307. Genetic linkage study suggests two loci in the family, 10p14-p12.1 and 5q31.3-q35.3. Within the chromosome 5 region, a smaller region of 9.37Mb spanning 5q33.1-q34, not shared by two unaffected individuals makes this region of 52 protein-coding genes important to be considered for further analysis. The whole genome was examined manually for the markers shared among the affected individuals that might have been missed in the LOD score analysis. This examination suggested possibility of such markers at a third location, 7p12.3-7q21.13. Among the three regions identified, the 10p14-p12.1 region is of primary interest for analysis in HWE307 on the basis of the best LOD score and haplotype segregation in all affected individuals. This is followed by the 5q31.3-q35.3 and 7p12.3-7q21.13 regions respectively. These three regions were focused at, in the whole exome sequencing dataset of four affected individuals from the family. This search identified five segregating variants: c.5396G>A in *ARHGAP21* and c.2T>C in *C10orf67* at 10p14-p12.1; c.578C>T in *C1QTNF2* at 5q31.3-q35.3; and c.383C>T in *WBSCR27* and c.50C>T in *DTX2* at 7p12.3-7q21.13 in HWE307.

A locus for hot water epilepsy, *HWE1*, has been previously reported on chromosome 10q21.3-q22.3 (Ratnapriya et al 2009b). A number of different epilepsy loci are known to map on chromosome 10. *ETL1* for familial temporal lobe epilepsy where seizures occur due to auditory sensory symptoms is located at 10q22.33, wherein a mutation in the *LGI1* (leucine rich glioma inactivated 1) gene has been identified (Winawer et al 2000). The same locus is also reported for generalized epilepsy with paroxysmal dyskinesia (Du et al 2005). A locus at the 10q22-q24 has been reported for an autosomal dominant lateral temporal lobe epilepsy (ADLTE) (Poza et al 1999), and familial temporal lobe epilepsy with aphasic seizures (Brodtkorb et al 2002); and 10q23-q25 for partial epilepsy (Mautner et al 2000). A locus for generalized tonic-clonic seizure susceptibility, *EIG4*, has been mapped to 10q25-q26 (Puranam et al 2005) and *EIG5* has been mapped to 10p11.22 (Kinirons et al 2008). However, the 18.7Mb locus at 10p14-p12.1 in HWE307 does not overlap with any of the known epilepsy loci.

The 36Mb region at 5q31.3-q35.3 harbors the protocadherin -alpha, -beta and -gamma gene clusters and the gamma aminobutyric acid (GABA) A receptor -alpha, -beta, -gamma subunit clusters. The region overlaps with *GEFSP3* locus for febrile seizures (Baulac et al 2001) and the *ECA2* locus for childhood epilepsy and absence seizures (Wallace et al 2001), both caused by mutations in *GABRG2* at 5q34. Mutation in *GABRA1* at 5q34 is reported in Dravet syndrome (Carvill et al 2014). Another locus for febrile seizures *FEB4* maps to 5q14.3 (Nakayama et al 2000) and a locus for pyridoxine-dependent epilepsy (PDE) is located at 5q23.2 (Bennett et al 2005).

The third locus at 7p12.3-7q21.13 overlaps with *EPM3* at 7q11.21, involved in progressive myoclonic epilepsy. Potassium channel tetramerization domain-containing 7 (*KCTD7*) has been identified as the candidate gene (Van Bogaert et al 2007). A locus on 7q32 is reported in photoparoxysmal response with or without myoclonic jerks (PPR3) (Pinto et al 2005). *ETL7* at 7q22.1 (Dazzo et al 2015) for autosomal-dominant lateral temporal epilepsy contains a mutation in reelin (*RELN*) gene. A locus at 7q35-q36 for cortical dysplasia focal epilepsy syndrome (CDFES) was identified with a variation in *CNTNAP2* (Strauss et al 2006). The same gene was also identified for causation of Pitt-Hopkins-like-syndrome-1 (PHLS1) (Zweier et al 2009) which is characterized by coarse facial features, short stature, seizures, hypertrichosis, short great toes and overbreathing (Orrico et al 2001).

Taken together, there are reported loci for different types of epilepsies that partially overlap with the regions identified in HWE307. Several families that transmit epilepsy in Mendelian manner have been identified with single gene mutations (Kullmann 2002). Both familial epilepsies and severe epilepsies due to *de novo* mutations can be monogenic (Helbig and Lowenstein 2013). Massive parallel sequencing is playing an important role in identification of the genetic cause in monogenic epilepsies. HWE307 transmits epilepsy in an autosomal dominant manner and suggests the involvement of a single locus and gene underlying the phenotype. Hence, it is rather improbable that all the three regions and the five genetic variants would be contributing to epilepsy in this family. However, finding multiple rare variants that co-segregate with clinical conditions in whole exome datasets is not unexpected (Cooper and Shendure 2011, Sirmaci et al 2012).

I have provided an overview of the five candidate genes identified in this family, majorly focusing on their brain expression, role in brain disorders/epilepsy if any, and the possible effect or contribution of the variant identified.

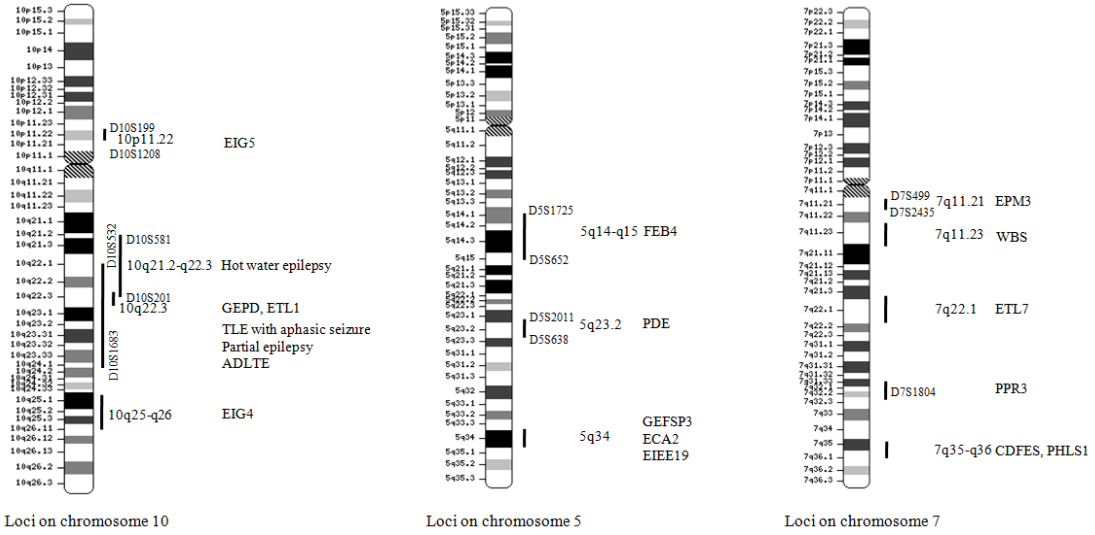


Figure 4.7: Genetic loci for epilepsy mapping to chromosomes 10, 5 and 7. The known epilepsy loci at the three chromosomal regions are shown along with the phenotypes these are involved in.

***ARHGAP21*, c.5396G>A:** *ARHGAP21* at 10p12.1 encodes a protein of 1958 amino acids. The protein comprises a pleckstrin homology (PH) domain, a RHOGAP domain, and a PDZ domain with a glycine-rich nucleotide-binding P-loop. Full length cDNA of ARHGAP21 has been amplified from brain cDNA library (Sanchez Bassères et al 2002) and has highest protein expression in brain and spinal cord. The protein localizes to nucleus, cytoplasm and

perinuclear regions depending on cell type (Bigarella et al 2009). ARHGAP21 has been elucidated in golgi complex organization (Dubois et al 2005), juvenile neuronal ceroid lipofuscinosis, JNCL (Schultz et al 2014), a lysosomal storage disorder characterized by early visual symptoms leading to blindness, progressive seizures, physical and mental decline with loss of brain volume. Although no specific function of ARHGAP21 is known in the brain, it's transcripts along with four other genes were found to undergo Nova-dependent differential splicing in the cortex and spinal cord (Ule et al 2005), suggesting some specific role they might perform in those brain regions. *ARHGAP21* NM_020824: c.T3491G; p.I1164R was identified as a prospective candidate in a screen for genes that affect brain structure and function (Karaca et al 2015). The c.5396G>A variant produces a non-synonymous change p.Arg1799Gln located in the non-domain containing, C-terminal tail of the protein. The residue is partly conserved across species, but the impact of this variant on protein function was inconclusive from bioinformatics prediction tools. This variant was found in 1 among 384 control chromosomes examined. However, considering that ARHGAP21 should have an important function in human brain, this variant has the capacity to disrupt the normal protein function. The variant should be analyzed in more control individuals and is a promising candidate for further examination.

***C10orf67,* c.2T>C:** *C10orf67* at 10p12.2 codes for a 185 amino acids long uncharacterized protein, predicted to localize in the mitochondria. The protein contains two domains of unknown function DUF4709 and DUF4724, and a calcium-binding coiled-coil domain. C10orf67 is expressed in the human brain (Xu et al 2016). The c.2T>C variant produces a change p.Met1Thr resulting in loss of the start codon. A variant of this kind can be anticipated to be important, leading to complete loss of the protein, unless otherwise. The variant was also absent among 384 control chromosomes. This makes the variant an important candidate for further study. However, the impact of this variant at this stage is difficult to envisage without much information available regarding the function of the protein.

***C1QTNF2,* c.578C>T:** *C1QTNF2* at 5q33.3 codes for a 330 amino acids long protein. C1QTNF2 is a secreted protein containing three functional domains: a collagen superfamily domain, an S-adenosylmethionine-dependent methyltransferases (SAM or AdoMet-MTase) domain and a third complement component C1q superfamily domain. No report is available regarding its expression in human brain, but the protein is expressed in mice brain mural cells (He et al 2016). The specific function of this protein has not been studied. The

c.578C>T variant results in a p.Thr193Ile change in the protein, residing within its C1q superfamily domain. C1q is a subunit of the C1 enzyme complex that activates the serum complement system. This residue is not well conserved across different species and fairly common in the ethnically matched control population. Based on the type of amino acid change, the lack of conservation and its presence among controls, it is likely that this variant does not pathogenically impact the protein function.

***WBSCR27*, c.383C>T:** *WBSCR27* at 7q11.2 encodes a 245 amino acids protein. This gene resides in the Williams-Beuren syndrome (WBS) critical region. WBS is a neurodevelopmental and multisystemic disease (Micale et al 2008) caused by the haploinsufficiency of around 28 genes within a microdeletion at 7q11.23 (Ewart et al 1993). The contribution of *WBSCR27* to this disorder is however not known. WBSCR27 belongs to the ubiE/COG5 methyltransferase family with unknown function (Merla et al 2010). It is predicted to be expressed in the endothelial cells and vascular smooth muscle cells of the central nervous system (Lee et al 2017). In a bioinformatic study, this gene along with eight others was identified for genetic modifications that could potentially be relevant in acquisition of human-specific traits (Hahn and Lee 2005). The c.383C>T variant produces a p.Pro128Leu change within the protein's methyltransferase domain. This residue is substantially conserved across species except in *C.lupus*, *B.taurus* and *D.rerio*. Bioinformatic prediction tools were unable to conclusively identify the impact of this variant on the protein function. The variant was absent among 382 control chromosomes, but needs to be screened in more control individuals. Considering the role of this molecule in a brain disorder and considerable conservation of the residue, this variant should be considered for further examination.

***DTX2*, c.50C>T:** *DTX2* at 7q11.23 encodes a 622 amino acids long protein. It comprises three distinct domains. Two WWE domains named after three of its conserved residues lie in the N terminus of the protein and predicted to mediate specific protein-protein interactions (Aravind 2001); a RING finger domain transfers ubiquitin to substrate proteins (Zheng et al 2000); and deltex_C superfamily domain at the C-terminus of the protein is a conserved regulator of Notch signaling (Matsuno et al 1998). DTX2 contains putative nuclear localization signals and is abundant in the cytoplasm and membranes of developing embryos. This protein regulates the anti-neural function of Notch probably through ubiquitin ligase activity. The c.50C>T variant corresponds to p.Ala17Val change in the

protein. This occurs two amino acids prior to the probable ubiquitin-mediated proteolysis domain. The Ala17 residue is conserved only in *H.sapiens* and *X.tropicalis*, being replaced by Val at that position in most other species. This could suggest that Ala17 does not have an important function in humans. The variant was also found twice among 384 control chromosomes. This variant seems unlikely to be having any deleterious effect. However, it needs to be examined in more control individuals to eliminate or consider it further.

In summary, I have identified five new, rare variants that co-segregate with the HWE307 phenotype. These data suggest evidence for *ARHGAP21* as the most likely epilepsy-causing gene in this family; *C10orf67* and *WBSCR27* are important candidate genes for further examination. *DTX2* and *C1QTNF2* are unlikely to be critical genes. Examination of all the variants in additional controls is critical to consider or eliminate them from further study. About 400 exons in the three regions remain uncovered in the exome sequencing experiment due to lack of probes. It would be important to sequence these regions to eliminate chances of missing out on an important and more relevant variant. Besides, the possibility of the role of non-protein coding variants in causation of the phenotype shall not be ignored.

In summary

While HWE has been reported from several different parts of the world, this disorder is predominant in the southern parts of India. The molecular mechanisms underlying this intriguing neurobehavioural phenotype are majorly unknown and genetic studies being carried out in the laboratory in the last five years or so, have begun to explore its underlying molecular causes. The sensory cortex is believed to respond to the stimuli of touch and temperature in the outcome of the phenotype. Cortical stimulation of genetically-predisposed brain regions might make it susceptible to seizures during hot water bath.

In this thesis, I have conducted genetic analyses of three multi-generation and multi-affected families with HWE to identify the potential causative genes for the disorder. An approach of whole exome sequencing analysis was used for the loci at 4q24-q28 (HWE227) and 10q21.3-q22.3 (HWE150). These studies identified *ZGRF1* as a potential gene at 4q24-q28. *FUT11* is a suggestive finding for the candidate gene at 10q21.3-q22.3. Genetic analysis in an additional set of HWE patients identified six additional rare variants for *ZGRF1*, and two rare variants for *FUT11*. Further, a combination of genome wide linkage analysis and whole exome sequencing was used to identify potential loci at 10p14-p12.1, 5q31.3-q35.3 and 7p12.3-7q21.13, respectively; and five candidate gene variants in HWE307. Taken together, the most persuasive genetic evidence was obtained for *ZGRF1* as a gene for epilepsy among the three families examined. ZGRF1 was pursued further to explore certain cell biological and functional aspects of the protein.

Interestingly, *ZGRF1* was a new gene per se at the time of its identification for its potential involvement in HWE in these studies. The biological roles of ZGRF1 were mostly unknown. Cell biological studies for ZGRF1 suggest its involvement in the nonsense-mediated mRNA decay (NMD) pathway, DNA damage and repair pathway, and cell cycle/cell division. While the wild-type and mutant ZGRF1 proteins could not be examined for NMD and DNA damage response due to technical limitation in the over-expression experiments, studies of the endogenous protein suggested its role in these two pathways. The mutant proteins presented cellular defects across different mitotic stages, and this finding is helpful to guide our thinking about the role of ZGRF1 in epilepsy. Although this study is limited to immunofluorescence experiments, findings from the same have marked the first step to explore more about the three different molecular aspects of ZGRF1. A globin NMD assay for the wild-type and mutant proteins would provide evidence for any

dysregulation of the NMD pathway for mutated proteins. The mutant proteins can be explored for affecting the DNA damage/repair pathway using comet assay. Although this study provides indirect evidence for DNA/ RNA binding ability of ZGRF1, biochemical assays to confirm the same shall be undertaken. Further, whole exome RNA sequencing studies would provide an understanding of the interacting and regulatory partners of ZGRF1.

While cell based studies were initiated for FUT11, these could not identify any difference in localization of the wild-type and mutant FUT11 proteins. This study provides evidence for a distinctive localization of FUT11 as compared to other fucosyltransferases of its kind. Biochemical and functional examination of the wild-type and mutants could not be conducted. It would be helpful to perform a minigene assay to understand contribution of the two intronic variants, if any. Following which, whole cell fucosylation can be assayed for wild-type and mutant proteins by probing them with *Aleuria aurantia* lectin. Alternatively, radiolabelled fucose administered to cells can be assayed for proteins bound to labelled fucose in the background of wild-type and mutant FUT11.

Genetic studies undertaken in the third HWE family 307 could progress only as far to identify three potential loci, with the best supporting evidence for 10p14-p12.1 as the critical genomic region. Following whole exome sequencing analysis, five candidate gene variants were identified in HWE307. This family needs to be examined further for potential variants in exons/UTRs which remain largely uncovered in the sequencing experiments conducted. Further genetic examination of the three regions and the potential genetic variants are required to identify the causative gene in this family.

In summary, I have studied the genetic aspects of hot water epilepsy in three large families. Each of these studies could be taken to different levels of exploration. Further work can be undertaken in these families to enhance genetic and molecular aspects of this curious neurobehavioural disorder. This work contributes to the growing research in sensory epilepsies, in general and hot water epilepsy, in particular; and shall help steer our thoughts towards understanding of genetic and cellular mechanisms underlying the disorder.

Appendix I for Chapter 2

Table A2.1. Primer sequences for amplifying uncovered exons at 4q24-q28

Amplicon	Forward primer 5' → 3'	Reverse primer 5' → 3'
CYP2U1-Ex1-i	aaagggtcgtaaccggactt	cctatgaccgagggatcaat
CYP2U1-Ex1-ii	gtgggcaacttcggtcac	ggaaggggcacaggaact
CCDC109B-Ex1	tgctgctgttcattccagag	ccagcaaaacgcagttgac
TET2-Ex2	tatccaacaaccagcatgt	actgctttgtgtgtgaagg
NPNT-Ex1	ctgggggttcctcgagact	acttctggatggaggagacg
NPNT-Ex3	taagtctccgggaatgtaac	gtcctttagcttagaatgga
NPNT-Ex4	ccagaagaaccagaaactattg	agagcacagagaatcagcac
CASP6-Ex1	cgactgggaggagctttaac	cgcttcaatccaagagtgc
AP1AR-EX1	cactgcctttgttccctagc	tgagtgtagcagggctctcg
ANK2-Ex1	ctccatcagtgcctcctata	ccaatgataattgcctagac
ANK2-Ex14	cagcctgagcaacaagagtg	gtgcatttttcgtcgtgtgt
SYNPO2-Ex5	tggttccaaatgtaaagctcct	ttaaggagcccaagaaatagg
TRPC3-Ex1	agcctctaactgctggatcg	aggtctgtcccctccaaatc
C4orf32-Ex1	gtctgtcggcctctctctca	gcgacgacttcccagagc
PRDM5-Ex1	ctccgggtttgaggccct	tggcgcagcgagtaaaggc
FGF2-Ex1-i	gaatgccaaagccctgcc	tgatgctcccggctgccat
FGF2-Ex1-ii	acagaagagcggccgagc	tgcaggctggaggggagaga
SEC24B-Ex1	cggagaagcttgggtacctg	ctggggtcgcattacataaa
CENPE-Ex1	cctgtttagcagtggtcacg	ggctcctggaaacatcgtag
LEF1-Ex1	ctgtaccgcccccaactcta	ccgctcaaactggattcaac

Table A2.2. Primer sequences for amplifying uncovered UTRs at 4q24-q28

Amplicon	Forward primer 5' → 3'	Reverse primer 5' → 3'
SLC9B2-5'utr-i	gcgcgtcttaacacacacaa	ccggaaagcttacccagag
SLC9B2-5'utr-ii	gttggtttcgggaacgataa	ctgcagataaacggtctcagg
INTS12-5'utr	tgggaaagaggaaaggatga	ggtccacccctttcaatacctc
GSTCD-3'utr-i	aatcccagcactttgagagg	ccaagatcgagccactgaat
GSTCD-3'utr-ii	gcctatagaaaacaacatgc	cctcaaaacctacctgccttt
NPNT-5'utr	ctcctttcctcctccactcc	ctcacctcccgtcgaactc
NPNT-3'utr	ccatgaacccccaactgtat	ccctgccaagaaatgctc
CYP2U1-5'utr	acttcggggcaaacttcag	tacgaggccgcatagcag
CYP2U1-3'utr	aaatgtagatggtccctctgg	ggatcaagtggtggcagaat
HADH-5'utr	tcatagaacaaggggccagt	atgaactgcctggtgacga
LEF1-5'utr-i	caatcaccacctccttctcg	gaggaggaggggaagagaaa
LEF1-5'utr-ii	gctgtgactccccgagact	ctgcggtagctggcgact

LEF1-5'utr-iii	atccggggtaactacagtgg	cggcggctctgtaatctc
RPL34-3'utr	cagggctatgtccaactttca	tcccttccattcatacagag
AGTX2L1-5'utr	caacgccagagccagact	gccccctctctgcacttact
SEC24B-5'utr	cggagaagcttgggtacctg	cctcaccgttctgctggt
CCDC109B-5'utr	tgctgctgttcattccagag	aaactttgcaccagcctacc
CASP6-5'utr	cgactgggaggagctttaac	ggagcaagacgcagacct
PLA2G12A-5'utr	ctcttgtgaaccaggaccaa	atggtcttcagggtggctct
LRIT3-3'utr	ggttaaaaacatcggggaac	ttctgaggttttgggtggtc
EGF-3'utr-i	tagctcagtgcagcctcaaa	tcaccaaagtacaaggggttg
EGF-3'utr-ii	aattttgggggctttgaatc	gggtagccgtgttctcatgt
PITX2-5'utr-i	tgaaaaaggaagggcaagaa	ttctgatggggctctgatct
PITX2-5'utr-iia	agggttcagaagtaaggcaca	tgaggaaagaggtcacagca
PITX2-5'utr-iib	tcactcatcctctcccatcc	attccaccaaactccactgc
PITX2-5'utr-iic	ctgtgctcgctcctggat	aagccacaaatcacctacgg
PITX2-5'utr-iii	gtctggactaataaaatcccatc	aagacaagaaaagaaggtccag
C4orf32-5'utr	cagtctgtcggcctctctct	ctcacctgctcgctctcct
AP1AR-5'utr	caggaattgaaccgtctcg	gggacggtgaaagggatt
TIFA-5'utr	tgggtacctggtgagaaagg	cccaaaatctcctggacaaa
TIFA-3'utr	gcagtactggaataggttctag	tcactacctggcaatgctaca
LARP7-5'utr	gagagtgtccacgtccctta	ctcacatttggccatcagaa
TRAM1L1-5'utr	gtggcagcgagtttgaagat	gacgatgtccgcatgattc
SEC24D-5'utr	aaagggaaaaagcgaaaagg	ggctaaagaacaaagaatgag
MYOZ2-3'utr	gtgtggaagttggtgactgtt	ggttggcagaatagggacag
USP53-5'utr	acctgccagatggtaggaaa	tccctcccttcataattgctt
USP53-3'utr	gctgcccttctgaacaaagat	cccattagaaagtgcttcttca
PRDM5-5'utr	gaaaacccagagctggacaa	gtcacccacctttcgcact
QRFPR-3'utr	gcattcagtgatggcaacat	ccgggttcacaccattctc
ANXA5-5'utr	ggagaccaactgggacgag	catggcgactactcaggtca
TMEM155-5'utr	ggatttcctcagctccttgg	agcttagggtccctctcgaa
CCNA2-5'utr	ccctgctcagtttcctttgg	tccgggttgatattctcctg
TRPC3-5'utr	ctgctcccacggtttgat	cacccttgcttgttctttgc
BBS12-5'utr	ccccttttatgcctcctaca	tcaagagcaactttccaagtg
FGF2-5'utr	aagttgagtcacggctggtt	ggttcacggatgggtgtct
FGF2-3'utr-i	gcttacctagagcaatgatc	agacacagcggttcgagaagt
SPRY1-5'utr	cccgaggtggatgttactga	accttgagcccccagaag
ANKRD50-5'utr	agccggcgggcaaagaaagg	tgggtggttgggtcaggtcca

Table A2.3. Primer sequences to amplify variants with MAF≤0.005

Primers	**Primer sequence 5' → 3'**	**Length**
SLC9B1-c.*211C>T-F	gggcctaaaatgcttacacg	20
SLC9B1-c.*211C>T-R	actggacatcatgggagttc	20
SLC9B1-c.*192C>T-F	gggcctaaaatgcttacacg	20

SLC9B1-c.*192C>T-R	actggacatcatgggagttc	20
SLC9B1-c.*185T>A-F	gggcctaaaatgcttacacg	20
SLC9B1-c.*185T>A-F	actggacatcatgggagttc	20
SLC9B1-c.1524G>A-F	tttgctaggaagaacatggaac	22
SLC9B1-c.1524G>A-R	tcctggattctctgtacagtcc	22
SLC9B1-c.1333-3C>T-F	catattcctgggcatgaag	19
SLC9B1-c.1333-3C>T-R	gctcatattttgtgactagag	21
SEC24B-c.134-10_134-9insCTTTT-F	gggaggactacaagggtgtg	20
SEC24B-c.134-10_134-9insCTTTT-R	ggcagagtacatggctggac	20
ZGRF1-c.1805C>T-F	gcagtgcagaaaatgatggt	20
ZGRF1-c.1805C>T-R	attcccatgtcaaaacccac	20
LARP7-c.552+15_552+16insA-F	tttgggaaatgtggcaatg	19
LARP7-c.552+15_552+16insA-R	cggcctttcttcttcttttc	21

