On molecular genetic analysis of juvenile myoclonic epilepsy

A thesis submitted for the degree of

Doctor of Philosophy

by

Shveta Jaishankar

Molecular Biology and Genetics Unit
Jawaharlal Nehru Centre for Advanced Scientific Research
Bangalore 560 064, India

May 2021

Dedicated to

Amma, Appa and Adi

Declaration

I hereby declare that this thesis entitled "***On molecular genetic analysis of juvenile myoclonic epilepsy***" 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 have 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.

Shveta Jaishankar

Place: Bangalore

Date: 28th May 2021

Certificate

This is to certify that the work described in this thesis entitled "***On molecular genetic analysis of juvenile myoclonic epilepsy***" is the result of the investigations carried out by Ms. Shveta Jaishankar 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: 28th May 2021

Acknowledgements

The process of earning a doctorate and writing a dissertation is long and arduous and it certainly cannot be done singlehandedly. My work would be incomplete without thanking people who have supported and influenced me to take up and successfully complete my graduate study.

My deep gratitude goes, first to Prof. Anuranjan Anand, who expertly guided, constantly supported and encouraged me through the duration of my graduate study. I thank him for giving me an opportunity to work on genetics of human disorders. His advice on both research as well as on my career have been priceless, and I have learnt a lot from him through the course of my stay in his lab.

I would like to thank the faculty members of the molecular biology & genetics unit and neuroscience unit: Prof Maneesha Inamdar (Chair MBGU), Prof Udaykumar Ranga, Prof. MRS Rao, Prof. Namita Surolia, Prof. Hemalatha Balaram, Prof Tapas Kumar Kundu, Prof. Kaustuv Sanyal, Prof Ravi Manjithaya, Prof Sheeba Vasu, Prof James Chelliah and Dr. Kushagra Bansal for their suggestions, discussions during departmental seminars and interactions during course work. I must thank Dr. Ramesh who has been influential in shaping my foundation in research methodologies. I extend my thanks to our clinical collaborators from the NIMHANS, SCTIMST and other epilepsy specialty clinics for contributing to our sample collection without which our research work would have been impossible. I thank Dr. Thangaraj at CCMB, Hyderabad and Sudhakar with whom I had an opportunity to work and collaborate with respect to male infertility studies. I thank Prof. Tapas Kumar Kundu for kindly providing the pMIR-REPORT-Luciferase plasmids required for my experiments. I thank Clevergene, especially Dr. Reddy and Amrita for their technical support extended to my study. I would also like to acknowledge Dr. Gurudutta at Centre for Human Genetics, Bangalore for his insightful comments and suggestions.

My thanks to my present and past lab members who have been instrumental in creating a stimulating and comfortable environment to work in - Nishtha, Praveen, Manpreet, Kalpita, Shalini, Sourav, Rammurthy, Mohan, Chandrashekar, Pooja, Shri, Sambhavi, Girija, Sukanya, Yashwini, and Nazia. Special thanks to Girija and Sourav for their help in sequencing and bioinformatic analysis. Shalini and Mohan for being wonderful friends both inside and outside lab settings.

The financial support of the JNCASR and ASHG is gratefully acknowledged. I also acknowledge support from the confocal, sequencing, and central facilities at MBGU and NSU, computer lab, library, academic and administrative sections. I specially thank central facilities: flow cytometry (Dr. Narendra), bio-imaging (Suma, Sunil, Keerthana) for their co-operation and their help in conducting my experiments successfully. My stay at JNCASR has been made comfortable by the hostel facilities that include, mess, book club, sports facilities and utility stores.

Getting through my dissertation required more than academic support, and I have many people to thank for. My batchmates with whom I began my studies in JNCASR- Sunaina, Lakshmi, Vikas, Surabhi, Manaswini, Avani, Joydeep, Amrutha, Malini and Dhanya. My colleagues and friends in JNCASR- Mahesh, Prabhu, Rishi, Rebu, Sutanuka, Anjali, Piyush, Amol, Krishnendu, Sridevi, Greeshma, Vybhav, Sunil, Samarth, Atif, Chakradhar who have helped me learn, explore new things, and have had fun times with.

Finally, I must express my very profound gratitude to my parents, my brother and my extended family for their love, unfailing support, constant encouragement throughout my years of study.

Abbreviation

A	adenine
C	cytosine
G	guanine
T	thymine
θ	recombination fraction
°C	degree Celsius
cm	centimetre
µg	microgram
µl	microlitre
µM	micromolar
kb	kilobases
kDa	kiloDaltons
Mb	megabases
mg	milligrams
mM	millimolar
M	molar
ml	millilitre
mm	millimetre
ng	nanogram
nm	nanometre
pmol	picomole
BCA	bicinchoninic acid
AED	anti-epileptic drugs
bp	base pairs
BSA	bovine serum albumin
BWA	burrows wheeler aligner
cDNA	complementary deoxyribose nucleic acid
ChIP	chromatin immunoprecipitation
CNV	copy number variation
CO_2	carbon dioxide
C-terminal	carboxy-terminal
DAPI	4', 6-diamidino-2-phenylindole
DEPC	diethyl pyrocarbonate
DMSO	dimethyl sulfoxide
DNA	deoxyribose nucleic acid
dNTP	deoxyribonucleotide triphosphate
dpc	days postcoitum
dpf	days post-fertilization
EDTA	ethylene diamine tetra-acetate
EEG	electroencephalogram
FBS	fetal bovine serum
g	unit of acceleration
g	grams
GABA	gamma aminobutyric acid
GFP	green fluorescent protein
GGE	genetic generalized epilepsy
GTCS	generalized tonic clonic seizure

HCl	hydrochloric acid
HMG	high-mobility group
ILAE	International League Against Epilepsy
LINE	long interspersed nuclear elements
LOF	loss of function
LOD	logarithm of odds
MAF	minor allele frequency
miRNA	micro ribonucleic acid
MRI	magnetic resonance imaging
mRNA	messenger ribonucleic acid
MTR	missense tolerance ratio
MW	molecular weight
ncRNA	non-coding ribonucleic acid
NMD	nonsense mediated decay
NMDA	N-Methyl-D-aspartic acid or N-Methyl-D-aspartate
N-terminal	amino-terminal
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PFA	paraformaldehyde
pH	power of hydrogen
pLI	probability of being loss of function intolerant
PTCs	premature translation-termination codons
qPCR	quantitative polymerase chain reaction
RNA	ribonucleic acid
rpm	revolutions per minute
RT-PCR	reverse transcription polymerase chain reaction
RVIS	residual variation intolerance score
SEM	standard error of the mean
SD	standard deviation
SDS	sodium dodecyl sulphate
SIFT	sorting intolerant from tolerant
SINE	short interspersed nuclear elements
snoRNA	small nucleolar ribonucleic acid
SNP	single-nucleotide polymorphism
SNV	single-nucleotide variation
SRY	sex-determining region Y
SUDEP	sudden unexpected death in epilepsy
TAE	tris-acetate-EDTA
Taq	Thermus aquaticus
TSS	transcription start sites
U	units
UTR	untranslated region
VPA	Valproate
WGS	whole genome sequencing
WT	wild-type

Table of Contents

Chapter 1

Introduction

Epilepsy is a neurological disorder characterized by occurrence of recurrent seizures that are sudden abnormal hypersynchronous electrical discharges of neurons located predominantly in the central nervous system. Clinical manifestations of epilepsy comprise loss of consciousness, loss of sensation, peculiar behaviour and convulsions or a combination of these. Practical definition of epilepsy states that incidence of at least two unprovoked seizures occurring 24 hours apart or one unprovoked seizure with a probability of further seizures similar to the general recurrence risk after two unprovoked seizures, occurring over the next ten years (Fisher et al. 2014).

Forty-six million people worldwide are burdened with epilepsy (Beghi 2020). Of these, 24 million have active (repeated seizure occurrence or those with treatment requirement) epilepsy. The prevalence of active epilepsy is slightly higher in men, where it is 329 per 100,000 population, and is 319 per 100,000 population in women (Beghi 2020). The prevalence of epilepsy increases with age, ages between 5-7 years and those above 80 have the highest number of epileptic cases (Beghi 2020). The mortality rate of active epileptic cases is 1·74 per 100,000 population: in women it is 1·40, and in men, it is 2·09. Mortality is either due to SUDEP (Sudden Unexpected Death in Epilepsy) or indirectly due to seizure-related accidents such as drowning, asphyxiation, etc. Epilepsy incidences are higher in lower- and middle-income countries over higher-income countries, which is 139 vs. 49 per 100,000 person-years, reflecting differences in risk of infections, population structure, and access to healthcare. About 80% of active epilepsy cases are present in low-/middle -income countries (GBD 2019, Beghi 2020).

The age of the patient influences the aetiology of epilepsy. In newborns, perinatal hypoxia and ischemia, intracranial haemorrhage, and trauma, maternal neurotoxic drug-use triggers seizures or epilepsy, in addition to other factors that they share with other age groups. Metabolic defects such as inborn errors of metabolism cause epilepsies in addition to causing other comorbidities. Structural defects that occur due to developmental cortical malformation, hippocampal sclerosis, tuberous sclerosis, brain tumours, and traumatic brain injury lead to seizure activity. Infections that affect the central nervous system, such as bacterial

meningitis, viral encephalitis, neurocysticercosis, and cerebral abscess, increases the risk of contracting epilepsy. Autoimmune disorders such as multiple sclerosis, systemic lupus erythematosus, celiac disease, and Rasmussen encephalitis have epileptic seizures as one of the clinical manifestations. Certain endocrine disorders, haematological disorders, and several other systemic diseases may also cause seizures over a broad age range. Advances in sequencing and diagnostic technology have identified mutations associated with epileptic disorders. 15% of epileptic patients are family history positive. Epilepsy syndromes not associated with any other phenotypic abnormality other than seizures are majorly due to mutations influencing ion channels. Symptomatic (structural/metabolic) epilepsies were found to be associated mostly with mutations in genes involved in nervous system development (Shorvon, Andermann and Guerrini 2010, Kasper et al. 2015, Sazgar and Young 2019).