Table A2.4. Primer sequences of complete gene structure of ZGRF1

Amplicon	Primer sequence 5' → 3'	Length (bp)	% GC	T_m (˚C)	Product Size (bp)
ZGRF1-5'utr-a-F	tccggatcctgatagtctcg	20	55	60.5	432
ZGRF1-5'utr-a-R	tccttcccggttctcttctt	20	50	58.4	
ZGRF1-5'utr-b+Ex1-F	ctccttgattcatgtctgtgg	21	48	59.5	287
ZGRF1-5'utr-b+Ex1-R	gctaaatctggcaacacagc	20	50	58.4	
ZGRF1-Ex2-F	tccctccctctccttctctc	20	60	62.5	477
ZGRF1-Ex2-R	ccactgtgatatggggatca	20	50	58.4	
ZGRF1-Ex3-F	tctcttaaacagactgatgc	20	40	54.3	486
ZGRF1-Ex3-R	acttaacctttctgcactgc	20	45	56.4	
ZGRF1-Ex4-F	agtcttgcgattacgggtgt	20	50	58.4	572
ZGRF1-Ex4-R	aagtgccctttgaatgacaga	21	43	57.5	
ZGRF1-Ex5i-F	ggcaaaccagaaacaatcctt	21	43	57.5	606
ZGRF1-Ex5i-R	ccttgtggttgtttttgagga	21	43	57.5	
ZGRF1-Ex5ii-F	ctgaagtccgaatcatctag	20	45	56.4	620
ZGRF1-Ex5ii-R	catgtatttacctcctgagc	20	45	56.4	
ZGRF1-Ex5iii-F	gcagtgcagaaaatgatggt	20	45	56.4	616
ZGRF1-Ex5iii-R	attcccatgtcaaaacccac	20	45	56.4	
ZGRF1-Ex5iv-F	agggtgaacatttgccattc	20	45	56.4	589
ZGRF1-Ex5iv-R	gtgcttctgtgtctttggaa	20	45	56.4	
ZGRF1-Ex5v-F	agtagtgacaacagtgtcca	20	45	56.4	580
ZGRF1-Ex5v-R	atggggacttttcattgcta	20	40	54.3	
ZGRF1-Ex6-F	agtgctacattccagattga	20	40	54.3	260
ZGRF1-Ex6-R	agtcatgaaccaagatagct	20	40	54.3	
ZGRF1-Ex7-F	gcccagccctagtatttctt	20	50	58.4	276
ZGRF1-Ex7-R	tgttagccaggatggtctct	20	50	58.4	
ZGRF1-Ex8-F	ggcaagtgtgttttcacagg	20	50	58.4	371
ZGRF1-Ex8-R	gggggcaggagtcattaaa	19	53	57.5	
ZGRF1-Ex9-F	ataggagaatttcccggttt	20	40	54.3	470
ZGRF1-Ex9-R	gagatattatgcttcttgtc	20	35	52.3	

ZGRF1-Ex10-F	cccccaccatagctctacat	20	55	60.5	
ZGRF1-Ex10-R	gcttcaagggatatgacactga	22	45	60.1	
ZGRF1-Ex11a-F	gggacttcatcaaccctaagaa	22	45	60.1	616
ZGRF1-Ex11a-R	ttgcattgggatgtgttttg	20	40	54.3	
ZGRF1-Ex11b-F	gcctgaggaccaaaatgaaa	20	45	56.4	544
ZGRF1-Ex11b-R	gccatctgtccaaattgctc	20	50	58.4	
ZGRF1-Ex12-F	gggttaaaattgccattcg	19	42	53	526
ZGRF1-Ex12-R	gaagattgtaatttcccact	20	35	52.3	
ZGRF1-Ex13-F	gactctgtttaaaggctattc	21	37	55.4	389
ZGRF1-Ex13-R	tcaagactcactctttaagc	20	45	54.3	
ZGRF1-Ex14-F	ggagtacagtaatgccaagc	20	50	58.4	500
ZGRF1-Ex14-R	cagagcaagagcttgtaact	20	45	56.4	
ZGRF1-Ex15-F	gatgacagttctgtgacagt	20	45	56.4	322
ZGRF1-Ex15-R	ggagataataaggctagactg	21	43	57.5	
ZGRF1-Ex16-F	cagagcactcccaaccttta	20	50	58.4	396
ZGRF1-Ex16-R	ggaactcactgtacacaaca	20	45	56.4	
ZGRF1-Ex17-F	gtttgggatctctgtgtgtg	20	50	58.4	455
ZGRF1-Ex17-R	ttacaggcatgagccactgc	20	55	60.5	
ZGRF1-Ex18-F	cagttctcttctaccctcct	20	50	58.4	627
ZGRF1-Ex18-R	gggtagtcagttctgttctc	20	50	58.4	
ZGRF1-Ex19-F	gtggaggttacagtgagctg	20	55	60.5	403
ZGRF1-Ex19-R	ccagcaagagtacatggaca	20	50	58.4	
ZGRF1-Ex20-F	gccttcaagatggaataatgc	21	43	57.5	474
ZGRF1-Ex20-R	catggatgaagctggaaacc	20	50	58.4	
ZGRF1-Ex21-F	tgacgagttaatgggtgcag	20	50	58.4	368
ZGRF1-Ex21-R	acatgcatgcaccttagtag	20	45	56.4	
ZGRF1-Ex22-F	ttcttgttaatgggcttgtgg	21	43	57.5	457
ZGRF1-Ex22-R	aagctggtcgcaaactccta	20	50	58.4	
ZGRF1-Ex23-F	tactcaggatacttcatggc	20	45	56.4	416
ZGRF1-Ex23-R	attatagggttccacttgtc	20	40	54.3	
ZGRF1-Ex24+25-F	catggttgtaattcatctcc	20	40	54.3	679
ZGRF1-Ex24+25-R	tcaggcaccaaaccaactac	20	50	58.4	
ZGRF1-Ex26-F	acagtggttgctttgagcag	20	50	58.4	450
ZGRF1-Ex26-R	agttacttcccccagcagag	20	55	60.5	
ZGRF1-Ex27+3'utr-F	gctgggggaagtaactgaca	20	55	60.5	541
ZGRF1-Ex27+3'utr-R	gttccaacagatgagttctgg	21	48	59.5	

Table A2.5. *ZGRF1*-cDNA primers to amplify and sequence confirm full-length *ZGRF1* cDNA

Amplicon	**Primer sequence 5' → 3'**	**Length (bp)**	**% GC**	**T_m (˚C)**	**Product size (bp)**
ZGRF1-1F	ggaaagccaagaatttattg	20	35	52.3	505
ZGRF1-1R	tacatctttcttgccaacag	20	40	54.3	
ZGRF1-2F	atatcctctggccgatctct	20	50	58.4	595
ZGRF1-2R	ctcagcacactcttcctgttg	21	52	61.5	

ZGRF1-3F	gctgagatgaagagcacaga	20	50	58.4	551
ZGRF1-3R	tggagatgtttcagttctggc	21	48	59.5	
ZGRF1-4F	gctcaggaggtaaatacatg	20	45	56.4	586
ZGRF1-4R	cagcatcagtccacttgaaag	21	48	59.5	
ZGRF1-5F	tgtgggttttgacatgggaa	20	45	56.4	561
ZGRF1-5R	ctggcagtttcaacctcaaca	21	48	59.5	
ZGRF1-6F	tccaaagacacagaagcaca	20	45	56.4	656
ZGRF1-6R	acctgcaagaagtcaatctgc	21	48	59.5	
ZGRF1-7F	tcagaggacacagctcacag	20	55	60.5	684
ZGRF1-7R	aatgcatcccagaaacagcc	20	50	58.4	
ZGRF1-8F	tccaggggatgtttcacttaa	21	43	57.5	672
ZGRF1-8R	gcaggttgactatgatggca	20	50	58.4	
ZGRF1-9F	gcagactttcacatcttgcct	21	48	59.5	591
ZGRF1-9R	tccagctcaaagtctagggt	20	50	58.4	
ZGRF1-10F	cagagaaaacagtatggcaa	20	40	54.3	692
ZGRF1-10R	aaatcaccactgccagcaag	20	50	58.4	
ZGRF1-11F	tcagcctaggagcaacattga	21	48	59.5	572
ZGRF1-11R	tccaactactcgaacctgct	20	50	58.4	
ZGRF1-12F	gaagacctgactcctacgga	20	55	60.5	683
ZGRF1-12R	tgaaagtccacagcactgag	20	50	58.4	
ZGRF1-13F	gcaagtggaatagcaggctc	20	55	60.5	423
ZGRF1-13R	cttctgtttttcttccacttg	21	38	55.4	

Table A2.6. Site-directed mutagenesis primer sets used in generating point mutations in *ZGRF1*

Primers	**Primer sequence 5' → 3'**	**Length**
ZGRF1-Leu9X-sense	gaatttattgttctata**g**actcatcaaaagatgaagaa	38
ZGRF1-Leu9X-antisense	ttcttcatcttttgatgagt**c**tatagaacaataaattc	38
ZGRF1-Arg326Gln-sense	accatgagaaataaaagcc**a**gtgggccatgtatttatcc	39
ZGRF1-Arg326Gln-antisense	ggataaatacatggcccac**t**ggcttttatttctcatggt	39
ZGRF1-Thr602Ile-sense	gttagtgacaaacctacagtga**t**atttcctgttaaagagactctg	45
ZGRF1-Thr602Ile-antisense	cagagtctctttaacaggaaat**a**tcactgtaggtttgtcactaac	45
ZGRF1-Glu660Gly- sense	gatgctgtatacggagataataaag**g**agatgctaataaacctattcaa	48
ZGRF1-Glu660Gly-antisense	ttgaataggtttattagcatct**c**ctttattatctccgtatacagcatc	48
ZGRF1-Arg1862X- sense	gaaaatggattggaacaaactctttttgat**t**gactttgcttaatggg	47
ZGRF1-Arg1862X- antisense	cccattaagcaaagtc**a**atcaaaaagagtttgttccaatccattttc	47
ZGRF1-Phe1940Leu- sense	taatgtggcagaagctacg**c**ttacactcaagctgattc	38
ZGRF1-Phe1940Leu- antisense	gaatcagcttgagtgtaa**g**cgtagcttctgccacatta	38
ZGRF1-Asp1984Gly- sense	gtggactttcaccatcctg**g**tattaaaactgtgcaggtg	39
ZGRF1-Asp1984Gly- antisense	cacctgcacagttttaata**c**caggatggtgaaagtccac	39

Appendix II for Chapter 3

Table A3.1. Primer sequences for amplifying uncovered exons at 10q21.3-q22.3

Amplicon	Forward primer 5' → 3'	Reverse primer 5' → 3'
SIRT1-Ex1i	ggagcggtagacgcaacag	ggcccattgtctccttcc
SIRT1-Ex1ii	agagatggtcccggcctc	acctcaggccagtaggagc
PBLD-Ex9	cttgaaactgatctgggagg	agactactacgcacatttgc
RUFY2-Ex1	ccgggttctgaggcatagt	aggagtaccagggccagtg
CCAR1-Ex10	ctatagctcattagtaactg	gaaaacaggtattcgggtct
CCAR1-Ex17	cagagcgagactccatctca	cgaggtttcactgtgttagctg
STOX1-Ex1	cgctgggctctggattctcc	cgacggaaagcgcggaaca
KIAA1279-Ex1	ccgactgcaaacattgagga	acgaattgcgccttacaacg
VPS26A-Ex1	gcccactcgacgtaatttg	gccagcttttcccagacc
SUPV3L1-Ex1	gcgttttcccgggcagctta	actcgatgggacctcggatc
HKDC1-Ex10	cctccctaccatcagcaaat	cttgccctccattagagacg
HKDC1-Ex17	tctcggcttctaaggcattc	ggatgcatcttcccatgc
HK1-Iso3-Ex1	gcatcaggactaggctggag	cgctcatattcccttccttg
HK1-Iso4-Ex1	aaccaatgggcgtggagga	ttgcgccctgcagtccaact
TACR2-Ex5	atcaaactcacccacgaagg	ccattcccacaagagtgatg
TACR2-Ex3	aggctacccaatgccacctt	gcagcttccgcaagcttct
TSPAN15-Ex1	agtgtcagtcccggagagaa	gcccgaatgtgaggtagaaa
COL13A1-Ex1	cgttttccagcgatacaagc	gtcctccgaatccctggtag
COL13A1-Ex3	tcgaagaccagatggaaacc	ccacaagaggcattctggag
TYSND1-Ex1i	ctctcgcttccagctgtgg	gagaagcgccactgttcc
TYSND1-Ex1ii	cttcctgctgagctgctg	cgtgaagcccacccattc
TYSND1-Ex1iii	ggcccactgctgcttaccga	ctacccactacacgcccaaa
PPA1-Ex1	ggtgggaacactagcagagc	cggcaagtatgcaatgtgag
NPFFR1-Ex1	agcccgaggggagggaaac	tctgtctcacatacacacgc
NPFFR1-Ex4i	gactactcagccaattcagg	gaaggcgtagacggtgacca
NPFFR1-Ex4ii	ttcttcacgctgtcctggct	gaccacgcatcctcacctaa
EIF4EBP2-Ex1	ctgttgctcctgaggctgct	gaagtaaagggtcccgaacg
PALD1-Ex7	ccctctgagtcgctctcact	aatttccagctctggctctg
PALD1-Ex11	ctctcagccactcccttgtc	tgcttaccccaaccctacag
PALD1-Ex12+13	actgtgaagctcgggttcc	gggcctcagattctccatct
PALD1-Ex14	ctgctaacctgcttggcttt	aatcggaaaacctcgtgaga
PALD1-Ex17	tcctccctcagtctgagaac	catggatcacttccacttcc
PALD1-Ex19	ccttgagacagtcctgtgagg	caactccgctcagtctctcc
PRF1-Ex1i	gtacaatgtgggggctgaac	agttggtgagcgccagag
PRF1-Ex1ii	gctccttcccagtggaca	cagcctccaagtttgattgg
ADAMTS14-Ex1	ggagggaagcagctaggc	tgtgcatactgcctcccacc
ADAMTS14-Ex18	cacaaggctctccacaggta	tgtgttgcctgtatggtcgt
ADAMTS14-Ex19	tgaaacccaaacccactctc	agccattcaggcagaaactc

ADAMTS14-Ex22	ctctgctgtgtgtcctgcat	gccctgctatgtgtgtcctc
PCBD1-Ex1	gaggaaggggttaccaaagg	gcaggggactcgaaaagact
UNC5B-Ex1	agagacccggagccagag	gagcagggacctttcagctt
UNC5B-Ex3	gggtggagacaggagacaag	aggccccacagatgtgaagt
UNC5B-Ex6	ctaggaaggccaggcttcag	ctcgggacaagcaccttct
UNC5B-Ex7	catgtgttctgggttcctca	gagaaggcaatgggaaggat
UNC5B-Ex10	taccgagaaggaagggagtg	ccctttccccatctgtgag
UNC5B-Ex13	cccgtgactggatctttcag	ggctgtttcagtgtgtgagg
C10orf105-Ex1	gctgatctgcctttgagctt	tgtgtgtcccagactgcttg
C10orf54-Ex1	ctcactcgctcgcactca	acggtcacacaggctgaga
PSAP-Ex1	ctcctattggcctcccttc	cccagagaaagccagagaaa
CHST3-Ex2i	ctagcagatgccaacagcac	ggcggaacatgaactgagtc
CHST3-Ex2ii	gtgtaccgcgacgtgctc	cggatgctctcgcagttg
CHST3-Ex2iii	ttcgccggcaagtataagac	actgtcctgttctccgcaat
SPOCK2-Ex1	gctgggctttttcagacaag	ctgtgacccccagtactgttt
SPOCK2-Ex6	aagaggacccttcaggctgt	gacctcacagctccctcaag
ASCC1-Ex3	atcccacgcaaaaactaacg	acagcgagctacagccaact
ASCC1-Ex10	tatttcgtggggttgtgagg	ttaaaagattggccggatgt
ASCC1-Ex11	gacaagtgacaaactagagg	ggtccaatgggtgtatatga
DDIT4-Ex1	tggtgagtgtcccttctgtg	gatgggataggaagccaagg
DDIT4-Ex2	agacacggcttacctggatg	gagttggcggagctaaacag
DNAJB12-Ex1	gcgtgcaccattattgacac	aaactacgcctcccagcag
MCU-Ex1	catcaactcagtcaagggct	agtcggggagaggttcaagt
ECD-Ex9	agataatggaaggcttcag	atcctggctaacacagtgaa
FAM149B1-ExF	gctgctcggagtctaggtga	cctgggaaaggaaaggaacg
DNAJC9-Ex1	gactcgccacacctcatttt	ggcgcgagaaataaagatca
MMS51-Ex1	cctgggtctgagagactgga	gtctcccactgccatcattt
PPP3CB-Ex3	tgcccaagtccctagagcct	ctgaaagcaaccgggagac
PPP3CB-Ex14	agactgaaatgaggatagga	cttatcagatagcacatgtc
FUT11-Ex1i	actcccgaaagctcactttg	tctgtgccgtagaagagcag
FUT11-Ex1ii	ctattcccccacttcccggga	gagtcctggacacctggagt
FUT11-Ex2	ggcatcaccaaccaatttct	tgattttcatacgttgccttg
CHCHD1-Ex1	cccggccaagtaaaagaaag	gaggaagccctggatctctt
ZSWIM8-Ex10i	ggggctgggtcaagaagta	cagctggctcagtaggaagg
ZSWIM8-Ex10ii	ctgacaggaagctggcact	gtatctgggggctcaggaag
VCL-Ex1	tcgaaaagggaccagtagga	ctcaggactcgctcaaggtc
ADK-Ex1	ccttccctccaatcagcac	ccacagacccgaaaacttct
KAT6B-Ex1	aaggacttgaggagccgaat	attcttggccaaccagtgag
KAT6B-Ex13	tggttggattccagtgttctg	ccatgcaacctcagttagga
DUPD1-Ex2	aacagtgcaggtggcctagg	cggggaaaccttgggctaag
DUPD1-Ex3	tgccttctcaacaaccctat	aagatggtttggaagagtcg
DUSP13-Ex3	tgctcaggatcacagagtgc	atcctcacttccagccacac
VDAC2-Ex1	ggattctgcctttggaggat	aggccctttaccctgatacg

COMTD1-Ex1+2	gtagtgccccgacaaggtg	ggactggcccttcggaac
COMTD1-Ex3+4	acgtggctgtgtgaccttgg	gtcgatcttgtgctccgcct
COMTD1-Ex5+6	ctgtggaggcaggtgagcg	gtgactggactgggtggggct
ZNF503-Ex1	tcgctttctgccctaagaag	attccacttcctcccccttt
ZNF503-Ex2i	ggctgggaggatcttcgt	ctgtcctgggcgtgaatg
ZNF503-Ex2ii	tgtttcgtcggagaagtcg	gcttactgcagcccagagac
ZNF503-Ex2iii	gcacttgaccccaccaag	gtatgggtccgcaagtgg
ZNF503-Ex2iv	cacatctgcaactgggtgtc	cgctgctgtcgggtagatag
DLG5-Ex1	tagggctgctgccgaggtgct	agcgagtccaccgaggccat
RPS24-Ex5	tggaaggcacctggacaat	ctgagtggagaagaaggtgg
ZMIZ1-Ex6	tcctgcctccactttgttct	tgactgagccccttcaactc
ZMIZ1-Ex7	aggttctgggtacagctcca	acagacctccacatgcttcc
ZMIZ1-Ex8	agtcactgccgcacagga	ccacacccctccaaagc
ZMIZ1-Ex9	caggagcaaatgaggagagg	gaacttcgccctgaacacat
ZMIZ1-Ex12	ccttcctctgagcatgcag	acagggcccagtcaatgat
ZMIZ1-Ex20	taccgggtggccacgtttcg	cttccccatcatggggcttc
PPIF-Ex1	acctgggcaagccaataaag	aatcattctgacgcccattc
ZCCH24-Ex4	tcatcgggtgtgtgtctctc	caagacaaggacgctaagagc
ZCCH24-Ex1	caaaagtttcctgcccaact	aggaagagcgggtcagacag
TMEM254-Ex1	acagggagactcgctctcag	ctctcattgcggagtctggt
TMEM254-Ex4	ttttgcatctgcttgggtct	cttggaagagcaggctcact
TMEM254-Ex5	attcccttccattgggtcat	tggctactgtctagggcaca
PLAC9-5'+Ex1	gtgatgccgaggaagacact	tccagctctctctccgtctc
ANXA11-Ex3	tgctgtgaggactgacttgg	ctttccctgtctccccatct
DYDC2-5'+Ex1	gacccagaaggagttaactg	gcagatttcactgcaggttg
SH2D4B-Ex1	agctgaggccaactgacaat	cccacaagacaagggtttca
SH2D4B-Ex5	ggtgaatcctgaggttcctg	tagccccaaactacacagca
CDHR1-5'+Ex1	cgctaattggtctcgtcagg	gccaggaagatggaaggact
CDHR1-Ex17i	ctgtgtgcattttcctggac	gagagctttctgggctgttg
CDHR1-Ex17ii	accatccgcattgagtgg	ggcggtagagatgaggacag
CDHR1-Ex17-Iso2	gtgccaggtattgttctgag	atgagagaagggctgtggaa
LRIT1-Ex4	tgttgataccttggctggt	agggtcagctcctcctttgt
LRIT1-Ex2	gtcatggcttggactctggt	gataccggtctccaggtgag
WAPAL-Ex18	tcattgcatgttatggcaaa	tgcatgacgaagaaatcagg
OPN4-Ex8	atggttcccagggtctgag	tacgaacttcatgggggtgt
OPN4-Ex3-Iso2	gcatcctcaagtccctcaaa	ggctcctgtcatcatccatc
LDB3-Ex4	agtgtcctcccctccaagtg	cggaggctcatctggtacat
LDB3-Ex7	aggcaggtggagctttctg	gcacacccctgtaattcacc
LDB3-Ex9	catttctctggctaggagtg	ttgttgcagtgaccgcagag
MMRN2-Ex6i	tggaggagaacaaggaggag	ggatctgctcgtagctcagg
MMRN2-Ex6ii	gccaagtgcaggcgctggat	tggaagagccgctggtgctg
ADIRF-Ex1	agcctgtatgacgggaacc	ctgcgtctcagtggaggac
FAM25A-Ex1	tcacctgtccagcctcctac	aggtctgcacccctcaact

Table A3.2. Primer sequences for amplifying uncovered UTRs at 10q21.3-q22.3

Amplicon	Forward primer 5' → 3'	Reverse primer 5' → 3'
LRRTM-3'utr-a	gacttaaacagagtatgacc	gataaaattctacgcagtaagc
LRRTM-3'utr-b	tccttctaacccccattcct	tctggaagttctcttgctctg
SIRT1-5'utr-b	ggaaggagacaatgggcc	acagccaaaaagttgcagac
HERC4-5'utr-a	ccctcttaaagcaacgcaaa	accctcctctgtcagccagt
MYPN-5'utr-a	tgcaggctaagaacatcgtg	atgcttgcctgggaaaacta
MYPN-5'utr-d	gcattcctcacacccagatt	agttgcagtgagccgagatt
MYPN-5'utr-f	gtgcccaaccagaactcaat	aggcttgtgcaccagaaaag
MYPN-5'utr-g	acaagcgttattactggtga	gccaaatcctcatgacactt
MYPN-5'utr-h	aataccttttcagaagccag	gagttgaaaatggtctgttc
MYPN-3'utr	gggttttggaaggagaggtc	aattcatttccctgcacagc
ATOH7-3'utr	tctccactgtgagcacttcg	tcacagcaatcaacccattc
PBLD-3'utr-b	gtccagacggaagggttgac	atacactcatgagtggagcc
RUFY2-5'utr	catgtggtctccttcacgtt	tcggccactatgcctcagaa
TET1-5'utr	cagtttgggtaaatccagct	cattgtttatctccggcaac
STOX1-5'utr-b	tgttccgcgctttccgtcg	agcctctgtggcagaagacg
DDX50-5'utr	cacactcattggttgagagg	tcatcagggacagacagttc
DDX21-5'utr-b	agacactgcgaaagcaaacc	tcgagacgggatttcactatg
DDX21-3'utr-b	caagccatgggtattaggtga	tgaacatgtggaagactgctc
TACR2-3'utr	cccaccaaaactcatgttga	tcaaagggtgctacacctca
TACR2-5'utr	cggaaggatggaatctgaaa	agatgacgatggcattaccc
C10orf35-5'utr	tccctcctccttcttctcct	gcagtccctttgtgctttgt
C10orf35-3'utr	ggcctgcattaagtggcaca	tctgaggtcaaatgctcgct
H2AFY2-5'utr	aaggtggatccagctaacag	tttttgcagagggtgggtcg
AIFM2-5'utr-b	tgaatggctcggcaggact	gagtatccctctcctcgca
TYSND1-3'utr-bi	acgcaggaagagtgcttagg	agttcccacgacctcctttt
SAR1A-5'utr-a	gcgctttgtcacgcctta	accgcctgcgtgtcactt
SAR1A-5'utr-b	catagagacatgaagggagc	caatctggaggagagtctag
LRRC20-5'utr-a	gcacacaaagcggttactcc	acacttctcgccctctctcc
EIF4EBP2-5'utr	cttcaccaggctcacaaaca	ggcgtggtgcaatagtcat
EIF4EBP2-3'utr	ccaagccaagagttttgtcc	gttggggtgtggtgctttta
NODAL-3'utr	acttgttgtgccccagtgaa	ccttgccactgtcttttcag
NODAL-5'utr+Ex1	aactaggtcctccgccagct	agcacttcccgagtccgct
PALD1-5'utr-a	gctgggctgagtgtgtcttt	cctcacctctccaaaggatg
ADAMTS14-3'utr	cctccaagggacaggtaaca	cccccagttctcttgctaca
SGPL1-5'utr	acattcactggagggtctgg	gtccacttgcttggacctca
SGPL1-3'utr	tagctctgacctgtcctgat	gttgcaatagaagtgactct
UNC5B-3'utr-a	caaaagcaaaaaggcaaagg	ttggctttatggattcagca
UNC5B-3'utr-b	agttgtgcagggagtaggtc	cccacgttgcagagctgact
UNC5B-3'utr-c	gcagagtgatggctgagaag	actgagtccatggcagggaa
UNC5B-3'utr-d	cctggcctttgaaagtctga	aggacactgaagcccagaga

UNC5B-3'utr-e	gcctggcagtccaaagacca	acacacacgctcgctcaggt
UNC5B-3'utr-f	ttctctgtgggcacatcagg	gagaagagcaacggagtcat
CDH23-5'utr	cgggggaagcaagggaaagt	aacaagtggttcggcagccg
CDH23-Ex11	tcctctagggtgtgatcctg	cttcccctgaacctttggt
CDH23-3'utr-a	aggtggtggacaaggatgag	tagagacggggtttccagtg
CDH23-3'utr-b-i	gggacctctgtgctgacagt	cccggcctgaaatcacttat
CDH23-3'utr-b-ii	ccctcccagtgtcatttgta	ccgatgagaatctgctgga
CDH23-3'utr-c	gctacttcatcacaggtgct	catcactaatccattgcagc
CDH23-3'utr-d	atgtccttgagatggccaag	attcccagctggtgtcaaga
C10orf105-5'utr-a	tggttccagtgggtagaagg	gaagcagagaatggggtctg
C10orf105-5'utr-b	agagggctccaaggagacag	ggccttgtttcccctctcta
C10orf105-3'utr-i	ttagataacaccatgtgcca	tgtattctctcctggttctc
C10orf105-3'utr-ii	ggacggtcaaggactgattc	aactggagggacagagtctcc
C10orf54-5'+Ex1	aaactgccgatcaagtggag	acggtcacacaggctgaga
CHST3-5'utr-a	cttcgccaacggttcccaag	ttctccctccaaagtctggc
SPOCK2-3'utr-a-i	cctgaccccagaaaacctct	agatcaaacagccagcgagt
SPOCK2-3'utr-a-ii	agatgagccgcagtctcct	ctgcagccaggactacctta
SPOCK2-Ex11	gtgactgctggtgtgtggac	atgccatgcacactcacact
SPOCK2 5'utr-b	ctggctggctttgtaaccgg	taatgtccttcctcccaccc
ASCC1-3'utr-a	cttctactttggctccctaa	gagaatagtcaaaaccagag
ASCC1-5'utr-a	ccatcatcaacaaatcgttcc	ggctgtaggggaaagtgcta
ANAPC16-5'utr-a	ggccctgaatcatcaaagag	tttctcctcctaccttctcca
ANAPC16-5'utr-b	aagcaaacagaacagccagt	caccatgttgcccctagact
ANAPC16-3'utr-i	ggaatctaaaaaccatctctcc	aacagcacatctcttagcct
ANAPC16-3'utr-iia	acattctacctgacgggagg	aaagtgccgggattacaggc
ANAPC16-3'utr-iib	tgagctgagatcacaccattg	tgctgtcgaatatggaacttct
DDIT4-5'utr-a	caggagagaacgttgcttacg	agccgctgtaagacaagagg
DNAJB12-3'utr-i	accaccgaggtccaagtatg	ccacctattcccacatggtt
DNAJB12-3'utr-ii	tattgtccgtgccctggtat	gcttctgcccacacttctct
DNAJB12-3'utr-iii	gtagcaggctccctccttct	ccatggcaggttttctgtct
MCU-5'utr-b	gatgttcccgcaggactggg	acactcacccaagagtgttc
P4HA1-5'utr-a	ttacatcccactcccgctta	gaggctgctgtggtaacctg
P4HA1-5'utr-b	gataggttgttttctttctc	cagaatagttatcagatctc
NUDT13-3'utr	acagttgctggcatttgcac	tcaccaatagggatcctttcc
FAM149B1-5'utr	ctgtggcacctttccttagc	gccttccgagtgtatctgga
USP54-3'utr	gaagagagcagatcagggcta	acccttaggagaacccacca
MYOZ1-5'utr-a	actttaagccctccctgctc	ctcccccaggatcaagaatc
SYNPO2L-3'utr	gctcttccgtctaaggcatga	tccatgtggggatatctgac
ZSWIM8-5'utr	taacgttgggagttggcttc	cctcggagatgaaggaacag
NDST2-5'utr-a	gctgagtggatgagcccag	cgggtggtgatggtggcaac
NDST2-5'utr-b	ctcagagatctcccgcacac	gggaaggggtgaatgctaag
C10orf55-5'utr-a	attggcaaacccacagagag	ctagacgggaggattgcttg
C10orf55-3'utr-a	gaccaggacgcagagaagca	tcgctgagcgctgcaagaca

C10orf55-3'utr-b	tctccgcactgtgctgcga	acccccttactgccgaaagc
C10orf55-3'utr-c	gctccctctgtgcgtctgtt	gtcctccagtctcccaattc
PLAU-5'utr	acaagcctctcgattcctca	cccagctgtctctctcctct
AP3M1-5'utr-b	ggactgggcatccttgttag	cttcctcagccaggctcagt
KAT6B-5'utr	gcgcggtatctaggagagag	gtcgcaagcccctgaatc
SAMD8-5'utr-a	ctcgcgtcttttccagtgtg	gttcgacaacctccatttgc
VDAC2-5'utr-b	gagatctacgtgaccctggtg	ggacaagaagtgcggtaagc
ZNF503-5'utr	cttgaccctccaatcacctc	ccggagctatttccagagag
RPS24-3'utr-d-ia	tgaaatccaccaggaatcca	gaagagcttctaaacttggc
RPS24-3'utr-d-ib	gactgactgacggaatgacg	gtggctcatgcctgtaatcc
RPS24-3'utr-d-ii	ggtgcaagcagagttgtgag	agcagctgggactacaggtg
ZMIZ1-5'utr-a	gacacttcaacttgctcggc	acccaatccccacgcttgca
PPIF-3'utr	tgtcctccctgccttagaga	cgaggtgggccttattgtta
SFTPA2-3'utr	ggctccactttccccttta	ccctcccatcaaacaaagaa
SFTPA1-3'utr	cactggctctgctttctcct	gtggctcctcctaccaaaca
TMEM254-5'utr-c	ccctctctggtcacagtcct	gactttccgcttcgtgagag
ANXA11-5'utr-a	ccgagacttcctggagcag	acgcctgtagtgcaaccag
ANXA11-5'utr-b	ggtgtcaaacttctgagctc	ctctaccctcacgaatggat
ANXA11-5'utr-c	acagagctggctaaatccag	agccaagattatgccactgc
DYDC1-5'utr-a	tctcctctctgctgcgcgtt	aacggtcagaaccgcctct
DYDC1-5'utr-b	taatcccatgcatgacggct	aatcctagcactttgggagg
DYDC2-5'utr-b	acccagcgaagctaggggaa	ctctactctgccgcctgct
DYDC2-5'utr-c	gctagtaaatgccagattgg	tcccaaagtgctaggattac
FAM213A-5'utr-a	tttaaagacaggctggagtgc	tcactccacaatgccagcta
FAM213A-5'utr-b	gttcgcggacttgaggagt	gcggctcctaagaaagttgg
FAM213A-3'utr	ctgtttaggcccactaaggca	cccagagcatgtcaattcta
TSPAN14-5'utr-a	ttcctcggcagacggcag	gagaacgctcggggaagtc
CDHR1-3'utr-a	gaccattttccctctccctaa	ccccttcactcagctctcat
LRIT1-3'utr	ctgcaacaaccgctaacact	gtggcagtataccagttctg
FAM190B-5'utr-a	aggcggagcttgggaaag	gactgcgtgcctggtctt
FAM190B-3'utr	gactatcttgagaaagctgg	acctaccttctacagcgtat
GRID1-5'utr+Ex1	agacacacagccagcccaga	gcccagggaaacgccaagtt
GRID1-Ex15	aagctgctaggatggctcaa	aaacactgctgcctccagac
GRID1-3'utr	gagaccagcatcttcaagag	atgactgtccctgctgatagc
WAPL-3'utr	cattggatttagggaaaggg	tgctccaaagatcacagctg
WAPL-5'utr-a	gcggacctggactacaactc	gaagtctcactgcctcgcc
LDB3-5'utr-a	ccagagcattctcaccaacc	ctgcagacaaagggtagaga
LDB3-3'utr-b-i	cgttcccagaattcagtttc	aggctggtttcaaactcctg
LDB3-3'utr-b-ii	gggagcacactttggagcta	tgtgcacacctttaacaggcta
BMPR1A-5'utr-a	tgcgctcggctcttttcgtgt	acccgcgagtgggagacaaa
SNCG-5'utr	atgaatgagggacaggttgg	catacatgacccccctccttg

Table A3.3. Primer sequences to amplify variants with MAF≤0.005

Primers	Primer sequence 5' → 3'	Length
HNRNPH3-c.112+4C>T-F	catttcctcttgtggcacct	20
HNRNPH3-c.112+4C>T-R	ttcacggttcatttcatcaca	21
CCAR1-c.518+102T>C-F	ctgttgcactgcctacaagc	20
CCAR1-c.518+102T>C-R	gcctgtcatcccagctactc	20
TYSND1-c.*1407T>A-F	gccctcatagctggtcatgt	20
TYSND1-c.*1407T>A-R	acgagtcaggaatgctggat	20
USP54-c.4999G>A-F	gtacccctgaggtcagcttg	20
USP54-c.4999G>A-R	cggcccacagtacagtttta	20
FUT11-c.1171G>A-F	cctcagaagctggcagagtt	20
FUT11-c.1171G>A-R	ccccagcactcttacctgtc	20
NDST2-c.1248+57C>T-F	cgtgtggtcatagggacaga	20
NDST2-c.1248+57C>T-R	ctcatagagctgcgtgtgga	20
DLG5-c.5165-58C>T-F	ccacgaggaaaaatcagagc	20
DLG5-c.5165-58C>T-R	atctccttcaccacgtccag	20
SFTPA2-c.*437A>C-F	agtgaggccttggggtagaa	20
SFTPA2-c.*437A>C-R	cagtcacagggttggcttg	19
SFTPA2-c.*362G>A-F	agtgaggccttggggtagaa	20
SFTPA2-c.*362G>A-R	cagtcacagggttggcttg	19
PLAC9-c.187C>T-F	atgtctcagcagccaccttc	20
PLAC9-c.187C>T-R	cacacaatgggtgctcagtaa	21
ANXA11-c.858+42_858+43insGTGT-F	tctgatgaagaccccagtcc	20
ANXA11-c.858+42_858+43insGTGT-R	gtgcgtcatttgcctcttg	19

Table A3.4. Primer sequences of complete gene structure of *FUT11*

Amplicons	Primer sequence	Length (bp)	% GC	Tm (˚C)	Product size (bp)
FUT11-1F	gcctcatctttactcccgaaa	21	48	59.5	483
FUT11-1R	tctgtgccgtagaagagcag	20	55	60.5	
FUT11-2F	ctattcccccacttcccggga	21	62	65.3	561
FUT11-2R	aggcctggacacctggagtc	20	65	64.6	
FUT11-3F	tccaggtgtccaggactcca	20	60	62.5	453
FUT11-3R	caggttgcttgtatgccagg	20	55	60.5	
FUT11-4F	cctttcacagcagtcttgtcc	21	52	61.2	483
FUT11-4R	tccctccatctctccttgag	20	55	60.5	
FUT11-5F	gagcatccactgcacaaacc	20	55	60.5	572
FUT11-5R	ggattgcttgatccaggagt	20	50	58.4	

Table A3.5. *FUT11*-cDNA primers to sequence confirm full-length *FUT11*-cDNA

Primer	Primer sequence	Length (bp)	% GC	Tm (˚C)
FUT11-1F	gggtggtgttggtccttcta	20	55	60.5
FUT11-1R	gcgactgaaggtggaggtaa	20	55	60.5
FUT11-2F	caacttcttgctgagccacg	20	55	60.5
FUT11-2R	gattgttcggcatccagtcc	20	55	60.5
FUT11-3F	ctcttggctttcttgtcccg	20	55	60.5
FUT11-3R	agccagttcgtagtcacaga	20	50	58.4
FUT11-4F	agctggcagagtttattgact	21	43	57.5
FUT11-4R	caaaatttcgtctgctctg	19	42	53

Table A3.6. Site-directed mutagenesis primer sets used in generating point mutations in *FUT11*

Primers	Primer sequence 5’ → 3’	Length
FUT11-L10X-sense	gccccattagggtggtgt**a**ggtccttctag	30
FUT11-L10X-antisense	ctagaaggacc**t**acaccaccctaatggggc	30
FUT11-V391I-sense	acggcttcgagtgtttc**a**tctgtgactacgaactg	35
FUT11-V391I-antisense	cagttcgtagtcacaga**t**gaaacactcgaagccgt	35
FUT11-H483R-sense	cgaaattttgggattacctac**g**tgaaatcttcatgaagaggca	43
FUT11-H483R-antisense	tgcctcttcatgaagatttca**c**gtaggtaatcccaaaatttcg	43

Appendix III for Chapter 4

Table A4.1. Fine mapping markers for chromosomal locus 10p14-p12.1

MARKER	PRIMER	LABEL	PRIMER LENGTH (bp)	CONC	PRODUCT SIZE (bp)
D10S602-F	CCTGACCTGTGGGTTGAGA		19	10n moles	233-255
D10S602-R	CAGTGTGCCGTGAGCAT	**VIC**	17		
D10S591-F	ACCTCGAAGGTCTGTTCTCC		20	10n moles	212-232
D10S591-R	GGCTTTATGGATCATATTAATCCAC	**NED**	25		
D10S1779-F	TCTGTCTTCAGCACACCC		18	10n moles	265-281
D10S1779-R	GCATATCTGTCCCACTCGATAC	**FAM**	22		
D10S570-F	GCATTCATCCAACAAGCATA		20	10n moles	287-305
D10S570-R	AATTAGTTCCATGGGCACAG	**VIC**	20		
D10S1664-F	AACCGTAATAACTTAGGTGCTC		22	10n moles	160-198
D10S1664-R	TGCTGAAGACAGGTAAAGAG	**NED**	20		
D10S191-F	CTTTAATTGCCCTGTCTTC		19	10n moles	124-152
D10S191-R	TTAATTCGACCACTTCCC	**FAM**	18		
D10S1661-F	ACGCTACTTGCCAGGTC		17	10n moles	250-272
D10S1661-R	ATTGCTTCCCTGAGAGTGT	**NED**	19		
D10S211-F	CTCCTGGTCTCATGCG		16	10n moles	194-211
D10S211-R	CAGGCTCCTACTACCGTC	**VIC**	18		
D10S600-F	AAGTCAGAAATGCAGCAC		18	10n moles	175-193
D10S600-R	AGGCAGACTCGGTGTC	**FAM**	16		
D10S1732-F	GGAACCTAAGGCATGTTGAT		20	10n moles	151-171
D10S1732-R	CCAAGACCCTGTCTGAAAAA	**VIC**	20		
D10S193-F	TATATGCAGTTTGGGATGGG		20	10n moles	213-231
D10S193-R	ATTGGGCTGTGCCTACACTT	**FAM**	20		

Table A4.2. Fine mapping markers for chromosomal locus 5q31.3-q35.3

MARKER	PRIMER	LABEL	PRIMER LENGTH (bp)	CONC	PRODUCT SIZE (bp)
D5S2098-F	CTGTATTCTGCATCANAGGG		20	10n moles	191-231
D5S2098-R	GCTAAGGGTAATCTTAAACTGCC	**FAM**	23		
D5S1984-F	CCAGCCCGCTTAGTGT		16	10n moles	213-233
D5S1984-R	TAGGAGGCTTCCCACATCT	**VIC**	19		
D5S500-F	ACCTATTCGACCTAATGACTAAAGA		25	10n moles	188-214
D5S500-R	ATCGGTGAAATGCAACTACTT	**NED**	21		
D5S2017-F	GAGTCCTCTCACTGATATTTTTGTA		25	10n moles	87-105
D5S2017-R	GTCTATCCTTCCAGATGGTTC	**FAM**	21		
D5S2090-F	CATGGGCATGTTTCAAAAT		19	10n moles	189-205

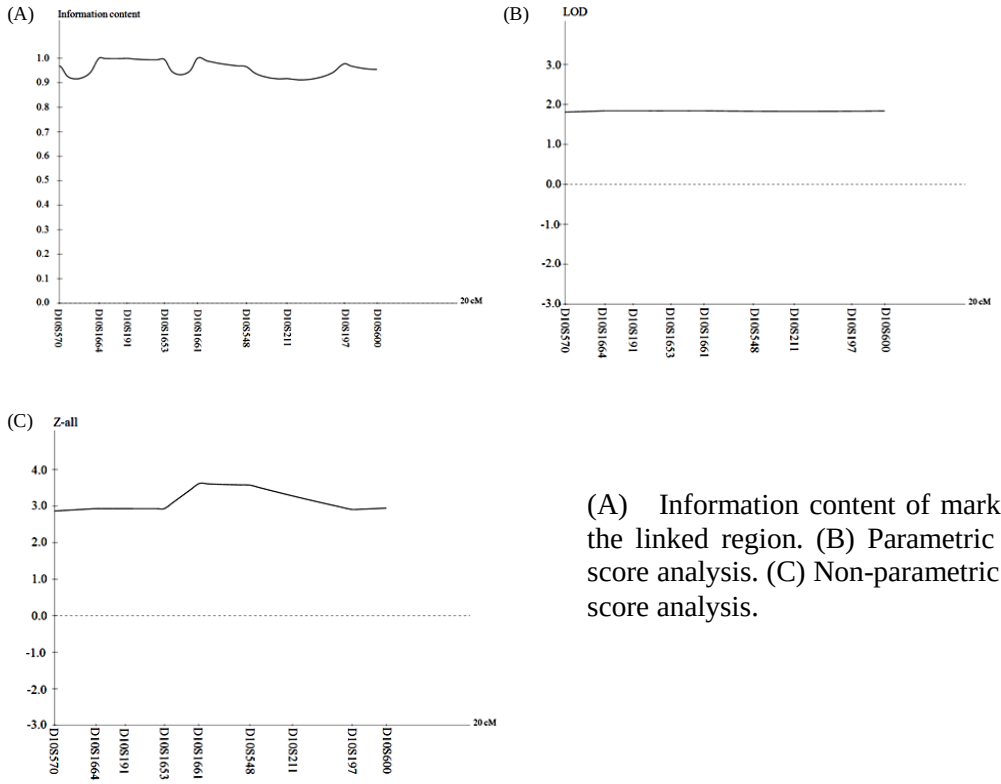
D5S2090-R	AGTACCTCCTTAGTAACTCTGGGC	**FAM**	24		
D5S2014-F	AGCTACTACCAGCAGCATTC		20	10n moles	138-162
D5S2014-R	CTACATTATTATTATTGTGTGTCCG	**FAM**	25		
D5S2112-F	AAATCCATTCACTATGCTGC		20	10n moles	296-312
D5S2112-R	AGCTATTGTCAGTCCACTATTGC	**FAM**	23		
D5S2066-F	TCGAGGCATTGAGTGTG		17	10n moles	221-263
D5S2066-R	CTGATAGAGGACAAGTGGTAGAAAC	**NED**	25		
D5S415-F	AGCNGAATATATGATTTATCATCAA		25	10n moles	115-147
D5S415-R	ATGTCTAAAATAGTTTGGGGATT	**VIC**	23		
D5S671-F	AATGGGTGGGCATCTCC		17	10n moles	197-211
D5S671-R	AACTTATTCCGGGGGGTG	**NED**	18		
D5S625-F	ATGATTCTGGGGATGTGTCTGTCTG		25	10n moles	224-238
D5S625-R	ATGCTGGGATTATAGGCCGTGC	**VIC**	22		
D5S429-F	TGTGTACCAGCATGGTTGAT		20	10n moles	160-186
D5S429-R	CTAGTTTAAGGTTTGCCAGTTTTC	**NED**	24		
D5S2069-F	TTACCAGGCCATGATTTAGG		20	10n moles	293-313
D5S2069-R	GGAAGCCATATGTTTGGAAT	**VIC**	20		
D5S498-F	CTACCAGCCTACCCCA		16	10n moles	171-189
D5S498-R	GATTATTATTCAGAGTTCTCCAG	**VIC**	23		
D5S2034-F	TTTAACAAAATATATAAAATGCCTG		25	10n moles	177-211
D5S2034-R	AATGAATCTTACAACAATTTGG	**VIC**	22		
D5S2030-F	GATCAGCAGCTGTGGTGACA		20	10n moles	150-176
D5S2030-R	ACTCCAGTGCAGCCACGTAA	**FAM**	20		
D5S2006-F	TGTATTCTTAAATTCTGTGAAGAGG		25	10n moles	125-155
D5S2006-R	ATCCATCATCTCCCGTAG	**NED**	18		

Table A4.3. Fine mapping markers for chromosome 3q28 for HWE307

MARKER	PRIMER	LABEL	PRIMER LENGTH (bp)	CONC	PRODUCT SIZE (bp)
D3S3689-F	TCGCACCACTGCACTC		16	10n moles	109-149
D3S3689-R	ACTTCAAAGTCTTAAAAGGCTGTAG	**FAM**	25		
D3S3668-F	CTTTTGGGAATTAAAAACTTCAGG		24	10n moles	235-257
D3S3668-R	GTCAGTAAAACATGGAAATATGAGC	**FAM**	25		
D3S1282-F	TCTTCTGAGATTAAGTTTATTGG		23	10n moles	140-154
D3S1282-R	ACAGGTTTTCACTTCATTTACT	**NED**	22		
D3S1556-F	TGCACACCTGTAGGCCC		17	10n moles	248-262
D3S1556-R	TGGAGGCATGAGCCATC	**VIC**	17		
D3S3730-F	GACTGGAAAATTCAGCCTCTA		21	10n moles	138-156
D3S3730-R	AAGATGAGTCCTGAGCATGT	**VIC**	20		

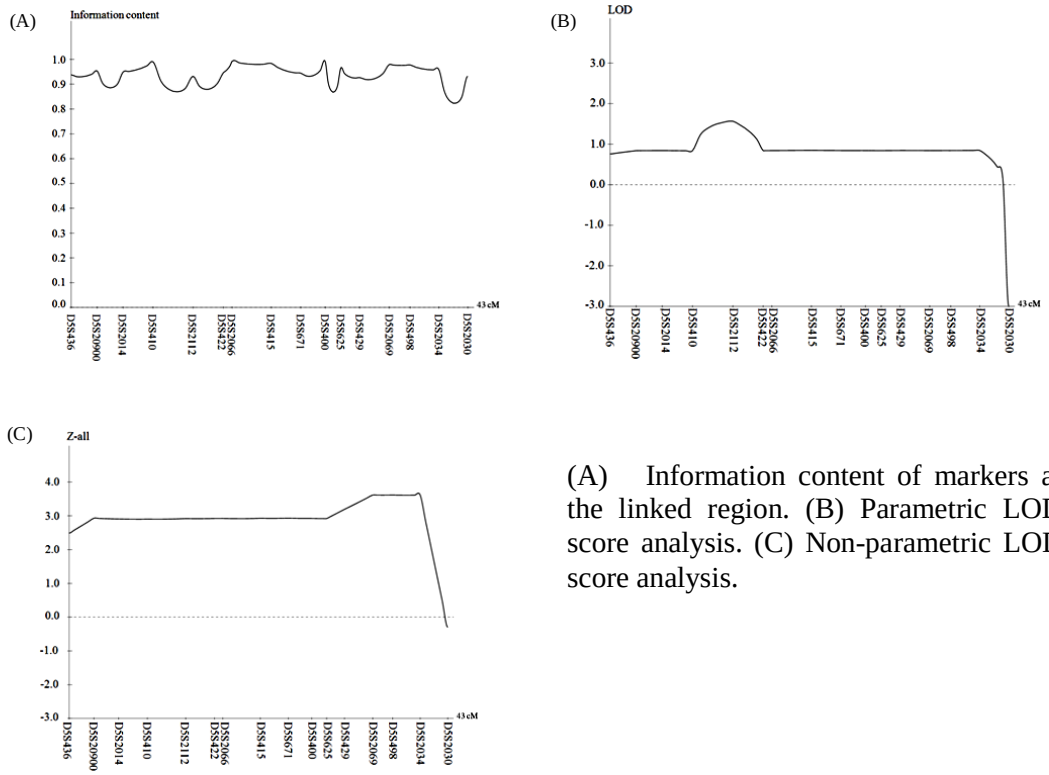
D3S3592-F	GCAGTTCTGAGTGATTTACCA		21	10n moles	159-173
D3S3592-R	TCATCTGAGGTGTCTGATTG	**VIC**	20		
D3S3686-F	AGGGTATTTCATTCCCATTG		20	10n moles	108-134
D3S3686-R	CCAGGTTACGCCAAGTG	**NED**	17		
D3S3530-F	ACAAACCATTTGCCACAGAC		20	10n moles	162-178
D3S3530-R	ACATTGCAGAGACGAAGCAC	**NED**	20		
D3S1314-F	AACTTACACATTTGGCCCTG		20	10n moles	144-170
D3S1314-R	TCAATCTGTGGAGTCATTGG	**FAM**	20		
D3S2748-F	CAGAGGGTTTCACTCTATCACA		22	10n moles	80-112
D3S2748-R	CATTCTATGTTTCTGTGTGTGC	**FAM**	22		
D3S1305-F	CTGCTGGAACTTAAAAGTGC		20	10n moles	189-198
D3S1305-R	AGAAATGAGATATTGTTTTCGC	**FAM**	22		
D3S1265-F	GAAGCAGGAGAATCACGGAG		20	10n moles	212-236
D3S1265-R	CGTTCTGTGTGTTGCTAGTATGTA	**VIC**	24		
D3S3550-F	TGCACTTATGGGAAGTATTT		20	10n moles	238-256
D3S3550-R	CAGCAGCAGCAAGATT	**NED**	16		

Figure A4.4. Multipoint analysis at 10p14-p12.1 using GeneHunter



(A) Information content of markers at the linked region. (B) Parametric LOD score analysis. (C) Non-parametric LOD score analysis.

Figure A4.5. Multipoint analysis at 5q31.3-q35.3 using GeneHunter



(A) Information content of markers at the linked region. (B) Parametric LOD score analysis. (C) Non-parametric LOD score analysis.