The most recent classification proposed by ILAE in 2017 is a revision of the 1989's categorization, whose primary purpose was clinical diagnosis, but this has greatly influenced disease research, drug development, and scientific communications. The aspects considered for classification are as follows, seizure types based on onset include generalized, focal, and unknown origins; etiological factors that include structural, infections, metabolic, immune, genetic, and unknown factors; and comorbidities that comprise learning, psychological, and behavioural defects. Classification is performed at three levels, the first being the seizure type followed by the diagnosis of epilepsy type that includes focal epilepsy, generalized epilepsy, combined focal and generalized epilepsy, and unknown epilepsy based on epilepsy seizure type, aetiology, and comorbidities (Figure 1). This finally leads to the diagnosis of an epilepsy syndrome (Fisher et al. 2017, Scheffer et al. 2017).

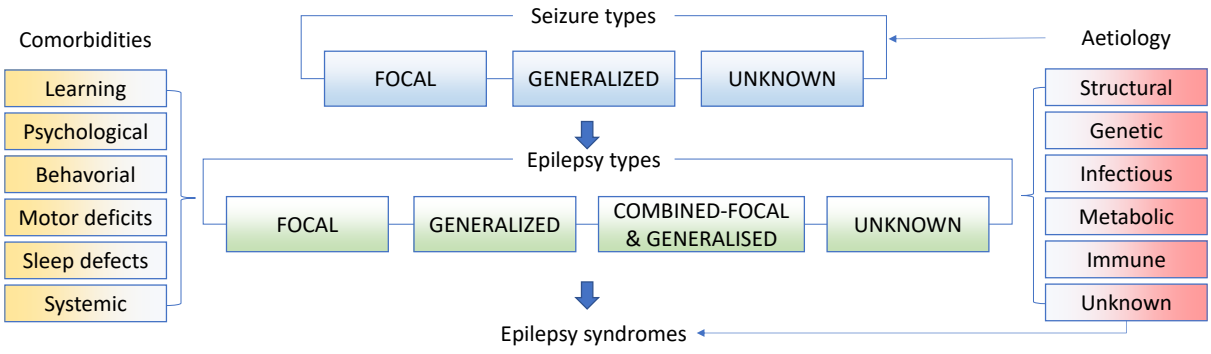


Figure 1: ILAE 2017 Classification (modified from Fisher et al. 2017)

Seizures are initiated in neurons due to high-frequency burst of action potential caused due to long-lasting depolarization of neuronal membranes by an influx of extracellular Ca^+ ions resulting in an influx of Na^+ ions leading to repeated hyperpolarization. This is followed by afterpotential hyperpolarization that involves GABA and K^+ channels. Neighbouring neurons also lose their ability to inhibit this burst due to the accumulation of K^+ and Ca^+ ions extracellularly and activation of NMDA receptors leading to propagation of the seizures. The extent and type of seizing activity is a result of several factors. Intrinsic factors include ion channel-numbers, -types, and -distribution that affect conductance, activation of secondary messenger system, protein expression changes determined by gene transcription, translation, and post-translational modification. Extrinsic factors include neurotransmitter concentration and type, extracellular receptor modification, spatial and temporal factors, neuronal networks, and properties of synaptic junctions, and the role of non-neuronal cell types such as astrocytes and oligodendrocytes. The basic mechanism behind the role of triggers such as sleep, fever, alcohol deprivation is not clearly understood, while rudimentary knowledge is available on seizure initiation and propagation (Shorvon et al. 2010, Kasper et al. 2015).

Epileptogenesis is the process by which the triggers leading to cortical hyperexcitability further lead to full-blown epileptic convulsions. The timeline between the initiating factors and the consequence can vary due to delays in alterations of neuron's properties, reduced inhibitory activity threshold, etc. Neuronal networks in thalamic-cortical regions are sites for initiation of generalized seizures and are mostly genetic in origin. In focal seizures, hemisphere-specific limbic and neocortical neurons are the origin sites and are usually caused due to structural abnormalities. Absence seizures are usually due to excessive inhibition of neuronal electric discharge, unlike other seizures (Shorvon et al. 2010, Kasper et al. 2015, Thijs et al. 2019).

Antiepileptic drugs can control seizures. In several cases, lifelong treatment is required, while in up to 70% of adults, antiepileptic therapy is discontinued once seizures are completely controlled. Medication is prescribed, keeping in mind the age, sex, types of seizures, clinical comorbidities, effect on precipitating factors, efficacy, toxicity, and interaction of drugs. Antiepileptic drug therapy is begun once epilepsy diagnosis is made or when epileptogenic factors such as lesions, infections, or trauma have been identified. Although several drugs have been developed to counter seizures, old medications are often used as the first line of therapy that includes valproic acid, phenytoin, carbamazepine, and ethosuximide. Twenty percent of patients with epilepsy are resistant to antiepileptic drug

therapy. Surgical procedures that include lesionectomy, region-specific lobotomy, transections to disconnect cortical neurons to prevent the spread of seizure activity are conducted once a surgical evaluation is done, and when the use of drugs is deemed unsuccessful. Vagus nerve stimulation by placing an electrode is a relatively recent treatment option for patients who are not candidates for surgical procedures (Thijs et al. 2019, Kasper et al. 2015).

1.1 Genetic generalized epilepsy

Genetic generalized epilepsy (GGE), earlier known as idiopathic generalized epilepsy, contributes to about a quarter of all epileptic cases. These are characterized mostly by bilateral generalized tonic-clonic seizures, myoclonic jerks, and absence seizures, while febrile, tonic, and atonic seizures are also observed. The seizures are reflected in electroencephalograms as generalized poly spike and wave discharges, suggesting a synchronized hyperexcited state of thalamocortical circuits (epilepsydiagnosis.org). The seizure activity occurs in both cerebral cortex hemispheres. In earlier times, the patient's cognition, intelligence, and brain imaging were considered normal. However, with advances in imaging technology, volumetric differences in various brain subregions in GGE cases, when compared to healthy individuals have been identified, while a small number of patients also exhibit temporary or permanent cognitive impairment. The risk and frequency of status epilepticus are ambiguous. The most common form of status epilepticus is absence status, which is prolonged, generalized, nonconvulsive seizures with variable consciousness levels. GGE patients with persisting GTCS due to abnormal lifestyle factors and poor compliance to therapy have an increased SUDEP risk. The antiepileptic medications effectively control seizures, although the remission rate varies across the different GGE syndromes (Kay and Szaflarski 2014, Nuyts et al. 2017, Nilo, Gelisse and Crespel 2020). The genetic component plays a crucial role in the aetiology of the disease. No clear-cut inheritance model has been accepted due to inconsistency in the presentation of epilepsy in family members. Both single gene and copy number variations have been identified to lead to genetic generalized epilepsy (Mullen and Berkovic 2018, Guerrini, Marini and Barba 2019).

Based on the age of onset and types of seizures presented, classic syndromes of GGE's include childhood absence epilepsy (CAE), juvenile absence epilepsy (JAE) and juvenile myoclonic epilepsy (JME), and epilepsy with generalized tonic-clonic seizures alone (EGTCS). These are also the most common forms of GGE. Rarer forms include generalized

epilepsy with febrile seizures plus (GEFS+), severe myoclonic epilepsy of infancy (SMEI), benign myoclonic epilepsy in infancy (BMEI), epilepsy with myoclonic absences (EMA), epilepsy with Eyelid Myoclonia and myoclonic-astatic epilepsy (MAE) (Table 1). This thesis's focus is the genetics of juvenile myoclonic epilepsy.

Table 1. Genetic generalized epilepsy syndromes

Syndromic diagnosis	**Age group**	**Seizure types**	**Genetics**	**Special features**
Severe myoclonic epilepsy of infancy (SMEI), Dravet syndrome	6 - 15 months	**Hemiclonic**, **GTCS**, Absence, Myoclonic, Atonic	*SCN1A, PCDH19*, mostly *de novo*	Drug-resistant, Cognitive and behaviour impairments, Considered EE
Benign myoclonic epilepsy in infancy (BMEI)	6 months to 2 years	**Myoclonic**, Febrile, GTCS	Unknown	Photic, noise or touch stimulation of myoclonic seizures, Self-limiting, Male preponderance
Generalized epilepsy with febrile seizures plus (GEFS+)	6 months to 6 years	**Febrile**, Myoclonic, Absence, Atonic, GTCS	Mendelian and complex/polygenic inheritance *SCN1A, SCN1B, GABRG2*,and *PCDH19.*	Febrile seizures resolve by puberty.
Epilepsy with myoclonic-atonic seizures, Myoclonic astatic epilepsy (MAE), Doose Syndrome	6 months to 6 years	**Myoclonic-atonic** Febrile, Absence, Tonic, GTCS	Complex/polygenic inheritance with variable penetrance, *SCN1A*,and *SLC2A1*.	Male predominance, Considered EE, Status epilepticus is common
Epilepsy with myoclonic absences (EMA)	1 to 12 years	**Myoclonic absence**, GTCS, Absence	Complex/polygenic inheritance, *SYNGAP1*	Male predominance, Frequent daily myoclonic absence seizures, Learning disability
Epilepsy with Eyelid Myoclonia, Jeavons syndrome	2 to 14 years	**Eyelid myoclonia**, Febrile, Myoclonic, Absence, GTCS	Complex/polygenic inheritance	Female predominance, Photosensitivity in all cases, Daily eyelid myoclonia induced by eye closure
Childhood absence epilepsy (CAE), Pyknolepsy	2 to 12 years	**Absence**, GTCS	Complex inheritance, *SLC2A1*, *GABRG2* and *CACNA1A*	Multiple daily absence seizures, Provoked by hyperventilation
Juvenile absence epilepsy (JAE)	8 to 20 years	**Absence**, GTCS	Complex inheritance, *GABRG2*,and *CACNA1A*	GTCS occur within 30min of awakening
Juvenile myoclonic epilepsy (JME)	8 to 25 years	**Myoclonic**, Absence, GTCS	Mendelian and complex/polygenic inheritance, *GABRA1, EFHC1, ICK, CASR, GABRD*,and *CACNB4*	Photosensitivity is common, Myoclonic seizures mostly on awakening
Epilepsy with generalized tonic-clonic seizures alone (EGTCS)	5 to 40 years	**GTCS**	Complex/polygenic inheritance, *CLCN2*	Provoked by sleep deprivation