Table A4.6. Primer sequences of genes examined at 10p14-p12.1 and 5q31.3-q35.3

10p14-p12.1			
Amplicon	**Forward primer 5' → 3'**	**Amplicon**	**Reverse primer 5' →3'**
CELF2-EX-1F	aaaaggcaactgggaaattg	CELF2-EX1-R	tgggtctcgatttcatctcc
CELF2-EX-2F	gggtggccctaaacctactc	CELF2-EX2-R	ctggtcctcacccggact
CELF2-EX3A-F	acccgcatcttgcattagtc	CELF2-EX3A-R	agctccgttcatcttgttgg
CELF2-EX3B-F	gcctccgctttgttttagttc	CELF2-EX3B-R	tgggctgcctagaaagtgac
CELF2-EX4-F	gcttgagatgttgttattgctg	CELF2-EX4-R	tttctaaccaggggatgcac
CELF2-EX5-F	aaagggtttgattcccgaac	CELF2-EX5-R	ggccacagagaagattccag
CELF2-EX6-F	acgtgcttggtctcacatgg	CELF2-EX6-R	ctgttggcttcacacctgac
CELF2-EX7-F	ttggtcagattagcaagtcagc	CELF2-EX7-R	actggagagcaagtggaagg
CELF2-EX8-F	agatcccccaaagaaaccac	CELF2-EX8-R	ctttgcccctcctacatcag
CELF2-EX9-F	ggatttacagccccagctc	CELF2-EX9-R	gtcatgcacagcgaaggtg
CELF2-EX10-F	ttgcatcagagagaactgaagg	CELF2-EX10-R	gaatgtgtgagaacggcttg
CELF2-EX11-F	caagcacagctcctcagctc	CELF2-EX11-R	gtggttggggattttctgag
CELF2-EX12-F	caccatttcctcatcacacg	CELF2-EX12-R	agcacacaattacactgacagg
CELF2-EX13-F	cattttcccccattatcacg	CELF2-EX13-R	gatcctcccatctcaacctg
CELF2-EX14-F	aaggcactccagcacttgac	CELF2-EX14-R	gcccaaatctgcttacaagg
CELF2-EX15-F	acggtggactcacgatcag	CELF2-EX15-R	aaactgggtaggggaagtgg
CAMK1D-5'utr+Ex1-F	agggaggcaagaaagtagca	CAMK1D-5'utr+Ex1-R	tgacaggctgggtttctgc
CAMK1D-Ex2-F	caactgtgccgtcgttattg	CAMK1D-Ex2-R	ttggttatcgtttgcacagc

CAMK1D-Ex3-F	tgggtctggaaaaagtctgg	CAMK1D-Ex3-R	ttacctggcccacttcaatc
CAMK1D-Ex4-F	gggcaacatctcaatctgaag	CAMK1D-Ex4-R	acacaaatgcccaccatctg
CAMK1D-Ex5-F	ctgtgtccaaaggtcaaggac	CAMK1D-Ex5-R	gctgtgattacatgccaagc
CAMK1D-Ex6-F	ttcaggccaaatagcagacc	CAMK1D-Ex6-R	ccttaaacatggccctgaac
CAMK1D-Ex7-F	ctgtgccttgccagattttc	CAMK1D-Ex7-R	tgttacagcagccaaccaag
CAMK1D-Ex8-F	tcatgctacacacgctcctc	CAMK1D-Ex8-R	ggggagaggaggaaagtcac
CAMK1D-Ex9-F	atgggtgggtttagggaatg	CAMK1D-Ex9-R	tgaggttctggggatttcag
CAMK1D-Ex10-F	gggacaacataagctttcc	CAMK1D-Ex10-R	gttgggattacaggcgtgag
CAMK1D-Ex11-F	cagcctggtgaaaagagtga	CAMK1D-Ex11-R	tccagtcctatgcaggaagg
CAMK1D-3'utr-a-F	ttgccaggcgctttctatac	CAMK1D-3'utr-a-R	ccaggcctttatccagacag
CAMK1D-3'utr-b-i-F	cagtgcactctggaagcaag	CAMK1D-3'utr-b-i-R	gctgggattacaggcatgag
CAMK1D-3'utr-b-ii-F	ctcaaaggagaatttaaagg	CAMK1D-3'utr-b-ii-R	gagagaaacagaaggagta
OPTN-5'UTR1-F	gtgacgccttagagcagtcc	OPTN-5'UTR1-R	cccaccctgtgttcactacc
OPTN-5'UTR2-F	ggtggttggtaacacagaagc	OPTN-5'UTR2-R	gagggagacgagggtatgac
OPTN-EX1-F	ccctttttatacacccatacacac	OPTN-EX1-R	cccaccagctaccacctatg
OPTN-EX2-F	caaggctaagcatggcatc	OPTN-EX2-R	ggcaaaacaccaatccagac
OPTN-EX3-F	tgtaaagatgggggtcttgc	OPTN-EX3-R	aatgttcaatttgccaggac
OPTN-EX4-F	ttccttgggttgcatgtc	OPTN-EX4-R	agtgtgagccaaacaggaac
OPTN-EX5-F	gcattgtaagctggcctctc	OPTN-EX5-R	ttggaggagcagacagtgag
OPTN-EX6-F	gagtttcacttgccttttacctc	OPTN-EX6-R	aattgacacagagcaggacaag
OPTN-EX7-F	ttggggtattgtcaaagttgg	OPTN-EX7-R	ttcatgctcacacattaactgg
OPTN-EX8-F	gcgtaagggttcagaaatatgg	OPTN-EX8-R	agacagagtgcaaacaatcagc
OPTN-EX9-F	cgtggggtgataaaggtagg	OPTN-EX9-R	tcagaagttacaaaccctagatgc
OPTN-EX10-F	ccacctcagcctctcaattc	OPTN-EX10-R	aagatccactgagcactttcc
OPTN-EX11-F	agcaggattgtgcatctgtg	OPTN-EX11-R	ggcgcgaacacagctattc
OPTN-EX12-F	tgaaccttggcagtgtagtttg	OPTN-EX12-R	gattcggtgggtaatggatg
OPTN-EX13-F	tgtgctcatgtcccactacg	OPTN-EX13-R	gccacttcctgggttcaag
FRMD4A-5'UTR1-F	tggagagatgggagggaag	FRMD4A-5'UTR1-R	cacggaatgtttaactgctacc
FRMD4A-EX1-F	tcctccctaagcacagttcc	FRMD4A-EX1-R	aaagagcctgcgggataaag
FRMD4A-EX2-F	ctccagatggtgatattccaaatag	FRMD4A-EX2-R	ccaactgttcaggaaacgtg
FRMD4A-EX3A-F	cgggataggttgactttacctc	FRMD4A-EX3A-R	tcaccaccaccctcactaatc
FRMD4A-EX3B-F	gccgttatctgcacactgag	FRMD4A-EX3B-R	catggcacctagaagatgagg
FRMD4A-5'UTR2-F	gcattgttgtccaggatcg	FRMD4A-5'UTR2-R	agggttgcagaagaaagcac
FRMD4A-5'UTR3-F	cttggtgccttttctccag	FRMD4A-5'UTR3-R	tcccaggatgaagtcagcag
FRMD4A-5'UTR4-F	tgtgaaggaaactgaagtggag	FRMD4A-5'UTR4-R	ttgcctactcgttgacatcc
FRMD4A-EX4-F	gaactgatccttccaactcctg	FRMD4A-EX4-R	tggtgggtcctagtcattctc
FRMD4A-EX5-F	ccttgacttggctgtgtgc	FRMD4A-EX5-R	gtaccaggcatgaccacaag
FRMD4A-EX6-F	tgggtcacagagcaagacttag	FRMD4A-EX6-R	tttatccctggcaatgaacc
FRMD4A-EX7-F	tggtagacccagttctttgg	FRMD4A-EX7-R	tggtgccgttagcttcataac
FRMD4A-EX8-F	gaggcagcatctcattatgttg	FRMD4A-EX8-R	tgctatggtttccgtcagtg
FRMD4A-EX9-F	ggcaaaggtgtgcattcag	FRMD4A-EX9-R	tgtttcagtggacaagttgagag
FRMD4A-EX10-F	ggtgaaaggcggttactgg	FRMD4A-EX10-R	tggtttcaggaagacacagg
FRMD4A-EX11+12-F	aggagctgttggtgccttg	FRMD4A-EX11+12-R	aagatcgcaccctgcacac
FRMD4A-EX13-F	gctccatccattccaagg	FRMD4A-EX13-R	agggaactcagatgggattg
FRMD4A-EX14-F	ccgaaattcccagtatgtgg	FRMD4A-EX14-R	cagacttttcccttccaggag
FRMD4A-EX15-F	ttcccatttcagcactggag	FRMD4A-EX15-R	tgaccatgaacaacgacagc
FRMD4A-EX16-F	atcagaaggtggcctttgc	FRMD4A-EX16-R	tgtagccctaagaggcaagc
FRMD4A-EX17-F	atgatgaggccaagggatg	FRMD4A-EX17-R	ttcctgggagcagttttctc
FRMD4A-EX18-F	gtagaaccacgcagtgcaag	FRMD4A-EX18-R	gaccttggaactcctctgtcc

FRMD4A-EX19A-F	aagggaaccaggaccaaatg	FRMD4A-EX19A-R	aaccaagcacctttgatgg
FRMD4A-EX19B-F	gctgggtattttactggatgc	FRMD4A-EX19B-R	aggaggggaaacttcaaagc
FRMD4A-EX19C-F	tgggaatataccgccttcttc	FRMD4A-EX19C-R	ttcactaataatctggccaacg
FRMD4A-EX20-F	tcacactcccttgattcctg	FRMD4A-EX20-R	atgcggttctgctctttctc
FRMD4A-EX21-F	agcataagcgttgacctgtatc	FRMD4A-EX21-R	taccgaggggaaccacatac
FRMD4A-EX22-F	tgggtgcagattttaggagtg	FRMD4A-EX22-R	aaggcagcggacagttagg
FRMD4A-EX23-F	caaaagaaacacccaacagc	FRMD4A-EX23-R	cactcccaaccaaagaggtg
FRMD4A-EX24-F	tgacttgtcttggcaactgg	FRMD4A-EX24-R	aacagtgtcggcagcagag
FRMD4A-EX25-F	ctggggaaaggctgagaac	FRMD4A-EX25-R	gcatgggtctcttggctaac
FRMD4A-EX26A-R	tgagtgagactctggcttgc	FRMD4A-EX26A-R	tgagccttgacgctgtagtg
FRMD4A-EX26B-F	tgtacctgcacagccagagc	FRMD4A-EX26B-R	atgattcccgttcctcactg
FRMD4A-EX27-F	gatgctgagaccctgctctg	FRMD4A-EX27-R	acacctcccagtgatgaacc
FRMD4A-EX28-F	gcatgacttgtatgatctgttgc	FRMD4A-EX28-R	agttccagttccccatgtatc
FRMD4A-EX29-F	ctccgaacaggaaagactcg	FRMD4A-EX29-R	ggtcaccctgtaactcaagtcc
CDNF-5'utr+Ex1-F	cccttcctatccccttccttg	CDNF-5'utr+Ex1-R	ttatagccagggccaggag
CDNF-Ex2-F	atgcccagccccatatagac	CDNF-Ex2-R	ttgctgcgtgctctaaaatg
CDNF-Ex3-F	caccatgtccagcctacctc	CDNF-Ex3-R	ccattcatcagccaagcaac
CDNF-Ex4-F	tcatagggtagtcagcaccaag	CDNF-Ex4-R	tgatgcattcccagttatcc
CDNF-3'utr-a-F	tgatctccaatgccagcac	CDNF-3'utr-a-R	ccatttctgataccaaagagagc
CDNF-3'utr-b-F	acaaaagacacaacagtcagc	CDNF-3'utr-b-R	atgtggccagaaaatggaac
HSPA14-5'utr+Ex1-F	gtgcgcactgtgcagtttc	HSPA14-5'utr+Ex1-R	gctctcaagcctgtggatg
HSPA14-Ex2+3-F	tttccacacagatggcaaag	HSPA14-Ex2+3-R	ggtactctctgtaatgactc
HSPA14-Ex4+5-F	cagtgatattgtgaaatacagtc	HSPA14-Ex4+5-R	ccaagatctaaaaatacctc
HSPA14-Ex6-F	gtgggcaagcaagtgtattg	HSPA14-Ex6-R	ggcctcttggcatttcatc
HSPA14-Ex7-F	ttcagtacgtgtgctttttgc	HSPA14-Ex7-R	ttccatccagaaaggtttgg
HSPA14-Ex8-F	tttggttgcagaacatctcag	HSPA14-Ex8-R	aaccacatcctcctttgcag
HSPA14-Ex9-F	ggccctggttgatcttgaac	HSPA14-Ex9-R	tggaccccaataaagtgaagac
HSPA14-Ex10-F	ctgtcattcagagaactatc	HSPA14-Ex10-R	gttcaccagagccattaatc
HSPA14-Ex11-F	acatacgatgaagatgactcag	HSPA14-Ex11-R	ctaaatttaagacctgtcag
HSPA14-Ex12-F	ctctttgcctagtgtcctg	HSPA14-Ex12-R	ctgcctatttttttccctct
HSPA14-Ex13-F	cagtgataaataagctttgg	HSPA14-Ex13-R	cctctgtaggtttttaatgtttc
HSPA14-Ex14+3'utr-F	ctaaatttagagacactgag	HSPA14-Ex14+3'utr-R	cttggttatctaacaagaatc
NMT2-EX1-F	ggggttagagagctggattg	NMT2-EX1-R	gcaagtgacagtagcgtctcc
NMT2-EX2-F	ttcctcacctacatcctaacacag	NMT2-EX2-R	tggcattcaacacacagtcac
NMT2-EX3-F	ttggaggaggattgttagtgtttag	NMT2-EX3-R	ggttatcaaaagccaatgtcac
NMT2-EX4+5-F	accactgcctagctttcctg	NMT2-EX4+5-R	acaggcaccagttattgtcc
NMT2-EX6-F	attcacccgagttcctgttg	NMT2-EX6-R	ttacaggtgcctaccaccac
NMT2-EX7-F	cccagcctgcattcttattc	NMT2-EX7-R	taatcaccaagtggcaaatg
NMT2-EX8-F	ctggacggataaaatgatgac	NMT2-EX8-R	ggcatttccaaacaaactcc
NMT2-EX9-F	tcttgttgagttgcagattcg	NMT2-EX9-R	ctctctacccttttagtacacatctttg
NMT2-EX10-F	gggcaagggagagaacagag	NMT2-EX10-R	gtctcgcataaattccacctg
NMT2-EX11-F	ataggcatgagccacagc	NMT2-EX11-R	caggcatggtggtggatg
NMT2-EX12-F	ggattacaggtgtgagctgag	NMT2-EX12-R	gattccataaatgttcccagtg
ITGA8-EXON1A-F	gttagctctgccaggaggtg	ITGA8-EXON1A-R	gtggctgctacccaggag
ITGA8-EXON1B-F	tagaccagccgcgaggag	ITGA8-EXON1B-R	atggcaggggagagaagg
ITGA8-EXON2-F	aatacgttcatgccctccac	ITGA8-EXON2-R	taaacgagcagcacaccaag
ITGA8-EXON3-F	acaggcatgagccaccatac	ITGA8-EXON3-R	aaatgccaacagccttatcg
ITGA8-EXON4-F	tgaccaaggcactcattagc	ITGA8-EXON4-R	cacgcttttgacttgctgtc
ITGA8-EXON5-F	gaatcacaacgctatttggtg	ITGA8-EXON5-R	tttaataggccgctatgtgc

ITGA8-EXON6-F	ccatgtctgcaaaactgcac	ITGA8-EXON6-R	aacgcttcatttgagaaactatg
ITGA8-EXON7-F	ctctgccttggagtttgagc	ITGA8-EXON7-R	gcggctggaaaaatagcag
ITGA8-EXON8-F	ttaaaattgacctcagaatatctcc	ITGA8-EXON8-R	ttcattcataacatgggattatgg
ITGA8-EXON9-F	cagacaggtgctggttgatg	ITGA8-EXON9-R	tgatgttcggtggattaggtg
ITGA8-EXON10-F	agatcattccgtgggctttc	ITGA8-EXON10-R	agctggatgggatctgtctg
ITGA8-EXON11-F	ttcactgggtcaggggttag	ITGA8-EXON11-R	gcagcagatgtggaaaatctc
ITGA8-EXON12-F	aagcagcgaaagcaatctg	ITGA8-EXON12-R	aaacgtaaagctgccttgag
ITGA8-EXON13-F	acttgagtccgggagtgtg	ITGA8-EXON13-R	tgaaagtcgctctggaatcg
ITGA8-EXON14-F	gaagctccaacatcctttcc	ITGA8-EXON14-R	ctggaagtttggtcatgcag
ITGA8-EXON15-F	ttggcacaaatgctgagaac	ITGA8-EXON15-R	ctcaaacttcggccacagac
ITGA8-EXON16-F	cctcctgtactattttccagtattacc	ITGA8-EXON16-R	ccaagcatgataaatgagctg
ITGA8-EXON17-F	cagctcatttatcatgcttgg	ITGA8-EXON17-R	tgaagtcagggtcaagaggag
ITGA8-EXON18-F	ccacataaatcagcctccatc	ITGA8-EXON18-R	tgcacaagaggaaaggcttc
ITGA8-EXON19-F	cctgtattgggaggtcactg	ITGA8-EXON19-R	gacccaaaccacaggctaac
ITGA8-EXON20-F	tgcctcttcctaccatctcttc	ITGA8-EXON20-R	ttgtggctcattgctgttatg
ITGA8-EXON21-F	tttctcttctgagttaagggtcatc	ITGA8-EXON21-R	gttgttgctgctaggtgctg
ITGA8-EXON22-F	gcagaacaaaggaaagagttacag	ITGA8-EXON22-R	ggtgagcttttaaggccaag
ITGA8-EXON23-F	ttagtaacgtgtgttctcctgtg	ITGA8-EXON23-R	gggtttactctgctttccaatg
ITGA8-EXON24-F	gctggagagcctacactagcag	ITGA8-EXON24-R	tttgagcccagaaagtcgag
ITGA8-EXON25-F	agcccatgatgctttcagag	ITGA8-EXON25-R	tttggatgtgctgctaattg
ITGA8-EXON26-F	cacactgctggtaatcacagg	ITGA8-EXON26-R	gcaagtgctccttaattcttacag
ITGA8-EXON27-F	ttcatcctcgccattaaactg	ITGA8-EXON27-R	tggcctcaagtaatccatctg
ITGA8-EXON28-F	acgcttagatggatggatgc	ITGA8-EXON28-R	gaaattggtgagctgaagagg
ITGA8-EXON29-F	ttcatgttcgagtgctacctc	ITGA8-EXON29-R	tgccaaagcaaacaagacac
ITGA8-EXON30A-F	tggggaatagagagcgagac	ITGA8-EXON30A-R	agggaggtgtcaaatgttcc
ITGA8-EXON30B-F	aaggtgaactaaggtgaaatgactg	ITGA8-EXON30B-R	ccatgacttttaaacactccaaag
ITGA8-EXON30C-F	tgctaaggaaaggaagattgg	ITGA8-EXON30C-R	ttcaagatgatgggaaaatagc
ITGA8-EXON30D-F	tacaggcatgagccaaactg	ITGA8-EXON30D-R	tcaggacacaaagttttccaac
RSU1-5'UTR-F	ctcgaagccgaataacgaac	RSU1-5'UTR-R	gatgacagcaagcgcagag
RSU1-EX1-F	gaacccttcccactccatc	RSU1-EX1-R	tcctactgcaaaccctctgc
RSU1-EX2-F	ttggtgagcgatagtgcttc	RSU1-EX2-R	agaccaacctgggtaacacg
RSU1-EX3-F	ggatcatggcactgtcttacac	RSU1-EX3-R	tgcacatactgctgagctacg
RSU1-EX4+5-F	aaagctgactccctcgtattg	RSU1-EX4+5-R	aggtggagagtagttgccaag
RSU1-EX6-F	ccaagaggcacttagggtttg	RSU1-EX6-R	ctgccatcaataggatactgtcac
RSU1-EX7-F	aacacagcgtagcacccagag	RSU1-EX7-R	tgttccctgcttttggtaatg
RSU1-5'UTR-F	ctcgaagccgaataacgaac	RSU1-5'UTR-R	gatgacagcaagcgcagag
CUBN-EX1-F	tttgacctctcacaagttgaag	CUBN-EX1-R	tgtttttccctgtttgcttctc
CUBN-EX2-F	agatgagcaggggcagttg	CUBN-EX2-R	gggaggctgaagacaggttg
CUBN-EX3-F	ccagtgtttattacgcatggtg	CUBN-EX3-R	tttcctgccttctgtccatc
CUBN-EX4-F	aggggagaggaatgaagagg	CUBN-EX4-R	gcaaactcagggtcagaagc
CUBN-EX5-F	cacaagagaggtgatttgaagaag	CUBN-EX5-R	atttgggaggccacagaatc
CUBN-EX6-F	attcctctcctgtgactgctg	CUBN-EX6-R	gctatgtagacgttaagcaagagc
CUBN-EX7-F	attgctggctgacgattc	CUBN-EX7-R	tcctcgcctatgtcctacttac
CUBN-EX8-F	ctctttccctggcatctttg	CUBN-EX8-R	agatcgtcgagcagaaccag
CUBN-EX9-F	tttgcccactttgttgtcag	CUBN-EX9-R	atgtgatctgcccacctcag
CUBN-EX10-F	gaaacaacccagaacattgc	CUBN-EX10-R	tttctcctgcctgcttgttg
CUBN-EX11-F	ctgcccgtcaaaatacacag	CUBN-EX11-R	cttgatctcctggctcgtg
CUBN-EX12-F	tgcaaaaagttagccaggtg	CUBN-EX12-R	gaaagtgggttcagcagagg
CUBN-EX13-F	gaatggcagggcttttatctc	CUBN-EX13-R	ccagcccctgtgacttatttc

CUBN-EX14-F	ataccttggtctggtggacttg	CUBN-EX14-R	tcttcagatcccaaaacatctc
CUBN-EX15-F	gcacagagcttgacacatgg	CUBN-EX15-R	gaagcaacaaacaggcacag
CUBN-EX16-F	cagacagagagaaggcattgg	CUBN-EX16-R	gggatcaaagggtattgcag
CUBN-EX17-F	tcccacaacacatcaaactg	CUBN-EX17-R	tcagccttatttgcctgcac
CUBN-EX18-F	ttgcagatttggctaacacg	CUBN-EX18-R	gcctgaggagataaccgaatataac
CUBN-EX19-F	ttcagagctgtttatcaagtcg	CUBN-EX19-R	aattctcgggacaaatctcac
CUBN-EX20-F	gggggtgactaaaatgtgc	CUBN-EX20-R	gactttgaatggtccctgttg
CUBN-EX21-F	cctcctcttgggcttgtatc	CUBN-EX21-R	cagatgtcaaaacttaatgcctac
CUBN-EX22-F	gtgttttaaggttgatgagaggtagag	CUBN-EX22-R	cacaggtttgcggatgatg
CUBN-EX23-F	ctccatagcttagatttagtgttgg	CUBN-EX23-R	tcaggctccaaatgacaaac
CUBN-EX24-F	aacatattctctcagtttgtgattgg	CUBN-EX24-R	tgttaaaaagaaatggctgctg
CUBN-EX25-F	aagatcatgccgttgcattc	CUBN-EX25-R	tcgaagatggtcactttcacc
CUBN-EX26-F	ttcatggctattcttttgtgatg	CUBN-EX26-R	aatttgaaggcccactctagg
CUBN-EX27-F	ctgctctccacaaattcaagg	CUBN-EX27-R	tatagatgcgtggggcagag
CUBN-EX28-F	cccacctcagcagatttacc	CUBN-EX28-R	tccactttgatgtcacttcctc
CUBN-EX29-F	tttcggtgagttgaatttgg	CUBN-EX29-R	ttcggaatcttggcagag
CUBN-EX30-F	ttcctaaggcacagggagtg	CUBN-EX30-R	tctcagtaggctgccctttg
CUBN-EX31-F	gggataaatggcttaatgtgg	CUBN-EX31-R	tagcgcagaaactgacttgc
CUBN-EX32-F	ccctagaaatgaccaaagacttg	CUBN-EX32-R	gagttgggataattgtaagtcacag
CUBN-EX33-F	gtgctcctgttttctcatgg	CUBN-EX33-R	gcaactgtagaaatgacttccctatc
CUBN-EX34-F	tcattaaccatccccacctc	CUBN-EX34-R	ttggcttaggctgctctttc
CUBN-EX35-F	tgaccctgaatgaagtggttc	CUBN-EX35-R	ttccaaccaatcctccacac
CUBN-EX36-F	ggcccttggtcttcagtttg	CUBN-EX36-R	tggtcgagagaaacttcatgc
CUBN-EX37-F	gagacagccaaataatgttgtg	CUBN-EX37-R	ttgcccttattgagccattc
CUBN-EX38-F	tccttttcaacacccttgc	CUBN-EX38-R	cagactatcccacatgatttgc
CUBN-EX39-F	catcttatgagtgattcccacag	CUBN-EX39-R	atatggtggaagggctttg
CUBN-EX40-F	attgcagctttccgaatcag	CUBN-EX40-R	aattccaaaccaaacggatg
CUBN-EX41-F	ggtaaagaaaagggcaggaag	CUBN-EX41-R	ctgaaggcagcacacttttg
CUBN-EX42-F	ccaccacaaactgagtagcc	CUBN-EX42-R	gtcccaagggtggagaacag
CUBN-EX43-F	cttggtggtaggtcttgctg	CUBN-EX43-R	ggtatgtggccttggaaaag
CUBN-EX44-F	acagatgaaaggaaatggagac	CUBN-EX44-R	tggccgactgtgttttcag
CUBN-EX45-F	tgacatggcttatgctgattc	CUBN-EX45-R	aattttggggagcagtatgg
CUBN-EX46-F	tggacttgtggtatcatttctg	CUBN-EX46-R	aattgagggagagaatcaccag
CUBN-EX47-F	tcttcataggcagggtcttg	CUBN-EX47-R	ggtgggcattacagtataactctc
CUBN-EX48-F	ctaaattgctgcctgtggac	CUBN-EX48-R	tgatgcacagagtgatagctg
CUBN-EX49-F	gggtaacagtggcttgtgg	CUBN-EX49-R	cagtgcgtttttcatctgc
CUBN-EX50-F	catgccttttacccataagc	CUBN-EX50-R	actgaggcaagaggatcacc
CUBN-EX51-F	caaatattgtgcaatgtctgctac	CUBN-EX51-R	aaatttctagatatcagccacaacc
CUBN-EX52-F	ggtgatttgcccacttcg	CUBN-EX52-R	gccaatgtatgtgctctagtgag
CUBN-EX53-F	gttgccatcccaaaccag	CUBN-EX53-R	atcaacccatgcttcctcag
CUBN-EX54-F	atggagagagcttgctgtgc	CUBN-EX54-R	aattttgcctgtggtagtgc
CUBN-EX55-F	aagtcaagaatgggacatgc	CUBN-EX55-R	tgactcatcctccccgtagac
CUBN-EX56-F	gggatttgcagtagatttgtgg	CUBN-EX56-R	tgccaatagcgaataacaagg
CUBN-EX57-F	cctgttttatcatacctgtgcttg	CUBN-EX57-R	ccatgaacctcactgacaatc
CUBN-EX58-F	ccctcttgctgtctttcatc	CUBN-EX58-R	ttcccaggttgaactgtcatc
CUBN-EX59-F	tgagttttgaatgctctactgg	CUBN-EX59-R	aaaaccaagcccttacatgg
CUBN-EX60-F	cccaggagggaatgatctg	CUBN-EX60-R	aatgttggaagccagtcacac
CUBN-EX61-F	ggcgtcataacaggtttttcc	CUBN-EX61-R	tcatgtctcatctcaggcaatc
CUBN-EX62-F	aggcagggaagacagaattg	CUBN-EX62-R	tttcatcatggcttcattacttc