Bold – Mandatory seizures, EE – Epileptic encephalopathy, GTCS – Generalized tonic-clonic seizures.

1.2 Epilepsy consortia studies

With the availability of NGS technologies, our abilities to sequence and study complete exome and genome sequences have been substantially enhanced. This has become a powerful tool to comprehensively study genetic variants contribution to a disorder in a single patient or in a large cohort. One of the earliest studies exploring genetic architecture of GGE was conducted for 118 European GGE patients by whole exome sequencing, and screening for potential epilepsy variant found in 878 GGE cases, but it did not identify any significant GGE-associated variant or pathways (Heinzen et al. 2012). Among the top-ranking variants, heterozygous *GREM1* p.Pro35Ala was the most enriched GGE associated variant located near the epilepsy CNV hotspot 15q13.3, while heterozygous missense variants in *PSME2, BTD* and *PEX6* and homozygous variant in *AGPAT3* were absent in controls and found in more than three epilepsy cases (Heinzen et al. 2012). Targeted Sanger sequencing of 237 channel genes in 152 GGE cases also did not identify any significantly associated variant or gene or gene family (Klassen et al. 2011). Combination of common and rare variants in complex patterns were inferred to contribute to disease causation (Klassen et al. 2011). Whole exome sequencing of 152 European GGE patients did not identify enrichment of any variant while gene set analysis identified missense variants in $GABA_A$ receptor genes to exhibit significant enrichment that was recapitulated in two other cohorts of 357 and 583 GGE European patients. These variants were equally distributed among the $GABA_A$ receptor genes. Function analysis revealed four ultra-rare, segregating, missense variants in *GABRB2* and *GABRA5*, previously unreported to cause epilepsy, to exhibit defective receptor function (May et al. 2018).

Given the complexity and heterogeneous genetics underlying epilepsy, world-wide collaborative studies were set up to explore epilepsy-associated genes, genetic variants, phenotypes in very large sample sizes. The first study analyzed 4000 epilepsy genomes, which included epileptic encephalopathy, genetic generalized epilepsy and focal epilepsy patients from multi-generational families, pairs, and sporadic individuals (Epi4KConsortia 2012). Analysis of familial genetic generalized epilepsy patients of European ancestry by a case- control study by exome sequencing identified enrichment of ultra-rare variations in known 43 dominant epilepsy and 33 epileptic encephalopathy genes (Epi4K Consortium 2017a). No single gene was study-wide significantly associated with the disorders, while the top ten ranking genes with functional variants included three known epilepsy genes (*KCNQ2*, GABRG2, and SCN1A) and the rest were ATP1A3,

CACNA1B, COPB1, KEAP1, SLC9A2. Variants with MAF > 0.1% (from ExAC and EVS database) were concluded to not contribute to familial GGE risk (Epi4K Consortium 2017a). Analysis of phenotypic elements of familial GGE patients from the Epi4K collections revealed several strongly heritable features particularly age of onset which was predominantly during adolescence (Epi4K Consortium 2017b, Ellis et al. 2019, Ellis et al. 2020). Absence epilepsy did not segregate strongly with JME but did so with MAE. Female preponderance was observed, while no "maternal effect" was identified (Epi4K Consortium 2017b).