CUBN-EX63-F	ttgcagagcacttccaacag	CUBN-EX63-R	gctaaattgggctgggttc
CUBN-EX64-F	cctgtaggcgtgtttgtgtg	CUBN-EX64-R	aatttccggtgtttggactg
CUBN-EX65-F	ctgcctggtgttttaattgc	CUBN-EX65-R	tggatgacacagcgagattc
CUBN-EX66-F	ggtgtagcccaaagaagagg	CUBN-EX66-R	acatggcaaaaccctgtttc
CUBN-EX67A-F	cttccccattttccctcaac	CUBN-EX67A-R	gaattacaggcacccaccac
CUBN-EX67B-F	ccccctgtattctcagcac	CUBN-EX67B-R	cctcctgtgtagctggaacc
CUBN-EX67C-F	aataatagccaggggcagtg	CUBN-EX67C-R	gcttggctccagaaatgaag
SLC39A12-5'UTR-F	ctattctgcgaggtgaaatcc	SLC39A12-5'UTR-R	tgactgtgggctagtaacaaagg
SLC39A12-EXON1-F	tcctgccatttggggttag	SLC39A12-EXON1-R	agatttgcctgtgattttgc
SLC39A12-EXON2-F	ctgcctgcttttcctttctg	SLC39A12-EXON2-R	acttctggcgaaaactgctc
SLC39A12-EXON3-F	ggaccagtgagatgagggttag	SLC39A12-EXON3-R	ttgaaaggccacataagctg
SLC39A12-EXON4-F	atgtccattctgtcatctctagg	SLC39A12-EXON4-R	tcactcaatctgaacccaagg
SLC39A12-EXON5-F	gtcgtaccggcttcatgtg	SLC39A12-EXON5-R	ttgaatggcaactccatcac
SLC39A12-EXON6-F	tccatatccctactggctgttac	SLC39A12-EXON6-R	atgtggcgagcctaatgttg
SLC39A12-EXON7-F	catacccaaaagcaaactgg	SLC39A12-EXON7-R	gaaggagctgtgtattcttgaaatc
SLC39A12-EXON8-F	tgtggcaactaccattttgg	SLC39A12-EXON8-R	aacagctttcaggtgcttgg
SLC39A12-EXON9-F	cattgtgtgtgtctatgaaacctg	SLC39A12-EXON9-R	atttgagagctggataggaattag
SLC39A12-EXON10-F	ttgcggaaacaaagtgatg	SLC39A12-EXON10-R	ggattttattccaaggtacaggag
SLC39A12-EXON11-F	cagccaaaggaatggaactc	SLC39A12-EXON11-R	gcaactgctatcacacacctg
SLC39A12-EXON12A-F	aaagacatggtagagtaggtattcagc	SLC39A12-EXON12A-R	aaagcaagagcttcatcatgg
SLC39A12-EXON12B-F	tcttaggcaaagtgtgtctctttc	SLC39A12-EXON12B-R	aggtctcccaaaatggcttg
ARL5B-EX1A-F	ccgtttgggactctcctacc	ARL5B-EX1A-R	aacgctctgtcccctgtctc
ARL5B-EX1B-F	tgatggggctgatcttcg	ARL5B-EX1B-R	tttctgggcggacacttatc
ARL5B-EX2-F	ctcccacattcttttacatgagaac	ARL5B-EX2-R	aagtgtttgtgaagggtttagc
ARL5B-EX3-F	aattcacaccacactgtttcc	ARL5B-EX3-R	gaatggcttgaactcagcag
ARL5B-EX4-F	tggggcacagtttttatttg	ARL5B-EX4-R	tttacctcaagactttcatcagtttag
ARL5B-EX5-F	ttttcagtttgaccagttcgac	ARL5B-EX5-R	gtttttaagttgccagtgattttac
ARL5B-EX6-F	tgtgacttgaatgttgacactg	ARL5B-EX6-R	ggagtgctggtcaagaattg
PLXDC2-EX1A-F	ccttccctcctcaaagtgtg	PLXDC2-EX1A-R	gcaaactgtcggttccactc
PLXDC2-EX1B-F	gcgctggctgtggaattag	PLXDC2-EX1B-R	acgagagtgccagagaggtc
PLXDC2-EX2-F	tccaatcagagatgttcgtacc	PLXDC2-EX2-R	accaagcctgtgtttgtctg
PLXDC2-EX3-F	tttcttacagttcgtatgtgtgtg	PLXDC2-EX3-R	cacttgaaaagcgattaccac
PLXDC2-EX4-F	atattctttagtacgaccgaggac	PLXDC2-EX4-R	aagaaaagcagaggtgatctagg
PLXDC2-EX5-F	ccagctaccatgacattttattttc	PLXDC2-EX5-R	tcagaccatttgcatcaagc
PLXDC2-EX6-F	cagatttgaaggccagatgc	PLXDC2-EX6-R	ccctttcttgacatttggag
PLXDC2-EX7-F	ggaactccctgtctcttctattc	PLXDC2-EX7-R	agttagctcaagggtgacagag
PLXDC2-EX8+9-F	tcagtcttctcacaggatgagg	PLXDC2-EX8+9-R	caagcatccctcctactttctc
PLXDC2-EX10-F	gcttctccagttgcgagttc	PLXDC2-EX10-R	tgcatctattaggggttgtagc
PLXDC2-EX11-F	aatgccttgttccatccatc	PLXDC2-EX11-R	caagaagccagaatcaatgg
PLXDC2-EX12-F	ctagtcatgttccgaccaagg	PLXDC2-EX12-R	agaagcccgtattgcctttc
PLXDC2-EX13-F	gcagaaaagttgtccaggaag	PLXDC2-EX13-R	catccatgtgtgtgaaatcag
PLXDC2-EX14A-F	ggcttggagggttgataagg	PLXDC2-EX14A-R	cagataagtccacgggagatg
PLXDC2-EX14B-F	gctgctgtagcctgaagaagac	PLXDC2-EX14B-R	tttgttctcactcttgcatgg
MLLT10-5'UTR1-F	tctgtctaagcagcgcattg	MLLT10-5'UTR1-R	agcgaatcccttcccaag
MLLT10-5'UTR2-F	tcttcccgacgctcacac	MLLT10-5'UTR2-R	cacatgcaagccgattagtg
MLLT10-EX1-F	gagtgactgagcggcaaag	MLLT10-EX1-R	cgcactttctcccacatgc
MLLT10-EX2-F	ggatgtgaactgttttacttctctg	MLLT10-EX2-R	agccagtgtaaccttcaaacttc
MLLT10-EX3-F	gcataagggcaggaatgaag	MLLT10-EX3-R	aggttgaggtgggaggatg
MLLT10-EX4A-F	cctttagccagatggatgttg	MLLT10-EX4A-R	ccttcttgaaacgccaaatg

MLLT10-EX4B-F	agtttctggtgctctgatgc	MLLT10-EX4B-R	gtcactgaagaacaaaatggacag
MLLT10-EX5-F	ctaatccgcctgcctcag	MLLT10-EX5-R	catactgcataaaatacccaatgc
MLLT10-EX6-F	gcaatggaagtaagacagcaatg	MLLT10-EX6-R	aaggtaacacagttttgcttgc
MLLT10-EX7-F	gactgatgaagataatgcctttgtg	MLLT10-EX7-R	tcagaaccagagtccacttattg
MLLT10-EX8-F	tgcaggttggggtgaaatac	MLLT10-EX8-R	ctccagccaaaaactcaagg
MLLT10-EX9-F	cggcattttctctttgtattgaaac	MLLT10-EX9-R	attcacaaatatggcattaaacc
MLLT10-EX10-F	cagaggaagtgatttttgacagaag	MLLT10-EX10-R	accaggtaagtggaggttgc
MLLT10-EX11-F	gattgtcatgtggaagtgaatagc	MLLT10-EX11-R	tgtcaccctgttgccatc
MLLT10-EX12-F	ggatgattttcttaggtgaactgac	MLLT10-EX12-R	ttcaactctggaagcacaaatc
MLLT10-EX13A-F	gctgtggcaaaaggtaggtg	MLLT10-EX13A-R	caggcccatgcttactttg
MLLT10-EX13B-F	accttaattggcctcccttc	MLLT10-EX13B-R	gccgaacatctttcccaaac
MLLT10-EX14-F	cattccacacactcccattaag	MLLT10-EX14-R	gcagtacattatctttgcaggag
MLLT10-EX15-F	agatttgcatcctacagtgactc	MLLT10-EX15-R	gatgaaccatattgacctatcagc
MLLT10-EX16-F	tggggattttaaccctgatg	MLLT10-EX16-R	cctaccaagtctgaagcaatatg
MLLT10-EX17-F	cctccctatcctgaacagatg	MLLT10-EX17-R	aagctctagcaatgcctaccac
MLLT10-EX18-F	ggtgtaggttcttttagacaactgc	MLLT10-EX18-R	ttggtgcccatacactcatc
MLLT10-EX19-F	aatttgcttttcagaaggtcac	MLLT10-EX19-R	gacattgctattttgcagcac
MLLT10-EX20-F	tgggaactgcaaagatagtgg	MLLT10-EX20-R	gagctaaaactggaacccaaac
MLLT10-EX21-F	ggaacctgatacttctggcatag	MLLT10-EX21-R	caggattcaagtccaagcag
MLLT10-EX22-F	aggaacactttgcctttgtg	MLLT10-EX22-R	tttctacatcttcatataactggtgac
MLLT10-EX23-F	ggttttgatacgggcttcag	MLLT10-EX23-R	actcctgagccttgttgagc
MLLT10-EX24-F	ttccttgaacagcagtaagagc	MLLT10-EX24-R	acacattgaaagccccaatc
MLLT10-EX25-F	ctgttttcagagtgtttggattcag	MLLT10-EX25-R	acgcaccacatctcagcac
MLLT10-EX26-F	tctcctatgtgtgtcctgttttg	MLLT10-EX26-R	ggggacgtataaattggcttag
MLLT10-EX27A-F	agtagccttgccaacttgc	MLLT10-EX27A-R	actgtaacccaaacaaaagagc
MLLT10-EX27B-F	gctttgttgcactgaaatgg	MLLT10-EX27B-R	tgtcccaccctgctgtctac
MLLT10-EX27C-F	ggaaccagatcaattcaaagc	MLLT10-EX27C-R	cagaaatgaggtaagaactgtgc
MLLT10-EX27D-F	tgtgacagtaccgcagagtg	MLLT10-EX27D-R	aaaacacaggccaagactcc
PIP4K2A-EXON1-F	cgcagcctaacggtccag	PIP4K2A-EXON1-R	gaggaggaggggaacgag
PIP4K2A-EXON2-F	agattatgatggcagagggaag	PIP4K2A-EXON2-R	gacagatggttatgtttatcacagg
PIP4K2A-EXON3-F	ctttctgcgggttgtacctg	PIP4K2A-EXON3-R	gaggtccccaatcaagagaac
PIP4K2A-EXON4-F	ttatctgggaatcggaaagc	PIP4K2A-EXON4-R	gccttggtgagaaatcgctac
PIP4K2A-EXON5-F	tggtaagcagtacaaatgtggtc	PIP4K2A-EXON5-R	ttgcctcacaggaaggaatac
PIP4K2A-EXON6-F	aatcgccctgagaaacattg	PIP4K2A-EXON6-R	ggctggagaagaatcacagg
PIP4K2A-EXON7-F	gagggatgagaagcccattc	PIP4K2A-EXON7-R	caggaagaaccacaatagcag
PIP4K2A-EXON8-F	ccggtggtattcaagccttc	PIP4K2A-EXON8-R	tgagagggctgctaaactgg
PIP4K2A-EXON9-F	ccaggagtgtaagagcatcg	PIP4K2A-EXON9-R	ggagtttggaggatgagtgc
PIP4K2A-EXON10-F	tctctggactccttccatcc	PIP4K2A-EXON10-R	agcagctcagtggcagaaag
PIP4K2A-EXON11-F	gatgtttttcagagcagagactg	PIP4K2A-EXON11-R	cactcattcattcggccatag
GAD2-EX1A-F	acctgcttggaggaaaacg	GAD2-EX1A-R	gtttagggacgtggcaacc
GAD2-EX1B-F	ctttccctcaaatgctctgg	GAD2-EX1B-R	ccgaagaagtttcctgttgc
GAD2-EX2-F	cctcgggaaacagataaagg	GAD2-EX2-R	gtagagcagggctgggtaag
GAD2-EX3-F	aacaaactgtgcggtgagtg	GAD2-EX3-R	tctgaagattctggcaggtc
GAD2-EX4-F	tctcattcattcccttcacc	GAD2-EX4-R	gcagtggaaggtgctttg
GAD2-EX5-F	tcttcgtggttagaggaaacg	GAD2-EX5-R	tctggtcccatttgtctttg
GAD2-EX6-F	tcgtgcattatgtcagtagttgg	GAD2-EX6-R	ttggaggcatgttcaatgg
GAD2-EX7-F	gtagcgcctttgagtctgag	GAD2-EX7-R	ctatattgcccaggctgctc
GAD2-EX8-F	caggctttggaggaacagac	GAD2-EX8-R	catggcatcaaggaaccac
GAD2-EX9-F	ctggaagcagtcaaggagttg	GAD2-EX9-R	acctgggagattgaggcttg

GAD2-EX10-F	gctgctgcttcctcaaattc	GAD2-EX10-R	tataatcagccttgggatgc
GAD2-EX11-F	gcccttcggtgtaaactctg	GAD2-EX11-R	cataaggcaaggaacacaagc
GAD2-EX12-F	ggccagacttcaccactacg	GAD2-EX12-R	ttaggcaccagaccttcacc
GAD2-EX13-F	catttgaccctgagccattg	GAD2-EX13-R	gaaagcaaatcttgggtaagg
GAD2-EX14-F	aaagggaacaaaagcacctg	GAD2-EX14-R	ggaccactttgagagtttcgtc
GAD2-EX15-F	ctctcaaagtggtccatataacg	GAD2-EX15-R	caacctggaagacacagtgag
GAD2-EX16-F	gcttgctgttgggttctttg	GAD2-EX16-R	ggcaatttggcactattcag
GAD2-3'UTR1-F	cctcagattggggttgtgc	GAD2-3'UTR1-R	gttgggggaatgttgatgtc
GAD2-3'UTR2-F	aggaacctcaaactgggtatc	GAD2-3'UTR2-R	tttccctacaaaagacccaaag
YME1L1-EX1-F	gcgctgctccttgtaaatg	YME1L1-EX1-R	cctgatccactcctcaatcc
YME1L1-EX2-F	agttttcgtctcaccagcag	YME1L1-EX2-R	ggagggacaatatccttaatcc
YME1L1-EX3-F	cagggtttggtgaaactgc	YME1L1-EX3-R	agcctggtgaagaaagcaag
YME1L1-EX4-F	ggacagtagtggtttgggtttg	YME1L1-EX4-R	cagcccacaataaacgctac
YME1L1-EX5-F	gagaatggcgtgaacctg	YME1L1-EX5-R	ccaaatactctttaggtcctgtg
YME1L1-EX6-F	tgtcaggcaaattagatagactagg	YME1L1-EX6-R	tggtgcccagaagagttagag
YME1L1-EX7-F	tgcaggaatgggatctagtg	YME1L1-EX7-R	gaaagggaaaccatcagcag
YME1L1-EX8-F	tgagaaagaatgccagattcc	YME1L1-EX8-R	tctgggaaactaaccactactagg
YME1L1-EX9-F	tttctaggctttgctttcttgc	YME1L1-EX9-R	cactgtagaatcaaatatcgacctc
YME1L1-EX10-F	aaggggccatttttgtatgc	YME1L1-EX10-R	aaagggtatgaggagcattctatg
YME1L1-EX11-F	gcagagacaagtacggtttgc	YME1L1-EX11-R	agcctccaacttcatcttgc
YME1L1-EX12-F	atgaatgcccaactgttttg	YME1L1-EX12-R	gggactactctgtttcctatacacatc
YME1L1-EX13-F	ctaggccacaaagcgagtcc	YME1L1-EX13-R	tccccattctccctaacctc
YME1L1-EX14-F	aaaactccatctgcgctctg	YME1L1-EX14-R	acctttctttagactcagggttc
YME1L1-EX15-F	gcacaatcttggcttactgc	YME1L1-EX15-R	cttgaaaacccagacactgaag
YME1L1-EX16-F	cctcttccccatagttttaatcc	YME1L1-EX16-R	agaattaggcagggcatgg
YME1L1-EX17+18-F	ttgggattatagccatgcac	YME1L1-EX17+18-R	aatcagaaccagctctcaaatc
YME1L1-EX19-F	cgtttctccgtgtttccag	YME1L1-EX19-R	actgttcttagcccgtgtcc
YME1L1-EX20-F	tgccttctcaggtgtttattctc	YME1L1-EX20-R	gagctgttttcaatcctgatagttc
RAB18-EXON1-F	ctgccgaaataaaacagcttg	RAB18-EXON1-R	agagcgacccagagaacttg
RAB18-EXON2-F	tgggggattgtgtaggagag	RAB18-EXON2-R	ttagtagcagtccctggcatac
RAB18-EXON3-F	gaagttggggtgtgaattgg	RAB18-EXON3-R	cctcctccccttaaaaattgtc
RAB18-EXON4-F	atcctccctactgtcttgctg	RAB18-EXON4-R	gcagttgctcagaatttgttg
RAB18-EXON5-F	cttgtcagtaagcgaacacatc	RAB18-EXON5-R	gcccagtttctgtaatactccaag
RAB18-EXON6+7-F	gctgttgttttgctttaatgg	RAB18-EXON6+7-R	cacaaggcaactaggaacaatac
RAB18-EXON8-F	gggaaaaattgctgatttgg	RAB18-EXON8-R	cgtggtgctaggagttctaaatac
MPP7-5'UTR1-F	tatcagcaggcccctcttac	MPP7-5'UTR1-R	agtttttcctccacctgtcg
MPP7-5'UTR2-F	agaacattcccagatccactg	MPP7-5'UTR2-R	cggtccacattcttaaccattag
MPP7-5'UTR3-F	tggtatcgggagacttttcg	MPP7-5'UTR3-R	aatggggcaaccacttagg
MPP7-5'UTR4-F	gtgcatacatttagagaacaacagg	MPP7-5'UTR4-R	tcaaggaatccaagatttgc
MPP7-EX1-F	ttgtcaagatcagaaactctcacag	MPP7-EX1-R	acccagagaaagaggcaatg
MPP7-EX2-F	ccctttggttctttggagaag	MPP7-EX2-R	cgtcggtgaatggaacaag
MPP7-EX3-F	gttgtaaacactgggccatag	MPP7-EX3-R	cgaaagcagaacaagagagtagc
MPP7-EX4-F	ggatctggactaccgaaggtc	MPP7-EX4-R	gagcttcccaataagagtattttgc
MPP7-EX5-F	ggtcaggaggcgtggaac	MPP7-EX5-R	cactggtaacctatcccttgg
MPP7-EX6-F	ggcatgatgtctatttttcctg	MPP7-EX6-R	aggctctcacatttcaacacag
MPP7-EX7-F	ggcatagaggctgaaacctg	MPP7-EX7-R	tgcaaatgaaaacactgagg
MPP7-EX8+9-F	gtttcctccgtttagtctatctaagag	MPP7-EX8+9-R	cagaaatccaaacccacagg
MPP7-EX10-F	atttgaaagtgggcgatttg	MPP7-EX10-R	ctgagaactgatgatcccaaag
MPP7-EX11A-F	cctttagcctatccctttgg	MPP7-EX11A-R	gccagtgaaaagtcatcacg

MPP7-EX11B-F	tgcttgtccttgctacattaagtc	MPP7-EX11B-R	ggtattcaagaggcagaaaaag
MPP7-EX12-F	attaaccagccgtggtgatg	MPP7-EX12-R	tccccacttgacaaaggttc
MPP7-EX13-F	ttgaggaaaaggtgaggatttc	MPP7-EX13-R	aactcacatgccccataaatac
MPP7-EX14-F	ggttagagagccttcaaaatagc	MPP7-EX14-R	tcattctgcaaagcagtgtc
MPP7-EX15-F	gggggtgatggaaaacattag	MPP7-EX15-R	gagcttcctccacttctctttc
MPP7-EX16-F	cctgtcatcctggaaactcc	MPP7-EX16-R	aaccctcctcctttgctttc
MPP7-EX17-F	attccccagcttcttttcag	MPP7-EX17-R	atcatgcctggctaatttcc
MPP7-EX18-F	tctggctgaggggaaagac	MPP7-EX18-R	gcgctgaactcccatacatac
5q31.3-q35.3			
GRIA1-EX-1F	ttcttttcccgtgctcagtt	GRIA1-EX-1R	cacagctttagggagccaag
GRIA1-EX-2F	agtttggcatggggagaag	GRIA1-EX-2R	caaagacgaatggcaactga
GRIA1-EX-3F	gatgggttggatggatgaac	GRIA1-EX-3R	ttttcagggaagcagtgagg
GRIA1-EX-4F	aagagtgggtttgggggtag	GRIA1-EX-4R	tcaaggctcccattgtttgt
GRIA1-EX-5F	gggtagggcacattgtaagc	GRIA1-EX-5R	cacatcagccagggacaac
GRIA1-EX-6F	cctgtggttttggaacaggt	GRIA1-EX-6R	tccctgaatttgttcgcttt
GRIA1-EX-7F	ccagaaaacacattcccaaa	GRIA1-EX-7R	caccaacaccaacaatgcag
GRIA1-EX-8F	tggaacactcttgggttctg	GRIA1-EX-8R	cagaaaaggtggactcacagg
GRIA1-EX-9F	tccttacttgcctgtctctgg	GRIA1-EX-9R	aaaggccagctttcagtcct
GRIA1-EX-10F	gctagagagccttccccagt	GRIA1-EX-10R	tcagcagcaacaacactcaa
GRIA1-EX-11aF	aggggttggacttcaaaggt	GRIA1-EX-11aR	tcactgtgccattcataggg
GRIA1-EX-11bF	ctccttccttgatcctttgg	GRIA1-EX-11bR	cgtgcaatggacagttgat
GRIA1-EX-12F	atgagaagggtgcaataggg	GRIA1-EX-12R	gggggtggtagaaagcttaga
GRIA1-EX-13F	tgagctaatctcgctcatgaat	GRIA1-EX-13R	ggcccataggtgttgaagaa
ADAM19-EX-1F	agcctcccctcctccatc	ADAM19-EX-1R	gcagaacgtgggaacaaag
ADAM19-EX-2F	gcagctcacagaaagtgcat	ADAM19-EX-2R	agagtgggccagaaacagaa
ADAM19-EX-3F	agcgctgtgctagatcctgt	ADAM19-EX-3R	ctgcttccttccagaacagc
ADAM19-EX-4F	ggatagtcctgcctggttca	ADAM19-EX-4R	tcccttcagaggagagacca
ADAM19-EX-5F	ggaaagctctttgcaggttg	ADAM19-EX-5R	tcaatggactccttcccaag
ADAM19-EX-6F	gaaccaatgatcaaaacccact	ADAM19-EX-6R	cctggattcaaagtaagcaacc
ADAM19-EX-7F	agtgccgaatgacaaaatca	ADAM19-EX-7R	gagaaagatgaatgggcaaca
ADAM19-EX-8F	ggtcctaccctccaaagctc	ADAM19-EX-8R	tgctcattccctttcagaaga
ADAM19-EX-9F	tcaagctccaccagaatgaa	ADAM19-EX-9R	ctctggaccagggagtcaga
ADAM19-EX-10F	cgtgatggccacttactcc	ADAM19-EX-10R	attgcccttcctgatgtcaa
ADAM19-EX-11F	gtggaattggtggagacctg	ADAM19-EX-11R	ttggccccatcagaagttta
ADAM19-EX-12F	aggagtgacagcacgagtga	ADAM19-EX-12R	tttttcatggggtattggaca
ADAM19-EX-13F	ttgctagagagctggggttc	ADAM19-EX-13R	ccttttccgctctgtgtttc
ADAM19-EX-14F	tggtttgggctaatgaggtc	ADAM19-EX-14R	cctgagtcaaggggatgaaa
ADAM19-EX-15F	tggagggaacaggaagaaga	ADAM19-EX-15R	gggccgagaatagatgactg
ADAM19-EX-16F	gggttccaaagcacatgact	ADAM19-EX-16R	ccaccaaggctcagaggtt
ADAM19-EX-17-18F	gccatgggtcctcatatgtt	ADAM19-EX-17-18R	caatgtggatgctctgcaac
ADAM19-EX-19F	tgcaaatcaaatgcctgaag	ADAM19-EX-19R	cagtcatagcaggggagagg
ADAM19-EX-20F	ccagagcagacttgggaaag	ADAM19-EX-20R	gggcaagactccatctcaaa
ADAM19-EX-21F	tgggctctggtttatctcaaa	ADAM19-EX-21R	agtttcaccttccccacctt
ADAM19-EX-22F	ccagcttgctctcctgactt	ADAM19-EX-22R	gggcaccaagaaacatgaat
ADAM19-EX-23F	gttctcagtgggtggtccat	ADAM19-EX-23R	agagcagaagcaatgggaag
PTTG1- 5'UTR1 + 5'UTR2-F	cctggctgcttaggtccttt	PTTG1- 5'UTR1 + 5'UTR2-R	gctcggagccatctaagc
PTTG1-5'UTR3 + Exon 1-F	cgtgactgttccgctgttta	PTTG1-5'UTR3 + Exon 1-R	gactccaattgcccaaaaag
PTTG1-Exon 2-F	gggggtgagaggtcaagttt	PTTG1-Exon 2-R	ctcagagggaagaaagctacg
PTTG1-Exon 3-F	tgctgacaggtgctggtact	PTTG1-Exon 3-R	gtccaaagcaacccagattc

PTTG1-Exon 4-F	caacccagtggaaaaaggtg	PTTG1-Exon 4-R	acctcccgcacaactcttc
PTTG1-Exon 5 + 3'UTR-F	aaatgtgtcaggaggggatg	PTTG1-Exon 5 + 3'UTR-R	gactctttgttaaacacctcagc
GABRB2-EX-1F	caagcttgctgcctcctc	GABRB2-EX-1R	tgttcaggggctgtacctct
GABRB2-EX-2F	tattctggatgctgctgtgg	GABRB2-EX-2R	gaagcaggcatagcgttttc
GABRB2-EX-3F	gtagggatccgatgctgaaa	GABRB2-EX-3R	aagagagcgcgcacaaag
GABRB2-EX-4F	agaagggttttgggcagagt	GABRB2-EX-4R	cccaatagctggctcatttc
GABRB2-EX-5F	atgtttgcttatcccccaga	GABRB2-EX-5R	gcacaatatttccatccaatga
GABRB2-EX-6F	tttccagctccagccttaga	GABRB2-EX-6R	tctggacatgagccaggact
GABRB2-EX-7F	ccaaacacataacagcctttagc	GABRB2-EX-7R	gcttccctgggacagaactc
GABRB2-EX-8F	ttgcagaagaaagtgtccttca	GABRB2-EX-8R	gataaggaaaacgcgtgcat
GABRB2-EX-9F	tgcaatagtaaaggcccaat	GABRB2-EX-9R	gcttggtttcctcatttgct
GABRB2-EX-10F	ttcccattctctgttagccttt	GABRB2-EX-10R	atctgcaaagatcccacaca
GABRB2-EX-11aF	aatcccttaagtgccgatca	GABRB2-EX-11aR	tgggaggccatgttttagtt
GABRB2-EX-11bF	tccctgacttgactgatgtga	GABRB2-EX-11bR	cccctctgagtaggctgcat
GABRA1-EX-1F	gaatgattccggaaatggag	GABRA1-EX-1R	tttcgcactttctcaggttg
GABRA1-EX-2F	gcatgaagtcaccgcctatt	GABRA1-EX-2R	cagctaggaaaagccacgtt
GABRA1-EX-3F	cagtcagccctggtggttat	GABRA1-EX-3R	gacaattttcttcccccgata
GABRA1-EX-4F	gagaaaggatttatttggaaagag	GABRA1-EX-4R	tgtcaccatttttattatttgacca
GABRA1-EX-5F	caatttcccatttgggtgtc	GABRA1-EX-5R	gccaagacttttcaggttgc
GABRA1-EX-6F	gcaaaaattatgcactgtctgc	GABRA1-EX-6R	cagcctgatgattgcctaca
GABRA1-EX-7F	gctcagcataaccctgcaat	GABRA1-EX-7R	tctacgcgtttttctcacaga
GABRA1-EX-8F	tgccctctggaaccatgata	GABRA1-EX-8R	gctcaactgatgcccaattt
GABRA1-EX-9F	ttcccagacctttggtactcat	GABRA1-EX-9R	gagagtggcaattccttgaaa
GABRA1-EX-10F	tgccattccatgaatcacag	GABRA1-EX-10R	tcatggcacttaattgtttacg
GABRA1-EX-11aF	aaaataagggccaccttgct	GABRA1-EX-11aR	actttgctctttcgcttgct
GABRA1-EX-11bF	aaaataagggccaccttgct	GABRA1-EX-11bR	tgggaattactgcgttgaga
GABRA1-EX-11cF	cagagagcctcagctaaaagc	GABRA1-EX-11cR	gggatcttcagctttggaaac
GABRA1-EX-11dF	tgtgaaggtgtttcaaagggta	GABRA1-EX-11dR	ttgggctcatgtgctatctg
GABRA1-EX-11eF	tcactttcatctgagcttttacca	GABRA1-EX-11eR	ggaaatgtgtggaatgacttga
GABRA1-EX-11fF	tgggtgtgaagtccacttatg	GABRA1-EX-11fR	gctcaagacttgggtttgga
GABRA1-EX-11gF	ccacaagtcacggtctaaaca	GABRA1-EX-11gR	gccatgagaagtctggaagg
GABRA6-EX-1F	tgacttgaggcaaacaagga	GABRA6-EX-1R	caccacttctggcctctttc
GABRA6-EX-2F	gaagtggtgttgcatgtatttg	GABRA6-EX-2R	gactcttctgggtcttcaaagg
GABRA6-EX-3F	tggtggtagagggctgagat	GABRA6-EX-3R	tcatgaatctggagccaagc
GABRA6-EX-4F	ttttgagattgggtgtataggc	GABRA6-EX-4R	cattgaaatcaacggttctga
GABRA6-EX-5F	ccatgaggtgaggtttctcc	GABRA6-EX-5R	tcttggtgactaaagcctcaa
GABRA6-EX-6F	ttgaggctttagtcaccaaga	GABRA6-EX-6R	gcttagctgagaatggctaagg
GABRA6-EX-7F	tcaactgcattctttttgcag	GABRA6-EX-7R	tggattgaaaaagtgcacca
GABRA6-EX-8F	tcatttgattttgccatagga	GABRA6-EX-8R	aactggccttgctcttgaaa
GABRA6-EX-9F	tttgtcaatggtgaaagagtgaa	GABRA6-EX-9R	acggaagctgcgttacagat
GABRG2-EX-1F	gggagaacgcgtaagtgtga	GABRG2-EX-1R	aattgtaaagccgcacatcc
GABRG2-EX-2F	ttttccactggtggtctgtg	GABRG2-EX-2R	ctcccactcataggcctgaa
GABRG2-EX-3F	gcttggtgcatgtgcatact	GABRG2-EX-3R	ttttgctggcatgtcttttg
GABRG2-EX-4F	gccaagatagctttgcttca	GABRG2-EX-4R	gatagcatgccaaccctgat
GABRG2-EX-5F	tcattggggatcactctgtg	GABRG2-EX-5R	gaatcaccctaatcggagca
GABRG2-EX-6F	cctttggtccaagatcctca	GABRG2-EX-6R	tttgaatccaattttcccatatt
GABRG2-EX-7F	cgggcaaaaatagttttcaa	GABRG2-EX-7R	tgcaggctaaatttaaagcaga
GABRG2-EX-8F	gaaactgcccatcagaagtca	GABRG2-EX-8R	ctgagccttagcctgcagat
GABRG2-EX-9F	cccaagctcagaactctcctt	GABRG2-EX-9R	gcttttgggcttggtgtaag

GABRG2-EX-10F	tggtgacattgtggaaaaaca	GABRG2-EX-10R	gcaggtttgaaacacagcaa
KCNIP1-EX-1aF	ccgggtctgtctcttaggg	KCNIP1-EX-1aR	gaggcttctggcttccttct
KCNIP1-EX-1bF	cttcagggcaccgtcctc	KCNIP1-EX-1bR	ctccctcctctgggtgttc
KCNIP1-EX-2F	cactcttttgacagcccaga	KCNIP1-EX-2R	ttatgagctcccttggcatt
KCNIP1-EX-3F	cttgggagaagtggaaatgc	KCNIP1-EX-3R	agcaaatggcaaggaagaga
KCNIP1-EX-4F	tcctctctttcttgtcgaagc	KCNIP1-EX-4R	ccccaatgccttgagattta
KCNIP1-EX-5F	cccaaagacacagcttcaca	KCNIP1-EX-5R	tgcctcattcagacagtgct
KCNIP1-EX-6F	tccccttctgcttgcttcta	KCNIP1-EX-6R	acagggttggacacaaatca
KCNIP1-EX-7F	gagctccagcaagaaagcag	KCNIP1-EX-7R	tgtgcaggtttctcatcacc
KCNIP1-EX-8aF	ggaaacccatgcttgacatt	KCNIP1-EX-8aR	gggcacgtaagcagcatatt
KCNIP1-EX-8bF	atggaaggtccctctgctta	KCNIP1-EX-8bR	tccctttcacagcttacctc

Table A4.7. Primer sequences for the genes sequenced for segregation analysis in HWE307

Forward primer (Gene name-Exon number)	**Primer sequence (5'→ 3')**	**Length**	**Reverse primer (Gene name-Exon number)**	**Primer sequence (5'→ 3')**	**Length**
C1QL3-5'UTR-F	cgtttgtcagcctcctcttc	20	*C1QL3*-5'UTR-R	ccaccaccacaactcgaata	20
C10orf67-1F	acgcctcacattctgacctc	20	*C10orf67*-1R	ctgctcgcttcctcctctac	20
ARHGAP21-26F	gcaccttcgttaactaaagtg	21	*ARHGAP21*-26R	cggtttacagctgaaagttc	22
JAKMIP2-5'UTR-F	tggtcgcttatcattggtca	20	*JAKMIP2*-5'UTR-R	caagttcaggccagaaatgtat	22
FAXDC2-7F	gggagaaagaggaatcagca	20	*FAXDC2*-7R	gccggtgtggtctaagttct	20
CNOT8-5'UTR-F	tgctgggcagaagtttcaat	20	*CNOT8*-5'UTR-R	cagagcaagaccccaacttt	20
C1QTNF2-3F	ccaagcactccttctccaag	20	*C1QTNF2*-3R	agcgtgatgtcgtaggtgaa	20
DDC-5'UTR-F	ggagggatgctgctcagtaa	20	*DDC*-5'UTR-R	cgacctaggttggtgcctat	20
WBSCR27-4F	gcaggtgtgaaccatgactg	20	*WBSCR27*-4R	ggaggatcagcaagggaaag	20
DTX2-3F	tctgcagggctactgatgtg	20	*DTX2*-3R	acccagatatgtgcgagagg	20

Appendix IV

Protocols

A. Genomic DNA isolation

Genomic DNA was extracted from blood using the phenol-chloroform extraction method (Sambrook and Russel 2001). Here, blood cells are lysed and treated with proteinase K, followed by phenol extraction and resuspension in TE buffer.

The genomic DNA isolation procedure is as follows:

1. 10ml of blood was diluted to 40ml with cold NKM buffer. The sample was vortexed and centrifuged at 3500xg at 4^{o}C for 30 minutes.
2. About 10ml of the supernatant was retained with the pellet and it was vortexed to dissolve any clumps.
3. Resuspension buffer was added to the dissolved pellet up to a volume of 40ml and centrifuged for 30 minutes at 3500xg at 4^{o}C.
4. 0.5ml of 10X TEN solution, 0.25ml of 2mg/ml proteinase K and 0.5ml of 10% SDS was added to ~4ml of the retained supernatant. The supernatant was mixed gently and incubated overnight at 37^{o}C to allow digestion.
5. DNA was extracted twice, once with adding equal volume of phenol-chloroform-isoamyl alcohol in the ratio of 25:24:1 and later with equal volume of chloroform-isoamyl alcohol in the ratio of 24:1.
6. Genomic DNA was precipitated using 2 volumes of ice-cold isopropyl alcohol and 5M NaCl (0.4M NaCl as final concentration).
7. Pellet was washed twice with 70% ethanol.
8. Pellet was air dried and the DNA was resuspended in 200µl of TE buffer.
9. Genomic DNA was stored at -20^{o}C.

B. Calcium phosphate mediated transfection of eukaryotic cells with plasmid DNA

1. Twenty four hours before transfection, exponentially growing cells were seeded at a density of ~2-3 X 10^5 cells/ml in culture dishes or coverslips, and grown in a humidified incubator of 5% CO_2 at 37°C.
2. Cells were grown to 40% confluence at the time of transfection.

3. One hour before transfection, the growth medium was replaced with transfection medium (antibiotic- and serum- free DMEM).
4. The transfection mix of 200µl comprised equal volumes each of $CaCl_2$-DNA solution and 2x HBS, and was prepared as follows (per coverslip/ 35mm dish):
    i. A 100µl solution of deionised sterile water, plasmid DNA and $CaCl_2$ was prepared. Volume of $CaCl_2$ was adjusted such that the final concentration in the culture dish after transfection was 12.5mM.
    ii. This solution was added slowly to 100µl of 2x HBS and mixed by pipetting. This transfection mix was added on the cells.
5. The culture dish was gently rocked to uniformly mix the solution in the medium.
6. Transfected cells were incubated in a humidified incubator of 5% CO_2 at 37°C.
7. Six hours after transfection, the medium was replaced with complete growth medium (10% heat-inactivated fetal bovine serum, 2mM L-glutamine, 100U/ml penicillin and 0.1mg/ml streptomycin).
8. Cells were grown for 48 hours from time of transfection, and processed accordingly.

C. Transformation of bacterial cells with ZGRF1 plasmids

1. A microfuge tube containing XL10-Gold competent cells was allowed to thaw in ice for 10 minutes, and was gently tapped to resuspend the cells.
2. 100-200ng of plasmid DNA was added to the cells.
3. The tube was incubated in ice for 1 hour with gentle tapping every 5 minutes during this time.
4. The cells were given a heat shock for 90 seconds at 42°C and immediately placed back in ice and incubated for 2 minutes.
5. 1ml of LB (Luria broth) medium was added to the cells. The cells were revived for 1 hour at 30°C (shaker incubator- speed maintained at 230rpm).
6. 100µl of the cells were plated on LB agar and incubated at 30°C for 24-36 hours, or till colonies appeared.

Appendix V

Reagents and solutions

NKM buffer

28ml of 5M NaCl (0.14M final)
1.5ml of 1M $MgCl_2$ (1.5mM final)
30ml of 1M KCl (30mM final)
Added H_2O to make up the volume to 1 litre
Filter sterilized and stored in 500ml sterile bottles at 4°C

TAE, 50X

242g Tris base
57.1ml glacial acetic acid
100ml 0.5M EDTA, pH8.0
Added H_2O to 1 litre
Filter sterilized and stored at room temperature

Resuspension buffer

20ml of 5M NaCl (0.10M final)
10ml of 1M Tris-Cl, pH7.5 (10mM final)
1.5ml of 1M $MgCl_2$ (1.5mM final)
Added H_2O to make up the volume to 1 litre
Filter sterilized and stored in 500ml sterile bottles at 4°C

Gel-loading buffer, 6X

0.25% bromophenol blue
0.25% xylene-cyanol FF
30% glycerol in water
Stored at 4°C

Proteinase K, 2mg/ml

49ml H_2O
0.5ml of 1M $CaCl_2$ (10mM final)
0.5ml of 1M Tris-Cl, pH7.5 (10mM final)
100mg proteinase K (2mg/ml final)
Stored in 1ml aliquots at −20°C

HEPES-buffered saline (HBS)

280mM NaCl
1.5mM Na_2HPO_4
50mM HEPES

TEN solution, 10X

9ml 1M Tris-Cl, pH7.5 (0.21M final)
24ml 0.5M EDTA, pH8.0 (0.29M final)
9ml 5M NaCl (1.07M final)
Filter sterilized and stored at room temperature

2.5M $CaCl_2$

11.025g of $CaCl_2.2H_2O$
30ml deionised H_2O

PBS buffer, 5X

40g NaCl
7.2g Na_2HPO_4
1.2g KH_2PO_4
1g KCl
Added 800ml of H_2O
Adjusted pH to 7.4 with HCl
Made volume up to 1 litre with H_2O
Sterilized by autoclaving and stored at 4°C

TE buffer

10mM Tris-Cl (pH8.0)
1mM EDTA (pH8.0)

References

Adamson B, Smogorzewska A, Sigoillot FD, King RW, Elledge SJ (2012) A genome-wide homologous recombination screen identifies the RNA-binding protein RBMX as a component of the DNA-damage response. Nat Cell Biol 14:318–328

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR (2010) A method and server for predicting damaging missense mutations. Nat Methods 7:248–249

Allen IM (1945) Observations on cases of reflex epilepsy. New Zealand Med. J. 44:135-142

Allen AS, Berkovic SF, Cossette P, Delanty N, Dlugos D, Eichler EE, Epstein MP, Glauser T, Goldstein DB, Han Y, Heinzen EL, Hitomi Y, Howell KB, Johnson MR, Kuzniecky R, Lowenstein DH, Lu YF, et al, Epi4K Consortium, Epilepsy Phenome/Genome Project (2013) De novo mutations in epileptic encephalopathies. Nature 501:217–221

Ames FR, Saffer D (1983) The sunflower syndrome. A new look at "self-induced" photosensitive epilepsy. J Neurol Sci 59:1–11

Applequist SE, Selg M, Raman C, Jäck HM (1997) Cloning and characterization of HUPF1, a human homolog of the Saccharomyces cerevisiae nonsense mRNA-reducing UPF1 protein. Nucleic Acids Res 25:814–821

Araujo J, Breuer P, Dieringer S, Krauss S, Dorn S, Zimmermann K, Pfeifer A, Klockgether T, Wuellner U, Evert BO (2011) FOXO4-dependent upregulation of superoxide dismutase-2 in response to oxidative stress is impaired in spinocerebellar ataxia type 3. Hum Mol Genet 20:2928–2941

Aravind L (2001) The WWE domain: a common interaction module in protein ubiquitination and ADP ribosylation. Trends Biochem Sci 26:273–275

Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res 38:W529–W533

Audenaert D, Van Broeckhoven C, De Jonghe P (2006) Genes and loci involved in febrile seizures and related epilepsy syndromes. Human mutation 27:391-401

Audenaert D, Schwartz E, Claeys KG, Claes L, Deprez L, Suls A, Van Dyck T, Lagae L, Van Broeckhoven C, Macdonald RL, De Jonghe P (2006) A novel GABRG2 mutation associated with febrile seizures. Neurology 67:687–690

Audenaert D, Claes L, Claeys KG, Deprez L, Van Dyck T, Goossens D, Del-Favero J, Paesschen, Van Broeckhoven C, De Jonghe P (2005) A novel susceptibility locus at 2p24 for generalised epilepsy with febrile seizures plus. J Med Genet 42:947–952

Auvin S, Lamblin MD, Pandit F, Bastos M, Derambure P, Vallée L (2006) Hot water epilepsy occurring at temperature below the core temperature. Brain Dev 28:265–268

Avery P, Vicente-Crespo M, Francis D, Nashchekina O, Alonso CR, Palacios IM (2011) Drosophila Upf1 and Upf2 loss of function inhibits cell growth and causes animal death in a Upf3-independent manner. RNA 17:624–638

Baboval T, Smith FI (2002) Comparison of human and mouse Fuc-TX and Fuc-TXI genes, and expression studies in the mouse. Mamm Genome 13:538–541

Bac P, Maurois P, Dupont C, Pages N, Stables JP, Gressens P, Evrard P, Vamecq J (1998) Magnesium deficiency-dependent audiogenic seizures (MDDASs) in adult mice: a nutritional model for discriminatory

screening of anticonvulsant drugs and original assessment of neuroprotection properties. J Neurosci 18:4363–4373

Baraban SC, Dinday MT, Hortopan GA (2013) Drug screening in Scn1a zebrafish mutant identifies clemizole as a potential Dravet syndrome treatment. Nat Commun 4:2410

Baulac S, Gourfinkel-An I, Couarch P, Depienne C, Kaminska A, Dulac O, Baulac M, LeGuern E, Nabbout R (2008) A novel locus for generalized epilepsy with febrile seizures plus in French families. Arch Neurol 65:943–951

Baulac S, Huberfeld G, Gourfinkel-An I, Mitropoulou G, Beranger A, Prud'homme J-F, Baulac M, Brice A, Bruzzone R, LeGuern E (2001) First genetic evidence of GABAA receptor dysfunction in epilepsy: a mutation in the γ2-subunit gene. Nat Genet 28:46–48

Bebek N, Baykan B, Gürses C, Emir Ö, Gökyiğit A (2006) Self-induction behavior in patients with photosensitive and hot water epilepsy: A comparative study from a tertiary epilepsy center in Turkey. Epilepsy Behav 9:317–326

Bebek N, Gürses C, Gokyigit A, Baykan B, Ozkara C, Dervent A (2001) Hot water epilepsy: Clinical and electrophysiologic findings based on 21 cases. Epilepsia 42:1180–1184

Becker DJ, Lowe JB (2003) Fucose: biosynthesis and biological function in mammals. Glycobiology 13:41R–53R

Belhedi N, Bena F, Mrabet A, Guipponi M, Souissi C, Mrabet H, Elgaaied A, Malafosse A, Salzmann A (2013) A new locus on chromosome 22q13.31 linked to recessive genetic epilepsy with febrile seizures plus (GEFS+) in a Tunisian consanguineous family. BMC Genet 14:93

Bell MR, Belarde JA, Johnson HF, Aizenman CD (2011) A neuroprotective role for polyamines in a Xenopus tadpole model of epilepsy. Nat Neurosci 14:505–512

Bennett CL, Huynh HM, Chance PF, Glass IA, Gospe SM (2005) Genetic heterogeneity for autosomal recessive pyridoxine-dependent seizures. Neurogenetics 6:143–149

Bhattacharya A, Czaplinski K, Trifillis P, He F, Jacobson A, Peltz SW (2000) Characterization of the biochemical properties of the human Upf1 gene product that is involved in nonsense-mediated mRNA decay. RNA 6:1226–1235

Bigarella CL, Borges L, Costa FF, Saad STO (2009) ARHGAP21 modulates FAK activity and impairs glioblastoma cell migration. Biochim Biophys Acta - Mol Cell Res 1793:806–816

Binnie CD (1988) Self-induction of seizures: the ultimate non-compliance. Epilepsy Res Suppl 1:153–158

Blume WT, Lüders HO, Mizrahi E, Tassinari C, van Emde Boas W, Engel J (2001) Glossary of descriptive terminology for ictal semiology: report of the ILAE task force on classification and terminology. Epilepsia 42:1212–1218

Boillot M, Baulac S (2016) Genetic models of focal epilepsies. J Neurosci Methods 260:132–143

Bond U (2006) Stressed out! Effects of environmental stress on mRNA metabolism. FEMS Yeast Res 6:160–170

Boratyn GM, Schäffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL (2012) Domain enhanced lookup time accelerated BLAST. Biol Direct 7:12

Brennan TJ, Seeley WW, Kilgard M, Schreiner CE, Tecott LH (1997) Sound-induced seizures in serotonin 5-HT2c receptor mutant mice. Nat Genet 16:387–390

Brodtkorb E, Gu W, Nakken KO, Fischer C, Steinlein OK (2002) Familial Temporal Lobe Epilepsy with Aphasic Seizures and Linkage to Chromosome 10q22-q24. Epilepsia 43:228-235

Brogna S, Ramanathan P, Wen J (2008) UPF1 P-body localization. Biochem Soc Trans 36:698–700

Byrns CN, Pitts MW, Gilman CA, Hashimoto AC, Berry MJ (2014) Mice lacking selenoprotein P and selenocysteine lyase exhibit severe neurological dysfunction, neurodegeneration, and audiogenic seizures. J Biol Chem 289:9662–74

Carvill GL, Weckhuysen S, McMahon JM, Hartmann C, Moller RS, Hjalgrim H, Cook J, Geraghty E, O'Roak BJ, Petrou S, Clarke A, Gill D, Sadleir LG, Muhle H, von Spiczak S, Nikanorova M, Hodgson BL, Gazina EV, Suls A, et al (2014) GABRA1 and STXBP1: Novel genetic causes of Dravet syndrome. Neurology 82:1245–1253

Celniker G, Nimrod G, Ashkenazy H, Glaser F, Martz E, Mayrose I, Pupko T, Ben-Tal N (2013) ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function. Isr J Chem 53:199–206

Ceulemans B, Koninckx M, Garmyn K, Wojciechowski M (2008) Hot water epilepsy: a video case report of a Caucasian toddler. Epileptic Disord 10:45–48

Chabrol E, Navarro V, Provenzano G, Cohen I, Dinocourt C, Rivaud-Péchoux S, Fricker D, Baulac M, Miles R, LeGuern E, Baulac S (2010) Electroclinical characterization of epileptic seizures in leucine-rich, glioma-inactivated 1-deficient mice. Brain 133:2749–2762

Chakrabarti S, Jayachandran U, Bonneau F, Fiorini F, Basquin C, Domcke S, Le Hir H, Conti E (2011) Molecular Mechanisms for the RNA-Dependent ATPase Activity of Upf1 and Its Regulation by Upf2. Mol Cell 41:693–703

Chen L, Toth M (2001) Fragile X mice develop sensory hyperreactivity to auditory stimuli. Neuroscience 103:1043–1050

Cheng Z, Muhlrad D, Lim MK, Parker R, Song H (2007) Structural and functional insights into the human Upf1 helicase core. EMBO J 26:253–64

Chi P, Van Komen S, Sehorn MG, Sigurdsson S, Sung P (2006) Roles of ATP binding and ATP hydrolysis in human Rad51 recombinase function. DNA Repair 5:381–391

Choi Y, Chan AP (2015) PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 31:2745–2747

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly 6:80–92

Clark MJ, Chen R, Lam HYK, Karczewski KJ, Chen R, Euskirchen G, Butte AJ, Snyder M (2011) Performance comparison of exome DNA sequencing technologies. Nat Biotechnol Suppl 29:908–914

Cooper GM, Shendure J (2011) Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. Nat Rev Genet 12:628–640

Cowart LA, Gandy JL, Tholanikunnel B, Hannun YA (2010) Sphingolipids mediate formation of mRNA processing bodies during the heat-stress response of Saccharomyces cerevisiae. Biochem J 431:31–38

Crowe SL, Kondratyev AD (2010) DNA Damage and Repair in the Brain: Implications for Seizure-Induced Neuronal Injury, Endangerment, and Neuroprotection. In: Acute Neuronal Injury. Springer US, Boston, MA, pp 243–275

Dai XH, Chen WW, Wang X, Zhu QH, Li C, Li L, Liu MG, Wang QK, Liu JY (2008) A novel genetic locus for familial febrile seizures and epilepsy on chromosome 3q26.2–q26.33. Hum Genet 124:423–429

Darby CE, Korte RA, Binnie CD, Wilkins AJ (1980) The Self-induction of Epileptic Seizures by Eye Closure. Epilepsia 21:31–42

Dazzo E, Fanciulli M, Serioli E, Minervini G, Pulitano P, Binelli S, Di Bonaventura C, Luisi C, Pasini E, Striano S, Striano P, Coppola G, Chiavegato A, Radovic S, Spadotto A, Uzzau S, La Neve A, Giallonardo AT, Mecarelli O, et al (2015) Heterozygous Reelin Mutations Cause Autosomal-Dominant Lateral Temporal Epilepsy. Am J Hum Genet 96:992–1000

De Jonghe P (2011) Molecular genetics of Dravet syndrome. Dev Med Child Neurol 53:7–10

De Kovel CGF, Pinto D, Tauer U, Lorenz S, Muhle H, Leu C, Neubauer BA, Hempelmann A, Callenbach PMC, Scheffer IE, Berkovic SF, Rudolf G, Striano P, Siren A, Baykan B, Sander T, Lindhout D, Trenité DGK-N, Stephani U, et al (2010) Whole-genome linkage scan for epilepsy-related photosensitivity: A mega-analysis. Epilepsy Res 89:286–294

De Luca G, Di Giorgio RM, Macaione S, Calpona PR, Costantino S, Di Paola ED, De Sarro A, Ciliberto G, De Sarro G (2004) Susceptibility to audiogenic seizure and neurotransmitter amino acid levels in different brain areas of IL-6-deficient mice. Pharmacol Biochem Behav 78:75–81

de Palma L, Boniver C, Cassina M, Toldo I, Nosadini M, Clementi M, Sartori S (2012) Eating-induced epileptic spasms in a boy with MECP2 duplication syndrome: insights into pathogenesis of genetic epilepsies. Epileptic Disord 14:414–417

DePristo MA, Banks E, Poplin R, Garimella K V, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43:491–498

Desmet FO, Hamroun D, Lalande M, Collod-Bëroud G, Claustres M, Béroud C (2009) Human Splicing Finder: An online bioinformatics tool to predict splicing signals. Nucleic Acids Res 37:1–14