A major study that aimed to cover 25000 epilepsy genomes began with analysis of 9170 affected samples that included 4,453 GGE patients of European descent by whole exome sequencing (Epi25 Collaborative 2019). No single gene exhibited study-wide statistical significance while known or potential epilepsy associated genes exhibited higher burden of ultra-rare variants whose allele count is not more than three. Ultra-rare nonsynonymous variants, notably protein truncating variants in GGE cases were largely present in genes intolerant to loss of function changes. Missense ultra-rare variants were at least 2-fold enriched in GGE patients over controls in known epilepsy genes and brain expression enriched genes (Epi25 Collaborative 2019). Singleton ultra-rare variants in *CACNA1G* and *UNC79* were top ten ranking GGE-associated, brain-enriched and novel epilepsy genes that function as ion channels or ion channel interactors. The rest top ranking genes with singleton variants were in *EEF1A2, GABRG2, ALDH4A1, SLC6A1, RC3H2, GABRA1, DNAJC13 and ZBTB2* (Epi25Collaborative 2019*)*. $GABA_A$ receptor subunits encoding genes exhibited higher burden of protein- truncating variants among GGE patients, while missense variants burden was elevated in $GABA_A$ receptor subunits, voltage-gated cation channel encoding genes and GABAergic pathway genes but not in ionotropic glutamate receptor subunits and nicotinic acetylcholine receptor subunits encoding genes. The GABA receptor subunit variant burden reiterates an earlier finding in an exome-based case-control study (May et al. 2018, Epi25 Collaborative 2019). Top ten ranking GGE associated genes with up to three variant allele counts were *EEF1A2, UNC79, RC3H2, GABRA1, GABRG2, IRAK3, F9, PARVG, ALDH4A1* and *SLC6A1* (Epi25 Collaborative 2019). The variant burden and genic association to GGE recapitulates the findings by the Epi4K study which had over representation of familial cases whereas the Epi25K study were mostly sporadic (Epi25 Collaborative 2019). Contribution of common genetic variants to GGE risk was estimated by polygenic risk scores which were significantly higher in GGE patients over control subjects

and focal epilepsy cohort of non-Finnish European ancestry (Leu et al. 2019). Polygenic risk scores in Finnish GGE cases were significantly higher over healthy subjects while the phenotypic variance was not efficiently justified by the scores when compared to non-Finnish European cohort. GGE risk in Japanese epilepsy cohort could not be estimated by polygenic risk scores (Leu et al. 2019). Copy number variations at epilepsy associated hotspots exhibited significant association to GGE unlike other epilepsy syndromes studied (Niestroj et al. 2020). More than 2 Mb deletion CNVs at epilepsy hotspots 16p13.11 and 22q11.2, as well as smaller CNVs at 15q13.3 and 16p13.11 were enriched in GGE cases. More than 2 Mb duplications were enriched in GGE cases with febrile seizure over GGE cases without febrile seizures. Network analysis of genes affected by CNVs in GGE cases identified *APP*, *SUMO3*, and *UBE3A* to centrally connect other genes. These three genes have not been previously associated with epilepsy but are known to cause neurological defects in presence of pathogenic variations (Niestroj et al. 2020).

An extension of Epi25k collaborative study involved additional 13,171 epileptic cases that included 5303 GGE patients that also included samples with non-European ancestry (Epi25 Collaborative 2021). Top ten ranking GGE associated genes that were ranked for genes with more than one ultra-rare pathogenic variant included four known epilepsy genes, *SLC6A1, SCN1A, GRIN2A, GABRA1* while the rest were *FBXO42, DAW1, NUP98, KCNK18, ZNF324B* and *C1orf112* (Epi25 Collaborative 2021). Only *SLC26A1, SCN1A* and *GABRA1* overlapped with top ten gene ranks from Epi4K and the first Epi25K findings. Ultra-rare missense variants in known epileptic encephalopathy genes particularly in sub-genic intolerant regions as defined by MTR scores were enriched in GGE patient cohort. ClinVar ultra-rare pathogenic variants were also enriched in GGE patients especially protein truncating variants and the missense variants located in sub-genic intolerant regions. Protein truncating variants, not missense variants were enriched in GGE cases in loss of function variant intolerant genes unreported to be associated with any disease, defined as those not reported in OMIM database. *NLGN2, HDLBP, RC3H2* and *XPO5* were top candidates for LOF intolerant, non-OMIM, GGE enriched protein truncating variants (Epi25 Collaborative 2021).

The findings of a few large-scale consortia studies so far indicate that no single gene has study- wise significance with respect to burden of rare pathogenic variants. This reflects the enormous complexity and heterogenous genetic background underlying epilepsy pathology and suggests requirement of even higher number of samples to achieve statistical

significance. As expected, several known epilepsy genes have been identified, while a few genes not previously associated with epilepsy or neurological phenotypes have also been highlighted suggesting that several novel genes are yet to be discovered. Milder epilepsy phenotypes have missense variant burden in genes associated with severe epilepsy phenotypes caused usually due to protein truncating variations, connecting the severity of the phenotype with the type of variation. No clear phenotype-genotype corelation could be made from these studies. The difference in top ranking epilepsy associated candidate genes across the studies is primarily due to difference in the analysis methodologies. The major limitation of these studies is lack of non-European samples.

1.3 Juvenile myoclonic epilepsy

1.3.1 History

JME was first described in 1867 by Théodore Herpin, a Parisian clinician, in a 14-year-old boy who presented myoclonic jerks. The next report was described in a thesis where five patients were diagnosed with JME clinical manifestations in 1899 by Leon Rabot. Janz and Christian in 1957 proposed that these clinical presentations be a distinct epileptic disorder along with identifying specific precipitating factors, on studying 45 epileptic cases and naming it *impulsive petit mal* due to which JME is also known as Janz syndrome. Lund first coined the term Juvenile myoclonic epilepsy in 1976. In 1982, Gastaut referred to JME as *Herpin–Rabot syndrome* after authors described epileptic patients who presented JME clinical symptoms described in essays titled *De la myoclonie épileptique* and *Des access incomplets D'epilepsie.* Delgado Escueta described this disease as juvenile myoclonic epilepsy of Janz in 1984. ILAE's 1989 proposal of classification of epileptic disorders termed it juvenile myoclonic epilepsy. ILAE's 2017 revision of the 2003's classification retained the term juvenile myoclonic epilepsy and classified it under genetic generalized epilepsy since greater than 50% of JME cases are family history positive, indicating genetic predisposition (Genton and Gelisse 2013, Yacubian 2017).

1.3.2 Epidemiology

The incidence of JME is 1 in 100,000 people, while the prevalence ranged from 0.1 to 0.2 per 1000 people. The occurrence of JME varies across countries, but generally accepted values lie between 1-10% among all epilepsies, and among generalized epilepsies, the value is about 18% (Jallon and Latour 2005). The prevalence varies from 3% (population-based

prevalence) to 12% (hospital/clinic-based prevalence) of all epilepsies (Nicoletti et al. 1999, Delgado-Escueta 2007). Literature suggests both male and female preponderance based on population and time of the study. Recent studies indicate higher representation among women up to 60% (Camfield, Striano and Camfield 2013, Christensen et al. 2005). 15-18% of CAE evolves into JME (Wirrell et al. 1996, Martínez-Juárez et al. 2006).

1.3.3 Seizure types and reflex traits

Three main seizure subtypes are observed in JME cases. Generalized tonic-clonic seizures (GTCS), earlier known as 'grand mal seizures' are characterized by tonic muscle contractions leading to clonic rhythmic jerking due to alternating muscle contraction and relaxation, resulting in unresponsiveness and muscle flaccidity throughout one's body. Myoclonic seizures or jerks are portrayed by an increase in muscle tone resulting in sudden and brief contractions in one or several parts of the body. Absence seizures are also known as 'petit mal seizures' are characterized by short temporary loss of consciousness without any changes in postural control. Although no other major clinical manifestation is observed, subtle signs of rapid blinking, chewing, or arms movement can be seen. Early morning, bilateral myoclonic jerks are the flagship seizures for JME diagnosis. GTCS is found to occur in 80-95% of JME cases. Absence seizures occur in about 31.9% of the patients (12.8% begin in the first decade of life and 19.1% exhibit juvenile-onset) (Panayiotopoulos, Obeid and Waheed 1989, Usui et al. 2005, Yacubian 2017). Reflex traits are also known to be associated with JME. They are photosensitivity (PS) present in 25-38%, eye closure sensitivity (ECS) in 15-20%, praxis induction (PI) in 30-50% and orofacial reflex myoclonia (ORM) is 26-40% of cases (Mayer et al. 2006, Guaranha et al. 2009, Guaranha et al. 2011, Yacubian and Wolf 2014, Yacubian and Wolf 2015, Gelžinienė, Endzinienė and Jurkevičienė 2015, Millichap and Millichap 2015, Wolf et al. 2015, Baykan and Wolf 2017).

1.3.4 Age of onset

Seizures begin during adolescence, around puberty, where myoclonic jerks are the first of the seizures to occur; after a gap of 3.3 years, there is a recurrence of seizures. In 25% of cases, GTCS is the first seizure to occur, while in 33% of cases, GTCS and myoclonic jerks co-occur. Absences occur first in about 15% of the cases. Myoclonus, usually the onset symptom, occurs around 12 - 14 years of age among girls and around 14 - 16 years of age among boys. In cases of CAE evolving into JME, the onset of absence seizures is around 4.5

± 2.5 years (Panayiotopoulos et al. 1989, Panayiotopoulos, Obeid and Tahan 1994, Yacubian 2017).

1.3.5 Precipitating factors

Several factors have been identified to trigger seizures in JME patients. Sleep deprivation contributed the highest risk in developing seizures either by itself or in combination with other factors, followed by menstruation in women. Stress caused by both mental and physical activities amounted to the third-highest risk contribution. Photic stimulation by viewing videos or flashing lights also triggered seizures. Other factors that promoted seizures include sudden auditory stimuli, alcohol intake or withdrawal, drug abuse, activities requiring concentration, complex mental calculations or decision-making thoughts and, sudden awakening or element of surprise. These risk factors' contribution varies with studies conducted across the world (Murthy, Rao and Meena 1998, da Silva Sousa et al. 2005).

1.3.6 Psychosocial and cognitive dysfunctions

Behavioural and psychiatric defects such as mood swings, anxiety, forgetfulness, etc. are observed in JME patients. Their contribution ranges from 25-75%. Cluster B personality disorders (PD), which consist of emotional instability, immaturity, and lack of discipline identified by MRI studies, also indicated prefrontal cortical abnormalities in JME patients. Varying aspects of cognitive function are affected in JME patients, such as learning difficulties, verbal fluency, and working memory, including impulsive decision-making capacity and reasoning. However, intellectual capacity is found to be unaltered and behavioural defects also vary with age. With disease progression, the severity of the symptoms presented increase. It is yet unclear whether personality disorders are a cause or effect of JME (de Araújo Filho et al. 2007, Wolf et al. 2015, Gilsoul et al. 2019).