Dewey CM, Cenik B, Sephton CF, Dries DR, Mayer P, Good SK, Johnson BA, Herz J, Yu G (2011) TDP-43 Is Directed to Stress Granules by Sorbitol, a Novel Physiological Osmotic and Oxidative Stressor. Mol Cell Biol 31:1098–1108

Dhandapany PS, Sadayappan S, Xue Y, Powell GT, Rani DS, Nallari P, Rai TS, Khullar M, Soares P, Bahl A, Tharkan JM, Vaideeswar P, Rathinavel A, Narasimhan C, Ayapati DR, Ayub Q, Mehdi SQ, Oppenheimer S, Richards MB, et al (2009) A common MYBPC3 (cardiac myosin binding protein C) variant associated with cardiomyopathies in South Asia. Nat Genet 41:187–191

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

Doose H, Waltz S (1993) Photosensitivity - Genetics and Clinical Significance. Neuropediatrics 24:249–255

Doose H, Petersen B, Neubauer BA (2002) Occipital sharp waves in idiopathic partial epilepsies--clinical and genetic aspects. Epilepsy Res 48:121-130

Doretto MC, Burger RL, Mishra PK, Garcia-Cairasco N, Dailey JW, Jobe PC (1994) A microdialysis study of amino acid concentrations in the extracellular fluid of the substantia nigra of freely behaving GEPR-9s: relationship to seizure predisposition. Epilepsy Res 17:157–165

Dravet C, Bureau M, Guerrini R, Giraud N, Roger J (1992) Severe myoclonic epilepsy in infants. In: Roger J, Dravet C, Bureau M, Dreifuss FE, Perret A, Wolf P, editors. Epileptic syndromes in infancy, childhood and adolescence. 2nd ed. London: John Libbey, pp 75-88

Du W, Bautista JF, Yang H, Diez-Sampedro A, You S-A, Wang L, Kotagal P, Lüders HO, Shi J, Cui J, Richerson GB, Wang QK (2005) Calcium-sensitive potassium channelopathy in human epilepsy and paroxysmal movement disorder. Nat Genet 37:733–738

Dubois T, Paléotti O, Mironov AA, Fraisier V, Stradal TEB, De Matteis MA, Franco M, Chavrier P (2005) Golgi-localized GAP for Cdc42 functions downstream of ARF1 to control Arp2/3 complex and F-actin dynamics. Nat Cell Biol 7:353–364

Duncan JS, Panayiotopoulos CP (1995) Typical absences and related epileptic syndromes. London: Churchill Livingstone 213-220

Eilbeck K, Quinlan A, Yandell M (2017) Settling the score: variant prioritization and Mendelian disease. Nat Rev Genet 18:599–612

Engel J, International League Against Epilepsy (ILAE) (2001) A proposed diagnostic scheme for people with epileptic seizures and with epilepsy: report of the ILAE Task Force on Classification and Terminology. Epilepsia 42:796–803

Engel J, Pedley TA (2008) Epilepsy: a comprehensive textbook. Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia

Eom T, Zhang C, Wang H, Lay K, Fak J, Noebels JL, Darnell RB (2013) NOVA-dependent regulation of cryptic NMD exons controls synaptic protein levels after seizure. Elife 2:e00178

Erdem E, Topcu M, Renda Y, Ciger A, Varli K, Zileli T (1992) Hot Water Epilepsy. Clin Electroencephalogr 23:152–158

Ewart AK, Morris CA, Atkinson D, Jin W, Sternes K, Spallone P, Stock AD, Leppert M, Keating MT (1993) Hemizygosity at the elastin locus in a developmental disorder, Williams syndrome. Nat Genet 5:11–16

Faingold CL, Riaz A (1995) Ethanol withdrawal induces increased firing in inferior colliculus neurons associated with audiogenic seizure susceptibility. Exp Neurol 132:91–98

Ferrier D (1873) Experimental researches in cerebral physiology and pathology. Br Med J 1:457

Fisher RS, Cross JH, French JA, Higurashi N, Hirsch E, Jansen FE, Lagae L, Moshé SL, Peltola J, Roulet Perez E, Scheffer IE, Zuberi SM (2017) Operational classification of seizure types by the International League Against Epilepsy: Position Paper of the ILAE Commission for Classification and Terminology. Epilepsia 58:522–530

Fisher RS, Acevedo C, Arzimanoglou A, Bogacz A, Cross JH, Elger CE, Engel J, Forsgren L, French JA, Glynn M, Hesdorffer DC, Lee BI, Mathern GW, Moshé SL, Perucca E, Scheffer IE, Tomson T, Watanabe M, Wiebe S (2014) ILAE Official Report: A practical clinical definition of epilepsy. Epilepsia 55:475–482

Fisher RS, Van Emde Boas W, Blume W, Elger C, Genton P, Lee P, Engel J (2005) Epileptic seizures and epilepsy: Definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE). Epilepsia 46:470–472

Fisher RS, Harding G, Erba G, Barkley GL, Wilkins A, Epilepsy Foundation of America Working Group (2005) Photic- and Pattern-induced Seizures: A Review for the Epilepsy Foundation of America Working Group. Epilepsia 46:1426–1441

Frickey T, Lupas AN (2004) Phylogenetic analysis of AAA proteins. J Struct Biol 146:2–10

Frings H, Frings M (1953) The Production of Stocks of Albino Mice With. Behaviour 5:305–319

Fuller JL, Smith ME (1953) Kinetics of sound induced convulsions in some inbred mouse strains. Am J Physiol 172:661–670

Gachon F, Fonjallaz P, Damiola F, Gos P, Kodama T, Zakany J, Duboule D, Petit B, Tafti M, Schibler U (2004) The loss of circadian PAR bZip transcription factors results in epilepsy. Genes Dev 18:1397–1412

Garcia-Cairasco N (2002) A critical review on the participation of inferior colliculus in acoustic-motor and acoustic-limbic networks involved in the expression of acute and kindled audiogenic seizures. Hear Res 168:208–222

Garofalo S, Cornacchione M, Di Costanzo A (2012) From genetics to genomics of epilepsy. Neurol Res Int 2012:876234

Glazova MV, Nikitina LS, Hudik KA, Kirillova OD, Dorofeeva NA, Korotkov AA, Chernigovskaya EV (2015) Inhibition of ERK1/2 signaling prevents epileptiform behavior in rats prone to audiogenic seizures. J Neurochem 132:218–229

Goddard GV, McIntyre DC, Leech CK (1969) A permanent change in brain function resulting from daily electrical stimulation. Exp Neurol 25:295–330

Gregory RP, Oates T, Merry RT (1993) Electroencephalogram epileptiform abnormalities in candidates for aircrew training. Electroencephalogr Clin Neurophysiol 86:75–77

Grone BP, Baraban SC (2015) Animal models in epilepsy research: Legacies and new directions. Nat Neurosci 18:339–343

Grosso S, Farnetani MA, Francione S, Galluzzi P, Vatti G, Cordelli DM, Morgese G, Balestri P (2004) Hot water epilepsy and focal malformation of the parietal cortex development. Brain Dev 26:490–493

Gu W, Fukuda T, Isaji T, Hang Q, Lee H, Sakai S, Morise J, Mitoma J, Higashi H, Taniguchi N, Yawo H, Oka S, Gu J (2015) Loss of α1,6-Fucosyltransferase Decreases Hippocampal Long Term Potentiation: Implications for core fucosylation in the regulation of ampa receptor heteromerization and cellular signaling. J Biol Chem 290:17566–17575

Guenther UP, Handoko L, Laggerbauer B, Jablonka S, Chari A, Alzheimer M, Ohmer J, Plöttner O, Gehring N, Sickmann A, von Au K, Schuelke M, Fischer U (2009) IGHMBP2 is a ribosome-associated helicase inactive in the neuromuscular disorder distal SMA type 1 (DSMA1). Hum Mol Genet 18:1288–1300

Guerrini R, Dravet C, Genton P, Bureau M, Bonanni P, Ferrari AR, Roger J (1995) Idiopathic Photosensitive Occipital Lobe Epilepsy. Epilepsia 36:883–891

Hahn Y, Lee B (2005) Identification of nine human-specific frameshift mutations by comparative analysis of the human and the chimpanzee genome sequences. Bioinformatics 21:i186–i194

Hall CS (1947) Genetic differences in fatal audiogenic seizures. J Hered 38:3–6

Harding GFA, Mackenzie R, Klistorner A (1998) Severe persistent visual field constriction associated with vigabatrin. BMJ 316:232–233

He L, Vanlandewijck M, Raschperger E, Andaloussi Mäe M, Jung B, Lebouvier T, Ando K, Hofmann J, Keller A, Betsholtz C (2016) Analysis of the brain mural cell transcriptome. Sci Rep 6:35108

He ZW, Qu J, Zhang Y, Mao CX, Wang ZB, Mao XY, Deng ZY, Zhou BT, Yin JY, Long HY, Xiao B, Zhang Y, Zhou HH, Liu ZQ (2014) PRRT2 mutations are related to febrile seizures in epileptic patients. Int J Mol Sci 15:23408–23417

Hebsgaard SM, Korning PG, Tolstrup N, Engelbrecht J, Rouze P, Brunak S (1996) Splice site prediction in Arabidopsis thaliana DNA by combining local and global sequence information. Nucleic Acids Research 24:3439-3452

Hedera P, Ma S, Blair MA, Taylor KA, Hamati A, Bradford Y, Abou-Khalil B, Haines JL (2006) Identification of a Novel Locus for Febrile Seizures and Epilepsy on Chromosome 21q22. Epilepsia 47:1622–1628

Helbig I, Lowenstein DH (2013) Genetics of the epilepsies: where are we and where are we going? Curr Opin Neurol 26:179–185

Hildebrand MS, Dahl HH, Damiano JA, Smith RJ, Scheffer IE, Berkovic SF (2013) Recent advances in the molecular genetics of epilepsy. J Med Genet 50:271-279

Hirano M, Quinzii CM, Mitsumoto H, Hays AP, Roberts JK, Richard P, Rowland LP (2011) Senataxin mutations and amyotrophic lateral sclerosis. Amyotroph Lateral Scler 12:223–227

Hom AC, Leppik IE, Rask CA (1993) Effects of estradiol and progesterone on seizure sensitivity in oophorectomized DBA/2J mice and C57/EL hybrid mice. Neurology 43:198–204

Hortopan GA, Dinday MT, Baraban SC (2010) Spontaneous Seizures and Altered Gene Expression in GABA Signaling Pathways in a mind bomb Mutant Zebrafish. J Neurosci 30:13718–13728

Hughes JR (2008) The Photoparoxysmal Response: The Probable Cause of Attacks during Video Games. Clin EEG Neurosci 39:1–7

Humphries DE, Wong GW, Friend DS, Gurish MF, Qiu WT, Huang C, Sharpe AH, Stevens RL (1999) Heparin is essential for the storage of specific granule proteases in mast cells. Nature 400:769–772

Illingworth JL, Ring H (2013) Conceptual distinctions between reflex and nonreflex precipitated seizures in the epilepsies: A systematic review of definitions employed in the research literature. Epilepsia 54:2036–2047

Ioos C, Villeneuve N, Fohlen M, Badinant-Hubert N, Jalin C, Cheliout-Heraut F, Pinard JM (1999) Hot water epilepsy: a benign and underestimated form. Arch Pediatr 6:755–758

Ioos C, Fohlen M, Villeneuve N, Badinand-Hubert N, Jalin C, Cheliout-Heraut F, Pinard JM (2000) Hot Water Epilepsy: A Benign and Unrecognized Form. J Child Neurol 15:125–128

Italiano D, Striano P, Russo E, Leo A, Spina E, Zara F, Striano S, Gambardella A, Labate A, Gasparini S, Lamberti M, De Sarro G, Aguglia U, Ferlazzo E (2016) Genetics of reflex seizures and epilepsies in humans and animals. Epilepsy Res 121:47–54

Iyer LM, Leipe DD, Koonin E V, Aravind L (2004) Evolutionary history and higher order classification of AAA+ ATPases. J Struct Biol 146:11–31

Jian X, Boerwinkle E, Liu X (2014) In silico prediction of splice-altering single nucleotide variants in the human genome. Nucleic Acids Res 42:13534–13544

Jiang W, Duong TM, Lanerolle NC de (1999) The Neuropathology of Hyperthermic Seizures in the Rat. Epilepsia 40:5–19

Kanemoto K, Watanabe Y, Tsuji T, Fukami M, Kawasaki J (2001) Rub epilepsy: a somatosensory evoked reflex epilepsy induced by prolonged cutaneous stimulation. J Neurol Neurosurg Psychiatry 70:541–543

Kang JQ, Shen W, Lee M, Gallagher MJ, Macdonald RL (2010) Slow Degradation and Aggregation In Vitro of Mutant GABAA Receptor 2(Q351X) Subunits Associated with Epilepsy. J Neurosci 30:13895–13905

Kang JQ, Shen W, Macdonald RL (2009) Two molecular pathways (NMD and ERAD) contribute to a genetic epilepsy associated with the GABA(A) receptor GABRA1 PTC mutation, 975delC, S326fs328X. J Neurosci 29:2833–44

Karaca E, Harel T, Pehlivan D, Jhangiani SN, Gambin T, Coban Akdemir Z, Gonzaga-Jauregui C, Erdin S, Bayram Y, Campbell IM, Hunter JV, Atik MM, Van Esch H, Yuan B, Wiszniewski W, Isikay S, Yesil G, Yuregir OO, Tug Bozdogan S, et al (2015) Genes that Affect Brain Structure and Function Identified by Rare Variant Analyses of Mendelian Neurologic Disease. Neuron 88:499–513

Kalsi G, Kuo PH, Aliev F, Alexander J, McMichael O, Patterson DG, Walsh D, Zhao Z, Schuckit M, Nurnberger JJ, Edenberg H, Kramer J, Vladimirov V, Prescott CA, Dick DM, Kendler KS, Riley BP (2010) A systematic gene-based screen of chr4q22-q32 identifies association of a novel susceptibility gene, DKK2, with the quantitative trait of alcohol dependence symptom counts. Hum Mol Genet 19:2497–2506

Karan KR, Satishchandra P, Sinha S, Anand A (2017) Rare SLC1A1 variants in hot water epilepsy. Hum Genet 136:693–703

Kasteleijn-Nolst Trenité DG (1989) Photosensitivity in epilepsy. Electrophysiological and clinical correlates. Acta Neurol Scand Suppl 125:3–149

Kasteleijn-Nolst Trenité DGA (2012) Provoked and reflex seizures: Surprising or common? Epilepsia 53:105–113

Kawaguchi Y, Okamoto T, Taniwaki M, Aizawa M, Inoue M, Katayama S, Kawakami H, Nakamura S, Nishimura M, Akiguchi I, Kimura J (1994) CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. Nat Genet 8: 221–228

Kawai H, Allende ML, Wada R, Kono M, Sango K, Deng C, Miyakawa T, Crawley JN, Werth N, Bierfreund U, Sandhoff K, Proia RL (2001) Mice expressing only monosialoganglioside GM3 exhibit lethal audiogenic seizures. J Biol Chem 276:6885-6888

Keipert JA (1969) Epilepsy precipitated by bathing: water-immersion epilepsy. J Paediatr Child Health 5:244–247

Kerjan G, Gleeson JG (2007) Genetic mechanisms underlying abnormal neuronal migration in classical lissencephaly. Trends Genet 23:623–630

Killam KF (1969) Experimental epilepsy in primates. Ann N Y Acad Sci 162:610–616

Kim HD, Choe J, Seo YS (1999) The sen1(+) gene of Schizosaccharomyces pombe, a homologue of budding yeast SEN1, encodes an RNA and DNA helicase. Biochemistry 38:14697–14710

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

Klauenberg BJ, Sparber SB (1984) A Kindling-like Effect Induced by Repeated Exposure to Heated Water in Rats. Epilepsia 25:292–301

Koepp MJ, Caciagli L, Pressler RM, Lehnertz K, Beniczky S (2016) Reflex seizures, traits, and epilepsies: from physiology to pathology. Lancet Neurol 15:92–105

Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeirsson TE, Gulcher JR, Stefansson K (2002) A high-resolution recombination map of the human genome. Nat Genet 31:241–247

Koressaar T, Remm M (2007) Enhancements and modifications of primer design program Primer3. Bioinformatics 23:1289–1291

Kramer MA, Cash SS (2012) Epilepsy as a disorder of cortical network organization. Neuroscientist 18:360–372

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58:1347–1363

Krushinsky LV, Molodkina LN, Fless DA, Dobrokhotova LP, Steshenko AP, Semiokhina AF, Zorina ZA, Romanova LG (1970) The Functional State of the Brain during Sonic Stimulation. In: Physiological Effects of Noise. Springer US, Boston, MA, pp 159–183

Krushinsky LV (1963) Etude physiologique des differents types de crise convulsives de l'epilepsie audiogene du rat. In Busnel, R. G. (ed.), *Psychophysiologie, Neuropharmacologie et Biechemie de la Crise Audiogene*, CNRS, Paris, pp 71–92

Kullmann DM (2002) Genetics of epilepsy. J Neurol Neurosurg Psychiatry 73 Suppl 2:II32-35

Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4:1073–1081

Kurata S (1979) Epilepsy precipitated by bathing-A follow-up study. Brain Dev (Domestic ed) 11:400–405

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

Lathrop GM, Lalouel JM (1984) Easy calculations of lod scores and genetic risks on small computers. Am J Hum Genet 36:460–465

Lee SJ, Kwon S, Gatti JR, Korcari E, Gresser TE, Felix PC, Keep SG, Pasquale KC, Bai T, Blanchett-Anderson SA, Wu NW, Obeng-Nyarko C, Senagbe KM, Young KC, Maripudi S, Yalavarthi BC, Korcari D, Liu AY, Schaffler BC, et al (2017) Large-scale identification of human cerebrovascular proteins: Inter-tissue and intracerebral vascular protein diversity. PLoS One 12:e0188540

Lee YC, Yen DJ, Lirng JF, Yiu CH (2000) Epileptic seizures in a patient by immersing his right hand into hot water. Seizure 9:605–607

Lenoir P, Ramet J, De Meirleir L, D'Allest AM, Desprechins B, Loeb H (1989) Bathing-induced seizures. Pediatr Neurol 5:124–125

Leviton A, Cowan LD (1982) Epidemiology of Seizure Disorders in Children; pp. 62–83. Neuroepidemiology 1:62–83

Li H, Durbin R (2009) 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

Lindner TH, Hoffmann K (2005) easyLINKAGE: a PERL script for easy and automated two-/multi-point linkage analyses. Bioinformatics 21:405–407

Magiorkinis E, Diamantis A, Sidiropoulou K, Panteliadis C (2014) Highights in the History of Epilepsy: The Last 200 Years. Epilepsy Res Treat 2014:1–13

Mani KS, Gopalakrishnan PN, Vyas JN, Pillai MS (1968) "Hot-water epilepsy"-a peculiar type of reflex epilepsy. A preliminary report. Neurol India 16:107–110

Mani KS, Mani AJ, Kamesh CK, Ahuja GK (1972) Hot-water epi lepsy. clinical and electroencephalographic features--study of 60 cases. Neurology India (Suppl II):237-240

Mani KS, Mani AJ, Ramesh CK (1974) Hot-water epilepsy-a peculiar type of reflex epilepsy: clinical and EEG features in 108 cases. Trans Am Neurol Assoc 99:224–226

Margoliash D, van Drongelen W, Kohrman M (2010) Introducing Songbirds as a Model System for Epilepsy Research. J Clin Neurophysiol 27:433–437

Martínez AR, Colmenero MIA, Pereira AG, Vilaplana FXS, Morón JA, Marfa MP (2011) Reflex seizures in Rett syndrome. Epileptic Disord 13:389–393

Matsuno K, Eastman D, Mitsiades T, Quinn AM, Carcanciu ML, Ordentlich P, Kadesch T, Artavanis-Tsakonas S (1998) Human deltex is a conserved regulator of Notch signalling. Nat Genet 19:74–78

Matthews WB, Wright FK (1967) Hereditary primary reading epilepsy. Neurology 17:919–921

Mautner VF, Lindenau M, Gottesleben A, Goetze G, Kluwe L (2000) Supporting evidence of a gene for partial epilepsy on 10q. Neurogenetics 3:31–34

McMillan DR, White PC (2004) Loss of the transmembrane and cytoplasmic domains of the very large G-protein-coupled receptor-1 (VLGR1 or Mass1) causes audiogenic seizures in mice. Mol Cell Neurosci 26:322–329

McNamara JO, Puranam RS (1998) Epilepsy genetics: an abundance of riches for biologists. Curr Biol 8:168–170

Meghana A, Sinha S, Sathyaprabha TN, Subbakrishna DK, Satishchandra P (2012) Hot water epilepsy clinical profile and treatment—A prospective study. Epilepsy Res 102:160–166

Merla G, Brunetti-Pierri N, Micale L, Fusco C (2010) Copy number variants at Williams–Beuren syndrome 7q11.23 region. Hum Genet 128:3–26

Merrill MA, Clough RW, Jobe PC, Browning RA (2005) Brainstem Seizure Severity Regulates Forebrain Seizure Expression in the Audiogenic Kindling Model. Epilepsia 46:1380–1388

Mervaala E, Andermann F, Quesney LF, Krelina M (1990) Common dopaminergic mechanism for epileptic photosensitivity in progressive myoclonus epilepsies. Neurology 40:53–56

Micale L, Fusco C, Augello B, Napolitano LMR, Dermitzakis ET, Meroni G, Merla G, Reymond A (2008) Williams–Beuren syndrome TRIM50 encodes an E3 ubiquitin ligase. Eur J Hum Genet 16:1038–1049

Misawa H, Sherr EH, Lee DJ, Chetkovich DM, Tan A, Schreiner CE, Bredt DS (2002) Identification of a Monogenic Locus (jams1) Causing Juvenile Audiogenic Seizures in Mice. J Neurosci 22:10088–10093

Miyao M, Tezuka M, Kuwajima K, Kamoshita S (1982) Epilepsy induced by hot water immersion. Brain Dev 4:158

Mocquet V, Neusiedler J, Rende F, Cluet D, Robin JP, Terme JM, Duc Dodon M, Wittmann J, Morris C, Le Hir H, Ciminale V, Jalinot P (2012) The human T-lymphotropic virus type 1 tax protein inhibits nonsense-mediated mRNA decay by interacting with INT6/EIF3E and UPF1. J Virol 86:7530–7543

Mofenson HC, Weymuller CA, Greensher J (1965) Epilepsy Due to Water Immersion. JAMA 191:600

Mollicone R, Moore SEH, Bovin N, Garcia-Rosasco M, Candelier JJ, Martinez-Duncker I, Oriol R (2009) Activity, splice variants, conserved peptide motifs, and phylogeny of two new alpha1,3-fucosyltransferase families (FUT10 and FUT11). J Biol Chem 284:4723–4738

Mooney CM, Jimenez-Mateos EM, Engel T, Mooney C, Diviney M, Venø MT, Kjems J, Farrell MA, O'Brien DF, Delanty N, Henshall DC (2017) RNA sequencing of synaptic and cytoplasmic Upf1-bound transcripts supports contribution of nonsense-mediated decay to epileptogenesis. Sci Rep 7:41517

Moraes MFD, Chavali M, Mishra PK, Jobe PC, Garcia-Cairasco N (2005) A comprehensive electrographic and behavioral analysis of generalized tonic-clonic seizures of GEPR-9s. Brain Res 1033:1–12

Moran J (1976) So-called water immersion epilepsy. Ir J Med Sci 145:140

Morar B, Zhelyazkova S, Azmanov DN, Radionova M, Angelicheva D, Guergueltcheva V, Kaneva R, Scheffer IE, Tournev I, Kalaydjieva L, Sander JW (2011) A novel GEFS+ locus on 12p13.33 in a large Roma family. Epilepsy Res 97:198–207

Moreno AB, Martínez de Alba AE, Bardou F, Crespi MD, Vaucheret H, Maizel A, Mallory AC (2013) Cytoplasmic and nuclear quality control and turnover of single-stranded RNA modulate post-transcriptional gene silencing in plants. Nucleic Acids Res 41:4699–4708

Morimoto T, Hayakawa T, Sugie H, Awaya Y, Fukuyama Y (1985) Epileptic Seizures Precipitated by Constant Light, Movement in Daily Life, and Hot Water Immersion. Epilepsia 26:237–242

Mueller DR, Khalesi Z, Benzing V, Castiglione CI, Roder V (2017) Does Integrated Neurocognitive Therapy (INT) reduce severe negative symptoms in schizophrenia outpatients. Schizophr Res 188:92–97

Murrey HE, Gama CI, Kalovidouris SA, Luo WI, Driggers EM, Porton B, Hsieh-Wilson LC (2006) Protein fucosylation regulates synapsin Ia/Ib expression and neuronal morphology in primary hippocampal neurons. Proc Natl Acad Sci 103:21–26

Nabbout R, Baulac S, Desguerre I, Bahi-Buisson N, Chiron C, Ruberg M, Dulac O, LeGuern E (2007) New locus for febrile seizures with absence epilepsy on 3p and a possible modifier gene on 18p. Neurology 68:1374-1381

N'Gouemo P, Faingold CL (1999) The periaqueductal grey is a critical site in the neuronal network for audiogenic seizures: modulation by GABAA, NMDA and opioid receptors. Epilepsy Res 35:39–46

Nakayama J, Hamano K, Iwasaki N, Nakahara S, Horigome Y, Saitoh H, Aoki T, Maki T, Kikuchi M, Migita T, Ohto T, Yokouchi Y, Tanaka R, Hasegawa M, Matsui A, Hamaguchi H, Arinami T (2000) Significant evidence for linkage of febrile seizures to chromosome 5q14-q15. Hum Mol Genet 9:87–91

Nechiporuk T, Fernandez TE, Vasioukhin V (2007) Failure of Epithelial Tube Maintenance Causes Hydrocephalus and Renal Cysts in Dlg5−/− Mice. Dev Cell 13:338–350

Need AC, McEvoy JP, Gennarelli M, Heinzen EL, Ge D, Maia JM, Shianna KV, He M, Cirulli ET, Gumbs CE, Zhao Q, Campbell CR, Hong L, Rosenquist P, Putkonen A, Hallikainen T, Repo-Tiihonen E, Tiihonen J, Levy DL, et al (2012) Exome Sequencing Followed by Large-Scale Genotyping Suggests a Limited Role for Moderately Rare Risk Factors of Strong Effect in Schizophrenia. Am J Hum Genet 91:303–312

Nguyen DK, Rouleau I, Sénéchal G, Ansaldo AI, Gravel M, Benfenati F, Cossette P (2015) X-linked focal epilepsy with reflex bathing seizures: Characterization of a distinct epileptic syndrome. Epilepsia 56:1098–1108

Olson H, Shen Y, Avallone J, Sheidley BR, Pinsky R, Bergin AM, Berry GT, Duffy FH, Eksioglu Y, Harris DJ, Hisama FM, Ho E, Irons M, Jacobsen CM, James P, Kothare S, Khwaja O, Lipton J, Loddenkemper T, et al (2014) Copy number variation plays an important role in clinical epilepsy. Ann Neurol 75:943–958

Omasits U, Ahrens CH, Müller S, Wollscheid B (2014) Protter: interactive protein feature visualization and integration with experimental proteomic data. Bioinformatics 30:884–886

Orrico A, Galli L, Zappella M, Lam CW, Bonifacio S, Torricelli F, Hayek G (2001) Possible case of Pitt-Hopkins syndrome in sibs. Am J Med Genet 103:157–159

Pacey LKK, Heximer SP, Hampson DR (2009) Increased GABA(B) receptor-mediated signaling reduces the susceptibility of fragile X knockout mice to audiogenic seizures. Mol Pharmacol 76:18–24

Pamenter ME, Hogg DW, Gu XQ, Buck LT, Haddad GG (2012) Painted turtle cortex is resistant to an in vitro mimic of the ischemic mammalian penumbra. J Cereb Blood Flow Metab 32:2033–2043

Panayiotopoulos CP (1996) Epilepsies characterized by seizures with specific modes of precipitation (reflex epilepsies). In: Epilepsy in Children. London: Chapman & Hall 355–375

Panayiotopoulos CP (2005) Reflex Seizures and Reflex Epilepsies, In: The Epilepsies: Seizures, Syndromes and Management Oxfordshire (UK): Bladon Medical Publishing

Parsonage MJ, Moran JH, Exley KA (1976) So-called water immersion epilepsy. Epileptology Proceedings of the 7th Internaional Symposium on Epilepsy. Stuttgart 50–60

Paulsen RD, Soni DV, Wollman R, Hahn AT, Yee MC, Guan A, Hesley JA, Miller SC, Cromwell EF, Solow-Cordero DE, Meyer T, Cimprich KA (2009) A Genome-wide siRNA Screen Reveals Diverse Cellular Processes and Pathways that Mediate Genome Stability. Mol Cell 35:228–239

Peiffer A, Thompson J, Charlier C, Otterud B, Varvil T, Pappas C, Barnitz C, Gruenthal K, Kuhn R, Leppert M (1999) A locus for febrile seizures (FEB3) maps to chromosome 2q23-24. Ann Neurol 46:671–678

Peter B, Wijsman EM, Nato AQ, Matsushita MM, Chapman KL, Stanaway IB, Wolff J, Oda K, Gabo VB, Raskind WH, Raskind WH (2016) Genetic Candidate Variants in Two Multigenerational Families with Childhood Apraxia of Speech. PLoS One 11:e0153864

Pinto D, Westland B, de Haan G-J, Rudolf G, da Silva BM, Hirsch E, Lindhout D, Trenité DG, Koeleman BP (2005) Genome-wide linkage scan of epilepsy-related photoparoxysmal electroencephalographic response: evidence for linkage on chromosomes 7q32 and 16p13. Hum Mol Genet 14:171–178

Piro RM, Molineris I, Ala U, Di Cunto F (2011) Evaluation of candidate genes from orphan FEB and GEFS+ loci by analysis of human brain gene expression atlases. PLoS One 6:e23149

Poduri A, Wang Y, Gordon D, Barral-Rodriguez S, Barker-Cummings C, Ulgen A, Chitsazzadeh V, Hill RS, Risch N, Hauser WA, Pedley TA, Walsh CA, Ottman R (2009) Novel susceptibility locus at chromosome 6q16.3-22.31 in a family with GEFS+. Neurology 73:1264–1272

Poduri A, Lowenstein D (2011) Epilepsy genetics-past, present, and future. Curr Opin Genet Dev 21:325–332

Polakova S, Molnarova L, Hyppa RW, Benko Z, Misova I, Schleiffer A, Smith GR, Gregan J (2016) Dbl2 Regulates Rad51 and DNA Joint Molecule Metabolism to Ensure Proper Meiotic Chromosome Segregation. PLoS Genet 12:e1006102

Poletaeva II, Fedotova IB, Sourina NM, Kostina ZA (2011) Audiogenic Seizures -Biological Phenomenon and Experimental Model of Human Epilepsies

Poulton C, Oegema R, Heijsman D, Hoogeboom J, Schot R, Stroink H, Willemsen MA, Verheijen FW, van de Spek P, Kremer A, Mancini GMS (2013) Progressive cerebellar atrophy and polyneuropathy: expanding the spectrum of PNKP mutations. Neurogenetics 14:43–51

Poza JJ, Sáenz A, Martínez-Gil A, Cheron N, Cobo AM, Urtasun M, Martí-Massó JF, Grid D, Beckmann JS, Prud 'homme JF, López De Munain A (1999) Autosomal Dominant Lateral Temporal Epilepsy: Clinical and Genetic Study of a Large Basque Pedigree Linked to Chromosome 10q. Ann Neurol 45:182-188

Puranam RS, Jain S, Kleindienst AM, Saxena S, Kim MK, 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

Puranam RS, McNamara JO (2001) Epilepsy and all that jazz. Nat Med 7:1103–1105

Puttagunta R, Gordon LA, Meyer GE, Kapfhamer D, Lamerdin JE, Kantheti P, Portman KM, Chung WK, Jenne DE, Olsen AS, Burmeister M (2000) Comparative maps of human 19p13.3 and mouse chromosome 10 allow identification of sequences at evolutionary breakpoints. Genome Res 10:1369–1380

Putnam TJ, Merritt HH (1937) Experimental Determination of the Anticonvulsant Properties of Some Phenyl Derivatives. Science 85:525–526

Qiang L, Zhao B, Shah P, Sample A, Yang S, He YY (2016) Autophagy positively regulates DNA damage recognition by nucleotide excision repair. Autophagy 12:357–68

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842

Racicot F, Obaid S, Bouthillier A, Guillon-Létourneau L, Clément JF, Nguyen DK (2016) Praxis-induced reflex seizures mainly precipitated by writing due to a parietal focal cortical dysplasia. Epilepsy Behav Case Reports 6:52–54

Radhakrishnan K, Silbert PL, Klass DW (1995) Reading epilepsy. Brain 118:75–89

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

Ramirez IBR, Pietka G, Jones DR, Divecha N, Alia A, Baraban SC, Hurlstone AFL, Lowe M (2012) Impaired neural development in a zebrafish model for Lowe syndrome. Hum Mol Genet 21:1744–1759

Raschle M, Van Komen S, Chi P, Ellenberger T, Sung P (2004) Multiple Interactions with the Rad51 Recombinase Govern the Homologous Recombination Function of Rad54. J Biol Chem 279:51973–51980

Ratnapriya R, Satishchandra P, Dilip S, Gadre G, Anand A (2009a) Familial autosomal dominant refex epilepsy triggered by hot water maps to 4q24-q28. Hum Genet 126:677–683

Ratnapriya R, Satishchandra P, Kumar SD, Gadre G, Reddy R, Anand A (2009b) A locus for autosomal dominant reflex epilepsy precipitated by hot water maps at chromosome 10q21.3-q22.3. Hum Genet 125:541–549

Reva B, Antipin Y, Sander C (2011) Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res 39:e118

Ribak CE, Khurana V, Lien NT (1994) The effect of midbrain collicular knife cuts on audiogenic seizure severity in the genetically epilepsy-prone rat. J Hirnforsch 35:303–311

Ricci S, Cusmai R, Fusco L, Vigevano F (1995) Reflex myoclonic epilepsy in infancy: a new age-dependent idiopathic epileptic syndrome related to startle reaction. Epilepsia 36:342–348

Robinson R, Gardiner M (2004) Molecular basis of Mendelian idiopathic epilepsies. Ann Med 36:89–97

Rogers KJ, Fu W, Akey JM, Monnat Jr RJ (2014) Global and disease-associated genetic variation in the human Fanconi anemia gene family. Hum Mol Genet 23:6815–6825

Roos C, Kolmer M, Mattila P, Renkonen R (2002) Composition of *Drosophila melanogaster* Proteome Involved in Fucosylated Glycan Metabolism. J Biol Chem 277:3168–3175

Roos RA, van Dijk JG (1988) Reflex-epilepsy induced by immersion in hot water. Case report and review of the literature.Eur Neurol 28:6-10

Sakurai A, Tamvacakis AN, Katz PS (2014) Hidden synaptic differences in a neural circuit underlie differential behavioral susceptibility to a neural injury. Elife 3:e02598

Salzmann A, Guipponi M, Lyons PJ, Fricker LD, Sapio M, Lambercy C, Buresi C, Bencheikh BOA, Lahjouji F, Ouazzani R, Crespel A, Chaigne D, Malafosse A (2012) Carboxypeptidase A6 gene (CPA6) mutations in a recessive familial form of febrile seizures and temporal lobe epilepsy and in sporadic temporal lobe epilepsy. Hum Mutat 33:124–135

Sambrook J, Russell RW (2001) Molecular cloning: A laboratory manual, 3rd ed. Cold Spring Harbor Laboratory press, Cold Spring Harbor, NewYork

Sanchez Bassères D, Vedelago Tizzei E, Duarte AA, Costa FF, Teresinha Olalla Saad S (2002) ARHGAP10, a novel human gene coding for a potentially cytoskeletal Rho-GTPase activating protein. Biochem Biophys Res Commun 294:579–585

Santa Maria SR, Kwon Y, Sung P, Klein HL (2013) Characterization of the interaction between the Saccharomyces cerevisiae Rad51 recombinase and the DNA translocase Rdh54. J Biol Chem 288:21999–22005

Santos-Silva R, Passas A, Rocha C, Figueiredo R, Mendes-Ribeiro J, Fernandes S, Biskup S, Leão M (2015) Bilateral frontoparietal polymicrogyria: A novel GPR56 mutation and an unusual phenotype. Neuropediatrics 46:134–138

Saracchi E, Castelli M, Bassi MT, Brighina E, Cereda D, Marzorati L, Patassini M, Appollonio I, Ferrarese C, Brighina L (2014) A novel heterozygous SETX mutation in a patient presenting with chorea and motor neuron disease. Amyotroph Lateral Scler and Frontotemporal Degener 15:1-2

Satishchandra P (2003) Hot-Water Epilepsy. Epilepsia 44:29–32

Satishchandra P, Kallur KG, Jayakumr PN (2000) Inter-ictal and ictal 99 m TC ECD SPECT scan in hot-water epilepsy. Epilepsia 42(Suppl 1):158

Satishchandra P, Shivaramakrishana A, Kaliaperumal VG, Schoenberg BS (1988) Hot-Water Epilepsy: A Variant of Reflex Epilepsy in Southern India. Epilepsia 29:52–56

Satishchandra P. Shivaramakrishana A, Kaliaperumal VG (1985) Hot water epilepsy-A variant of reflex epilepsy in parts of South India. J Neurol 232(suppl):212.

Savitha MR, Krishnamurthy B, Ashok DA, Ramachandra NB (2007) Self abortion of attacks in patients with Hot Water Epilepsy. Indian Pediatr 44:295–298

Scheffer I, Berkovic SF (1997) Generalized epilepsy with febrile seizures plus. A genetic disorder with heterogeneous clinical phenotypes. Brain 120:479–490

Scheffer IE, Berkovic S, Capovilla G, Connolly MB, French J, Guilhoto L, Hirsch E, Jain S, Mathern GW, Moshé SL, Nordli DR, Perucca E, Tomson T, Wiebe S, Zhang YH, Zuberi SM (2017) ILAE classification of the epilepsies: Position paper of the ILAE Commission for Classification and Terminology. Epilepsia 58:512–521

Schubert J, Siekierska A, Langlois M, May P, Huneau C, Becker F, Muhle H, Suls A, Lemke JR, De Kovel CGF, Thiele H, Konrad K, Kawalia A, Toliat MR, Sander T, Rüschendorf F, Caliebe A, Nagel I, Kohl B, et al (2014) Mutations in STX1B, encoding a presynaptic protein, cause fever-associated epilepsy syndromes. Nat Genet 46:1327–1332

Schultz ML, Tecedor L, Stein CS, Stamnes MA, Davidson BL (2014) CLN3 Deficient Cells Display Defects in the ARF1-Cdc42 Pathway and Actin-Dependent Events. PLoS One 9:e96647

Schwarz JM, Rödelsperger C, Schuelke M, Seelow D (2010) MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods 7:575–576

Selmer K, Eriksson AS, Brandal K, Egeland T, Tallaksen C, Undlien D (2009) Parental *SCN1A* mutation mosaicism in familial Dravet syndrome. Clin Genet 76:398–403

Semiokhina AF, Fedotova IB, Poletaeva II (2006) Rats of Krushinsky-Molodkina strain: studies of audiogenic epilepsy, vascular pathology, and behavior. Zh Vyssh Nerv Deiat Im I P Pavlova 56:298–316

Seneviratne U (2001) Bathing epilepsy. Seizure 10:516–517

Shaw NJ, Livingston JH, Minns RA, Clarke M (1988) Epilepsy Precipitated by Bathing. Dev Med Child Neurol 30:108–111

Shen J, Gilmore EC, Marshall CA, Haddadin M, Reynolds JJ, Eyaid W, Bodell A, Barry B, Gleason D, Allen K, Ganesh VS, Chang BS, Grix A, Hill RS, Topcu M, Caldecott KW, Barkovich AJ, Walsh CA (2010) Mutations in PNKP cause microcephaly, seizures and defects in DNA repair. Nat Genet 42:245–249

Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GLA, Edwards KJ, Day INM, Gaunt TR (2013) Predicting the Functional, Molecular, and Phenotypic Consequences of Amino Acid Substitutions using Hidden Markov Models. Hum Mutat 34:57–65

Silva S, Altmannova V, Luke-Glaser S, Henriksen P, Gallina I, Yang X, Choudhary C, Luke B, Krejci L, Lisby M (2016) Mte1 interacts with Mph1 and promotes crossover recombination and telomere maintenance. Genes Dev 30:700–717

Silva-Barrat C, Menini C (1990) Photosensitive Epilepsy of the Baboon: A Generalized Epilepsy with a Motor Cortical Origin. In: Generalized Epilepsy. Birkhäuser Boston, Boston, MA, pp 286–297

Singh G, Jakob S, Kleedehn MG, Lykke-Andersen J (2007) Communication with the Exon-Junction Complex and Activation of Nonsense-Mediated Decay by Human Upf Proteins Occur in the Cytoplasm. Mol Cell 27:780–792

Singh R, Scheffer IE, Crossland K, Berkovic SF (1999) Generalized epilepsy with febrile seizures plus: a common childhood-onset genetic epilepsy syndrome. Ann Neurol 45:75–81

Singleton MR, Dillingham MS, Wigley DB (2007) Structure and Mechanism of Helicases and Nucleic Acid Translocases. Annu Rev Biochem 76:23–50

Sirmaci A, Edwards YJK, Akay H, Tekin M (2012) Challenges in Whole Exome Sequencing: An Example from Hereditary Deafness. PLoS One 7:e32000

Skradski SL, Clark AM, Jiang H, White HS, Fu YH, Ptáček LJ (2001) A Novel Gene Causing a Mendelian Audiogenic Mouse Epilepsy. Neuron 31:537–544

Smogorzewska A, Desetty R, Saito TT, Schlabach M, Lach FP, Sowa ME, Clark AB, Kunkel TA, Harper JW, Colaiácovo MP, Elledge SJ (2010) A Genetic Screen Identifies FAN1, a Fanconi Anemia-Associated Nuclease Necessary for DNA Interstrand Crosslink Repair. Mol Cell 39:36–47

Sosnay PR, Cutting GR (2014) Interpretation of genetic variants. Thorax 69:295–297

Stensman R, Ursing B (1971) Epilepsy precipitated by hot water immersion. Neurology 21:559–562

Stephani U, Tauer U, Koeleman B, Pinto D, Neubauer BA, Lindhout D (2004) Genetics of Photosensitivity (Photoparoxysmal Response): A Review. Epilepsia 45:19–23

Stevens JR (1962) Central and peripheral factors in epileptic discharge. Clinical studies Arch Neurol 26:409–419

Stoll M, Corneliussen B, Costello CM, Waetzig GH, Mellgard B, Koch WA, Rosenstiel P, Albrecht M, Croucher PJP, Seegert D, Nikolaus S, Hampe J, Lengauer T, Pierrou S, Foelsch UR, Mathew CG, Lagerstrom-Fermer M, Schreiber S (2004) Genetic variation in DLG5 is associated with inflammatory bowel disease. Nat Genet 36:476–480

Strauss KA, Puffenberger EG, Huentelman MJ, Gottlieb S, Dobrin SE, Parod JM, Stephan DA, Morton DH (2006) Recessive Symptomatic Focal Epilepsy and Mutant Contactin-Associated Protein-like 2. N Engl J Med 354:1370–1377

Subrahmanyam HS (1972) Hot water epilepsy. Neurology 20:224–226

Szymonowicz W, Meloff KL (1978) Hot Water Epilepsy. Can J Neurol Sci 5:247–251

Tampere M, Mortusewicz O (2016) DNA Damage Induction by Laser Microirradiation. *Bio-protocol* 6(23): e2039

Tauer U, Lorenz S, Lenzen KP, Heils A, Muhle H, Gresch M, Neubauer BA, Waltz S, Rudolf G, Mattheisen M, Strauch K, Nürnberg P, Schmitz B, Stephani U, Sander T (2005) Genetic dissection of photosensitivity and its relation to idiopathic generalized epilepsy. Ann Neurol 57:866–873

Taylor I, Berkovic SF, Scheffer IE (2013) Genetics of epilepsy syndromes in families with photosensitivity. Neurology 80:1322–1329

Teillet MA, Naquet R, Le Gal La Salle G, Merat P, Schuler B, Le Douarin NM (1991) Transfer of genetic epilepsy by embryonic brain grafts in the chicken. Proc Natl Acad Sci U S A 88:6966–6970

Teixeira D, Sheth U, Valencia-Sanchez MA, Brengues M, Parker R (2005) Processing bodies require RNA for assembly and contain nontranslating mRNAs. RNA 11:371–382

Tezer FI, Ertas N, Yalcin D, Saygi S (2006) Hot water epilepsy with cerebral lesion: A report of five cases with cranial MRI findings. Epilepsy Behav 8:672–676

Thomas RH, Berkovic SF (2014) The hidden genetics of epilepsy - A clinically important new paradigm. Nat Rev Neurol 10:283–292

Ule J, Ule A, Spencer J, Williams A, Hu JS, Cline M, Wang H, Clark T, Fraser C, Ruggiu M, Zeeberg BR, Kane D, Weinstein JN, Blume J, Darnell RB (2005) Nova regulates brain-specific splicing to shape the synapse. Nat Genet 37:844–852

Ullal GR, Satishchandra P, Kalladka D, Rajashekar K, Archana K, Mahadevan A, Shankar SK (2006) Kindling & mossy fibre sprouting in the rat hippocampus following hot water induced hyperthermic seizures. Indian J Med Res 124:331–342

Ullal GR, Satishchandra P, Shankar SK (1996) Hyperthermic seizures: an animal model for hot-water epilepsy. Seizure 5:221–228

Van Bogaert P, Azizieh R, Désir J, Aeby A, De Meirleir L, Laes J-F, Christiaens F, Abramowicz MJ (2007) Mutation of a potassium channel-related gene in progressive myoclonic epilepsy. Ann Neurol 61:579–586

Velmurugendran CU (1985) Reflex epilepsy. J Neurol 232(suppl):212

Venkataramiah V, Ghorpade VAP (2013) Pseudo hot water epilepsy: A pilot study on behavioral psychotherapeutic management. Indian J Psychiatry 55:149–153

Verrotti A, Trotta D, Salladini C, di Corcia G, Chiarelli F (2004) Topical Review: Photosensitivity and Epilepsy. J Child Neurol 19:571–578

Voiculescu V, Haţegan D, Manole E, Ulmeanu A (1994) Increased susceptibility to audiogenic seizures following withdrawal of progesterone. Rom J Neurol Psychiatry 32:131–133

Walker JE, Saraste M, Runswick MJ, Gay NJ (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J 1:945–951

Wallace RH, Scheffer IE, Barnett S, Richards M, Dibbens L, Desai RR, Lerman-Sagie T, Lev D, Mazarib A, Brand N, Ben-Zeev B, Goikhman I, Singh R, Kremmidiotis G, Gardner A, Sutherland GR, George AL, Mulley JC, Berkovic SF (2001) Neuronal Sodium-Channel α1-Subunit Mutations in Generalized Epilepsy with Febrile Seizures Plus. Am J Hum Genet 68:859–865

Wallace RH, Marini C, Petrou S, Harkin LA, Bowser DN, Panchal RG, Williams DA, Sutherland GR, Mulley JC, Scheffer IE, Berkovic SF (2001) Mutant GABA(A) receptor gamma2-subunit in childhood absence epilepsy and febrile seizures. Nat Genet 28:49–52

Wang SHJ, Celic I, Choi SY, Riccomagno M, Wang Q, Sun LO, Mitchell SP, Vasioukhin V, Huganir RL, Kolodkin AL (2014) Dlg5 regulates dendritic spine formation and synaptogenesis by controlling subcellular N-cadherin localization. J Neurosci 34:12745–12761

Wang Y, Shao L, Shi S, Harris RJ, Spellman MW, Stanley P, Haltiwanger RS (2001) Modification of epidermal growth factor-like repeats with O-fucose. Molecular cloning and expression of a novel GDP-fucose protein O-fucosyltransferase. J Biol Chem 276:40338–40345

Wasterlain CG, Jonec V (1983) Chemical kindling by muscarinic amygdaloid stimulation in the rat. Brain Res 271:311–323

Wetzel W, Popov N, Lössner B, Schulzeck S, Honza R, Matthies H (1980) Effect of L-fucose on brain protein metabolism and retention of a learned behavior in rats. Pharmacol Biochem Behav 13:765–771

Winawer MR, Ottman R, Hauser WA, Pedley TA (2000) Autosomal dominant partial epilepsy with auditory features: Defining the phenotype. Neurology 54:2173-2176

Wolf P (2005) From precipitation to inhibition of seizures: rationale of a therapeutic paradigm. Epilepsia 46:15-16

Wolf P, Goosses R (1986) Relation of photosensitivity to epileptic syndromes. J Neurol Neurosurg Psychiatry 49:1386–1391

Xu Y, Yao Shugart Y, Wang G, Cheng Z, Jin C, Zhang K, Wang J, Yu H, Yue W, Zhang F, Zhang D (2016) Altered expression of mRNA profiles in blood of early-onset schizophrenia. Sci Rep 6:16767

Xue LY, Ritaccio AL (2006) Reflex seizures and reflex epilepsy. Am J Electroneurodiagnostic Technol 46:39–48

Yacubian EM, Wolf P (2015) Orofacial reflex myocloni. Definition, relation to epilepsy syndromes, nosological and prognosis significance. A focused review. Seizure 30:1–5

Yavuz EN, Ozdemir O, Catal S, Bebek N, Ozbek U, Baykan B (2012) Bromodomain-Containing Protein 2 gene in photosensitive epilepsy. Seizure 21:646–648

Zahir FR, Mwenifumbo JC, Chun HJE, Lim EL, Van Karnebeek CDM, Couse M, Mungall KL, Lee L, Makela N, Armstrong L, Boerkoel CF, Langlois SL, McGillivray BM, Jones SJM, Friedman JM, Marra MA (2017) Comprehensive whole genome sequence analyses yields novel genetic and structural insights for Intellectual Disability. BMC Genomics 18:403

Zhang G, Zhang X, Wang X, Li JP (2014) Towards understanding the roles of heparan sulfate proteoglycans in Alzheimer's disease. Biomed Res Int 2014:516028

Zheng N, Wang P, Jeffrey PD, Pavletich NP (2000) Structure of a c-Cbl–UbcH7 Complex: RING Domain Function in Ubiquitin-Protein Ligases. Cell 102:533–539

Zifkin BG, Kasteleijn-Nolst Trenité D (2000) Reflex epilepsy and reflex seizures of the visual system: a clinical review. Epileptic Disord 2:129–136

Zweier C, de Jong EK, Zweier M, Orrico A, Ousager LB, Collins AL, Bijlsma EK, Oortveld MAW, Ekici AB, Reis A, Schenck A, Rauch A (2009) CNTNAP2 and NRXN1 Are Mutated in Autosomal-Recessive Pitt-Hopkins-like Mental Retardation and Determine the Level of a Common Synaptic Protein in Drosophila. Am J Hum Genet 85:655–666

Web resources

BLAST tool at National Centre for Biotechnology Information, National Institute of Health: http://blast.ncbi.nlm.nih.gov/Blast.cgi

ClustalW2 at European Bioinformatics Institute: http://www.ebi.ac.uk/Tools/msa/clustalw2/

Entrez Gene database at National Centre for Biotechnology Information (NCBI, National Institute of Health (NIH):http://www.ncbi.nlm.nih.gov/gene

Entrez SNP database at NCBI, NIH: http://www.ncbi.nlm.nih.gov/snp/

GenBank database at NCBI, NIH: http://www.ncbi.nlm.nih.gov/genbank/

Human Genome Map Viewer database at NCBI, NIH: http://www.ncbi.nlm.nih.gov/mapview/

Online Mendelian Inheritance in Man database at at NCBI, NIH : http://www.ncbi.nlm.nih.gov/omim/

Primer3 Input software:http://frodo.wi.mit.edu/

UniSTS database at NCBI, NIH: http://www.ncbi.nlm.nih.gov/unists

Ensembl genome browser: https://asia.ensembl.org/index.html

1000 Genomes dataset: http://browser.1000genomes.org/index.html

Exome Aggregation Consortium (ExAC) dataset**:** http://exac.broadinstitute.org/

Exome Variant Server (EVS) NHLBI GO Exome Sequencing Project (ESP): http://evs.gs.washington.edu/EVS/

PolyPhen-2: http://genetics.bwh.harvard.edu/pph2/

SIFT: http://sift.bii.a-star.edu.sg/

Mutation Taster: http://www.mutationtaster.org/

Mutation Assessor: http://mutationassessor.org/r3/

FATHMM: http://fathmm.biocompute.org.uk/

NNSplice: https://omictools.com/nnsplice-tool

NetGene2: http://www.cbs.dtu.dk/services/NetGene2/

Human Splicing Finder: http://www.umd.be/HSF3/

Protter: http://wlab.ethz.ch/protter/start/