1.3.7 Neuropathology

The presence of structural and physiological abnormalities in the brain of JME patients is debatable. Studies reporting both the presence and absence of differences when matched with their control samples, are available. Brain autopsy studies have uncovered alterations in the subcortical white matter showing increased neuronal density. A reduced number of axons have been observed in the anterior horn of the spinal cord. Structural MRI studies using voxel-based morphometry (VBM), surface-based morphometry, and diffusion tensor imaging (DTI) have indicated abnormal cortical layers thickening or thinning and volume reduction in

the thalamus and hippocampus. White matter structural integrity was modified, affecting connectivity in the corpus callosum and with the primary motor cortex. Functional MRI studies have identified dysfunctional connectivity between motor and frontal cognitive neural networks (Gilsoul et al. 2019).

Molecular MRI and PET studies in JME patient brains have identified alteration in levels of several metabolites. N-acetyl aspartate (NAA) in the prefrontal cortex, frontal lobe, thalamus, and hippocampus was found to be reduced. Brain glutamate Glx (Glutamine and Glutamate) was found to be increased in the thalamus, striatum, and insula and reduced in the medial prefrontal cortex and the primary motor cortex. Inhibitory neurotransmitter GABA (γ aminobutyric acid) concentrations were reduced in the thalamus and increased in the prefrontal cortex. Dopamine transporter (DAT) binding sites numbers in the midbrain were reduced. Dopamine D2/D3 receptor binding sites have been demonstrated to be reduced in the posterior putamen, while serotonin 1A receptor binding was decreased in the dorsolateral prefrontal cortex, the raphe nuclei, and the hippocampus (Gilsoul et al. 2019).

1.3.8 Management and treatment

The main course of treatment is antiepileptic drugs (AED). Medications are selected such that they do not aggravate seizure promoting factors, interact with other medication prescribed for other underlying clinical conditions, and do not have harsh side effects. Other criteria for drug selection, in addition to efficacy, include types of seizures presented, the ability of the patient to follow the medication routine, dosage scheduling convenience and amount, use of contraception, plans for pregnancy, and previously used drugs (Mantoan and Walker 2011, Kasper et al. 2015).

Valproic acid (VPA) is the first drug of choice to control seizures and was also reported to be used in the earliest reported JME cases. In 85-95% of JME cases, valproate controls seizures. Its mechanism of action is unclear to date, although it is known to affect the metabolism of GABA, leading to an increase in inhibitory neurotransmitter GABA in the brain. Side effects include ataxia, sedation, gastrointestinal irritation, hair loss, hepatotoxicity, thrombocytopenia, hyperammonemia, and weight gain. VPA also is severely teratogenic and is not prescribed for women belonging to the child-bearing age group. It is ineffective in patients presenting the triad of seizures, i.e., GTCS, myoclonic jerks, absence seizure, and those with psychiatric complications. VPA is the first choice of AED in men with JME (Gelisse et al. 2001, Marson et al. 2007, Mantoan and Walker 2011, Kasper et al. 2015).

Prior to the availability of VPA, phenobarbital (PB) and primidone were prescribed that controlled seizures in 86% of the patients, while phenytoin (PHT) was effective in controlling seizures in 67% of the cases. Post the 1990s; several drugs have been developed to address the drawbacks of valproate. Pharmacokinetics and toxicity of the drugs are considered while prescribing drugs to JME cases. For women of child-bearing potentials, levetiracetam or lamotrigine are prescribed. Topiramate and zonisamide anticonvulsants are teratogenic and have side effects, including psychiatric problems and weight loss. Benzodiazepines such as clonazepam, clobazam, and carbamazepine, oxcarbazepine are drugs used in combination with other AEDs to improve the overall efficacy of the medication (Mantoan and Walker 2011, Crespel et al. 2013, Landmark et al. 2019b, Serafini et al. 2019, Silvennoinen et al. 2019).

The overall prognosis in JME is good. In addition to medication, several other factors are important for controlling seizures. Proper lifestyle choices that include proper sleep-wake cycle, avoiding alcohol consumption, and avoiding drug abuse are vital to prevent seizure recurrence. Compliance with medication is also crucial in controlling seizures. A sudden stoppage and change in AED consumption can reduce the medication's effect and may instead increase the occurrence of seizures. Therapeutic drug monitoring (TDM) is useful in drug adherence and individual precision medication. TDM studies have identified extensive pharmacokinetic variability between JME patients (Landmark et al. 2019a). Psychotherapy is beneficial in patients with social and cognitive difficulties. With age, the AED's' side effects and the triggering factors fade, increasing the efficacy of the medication. Lifelong consumption of AEDs in certain cases is discontinued when there is complete control over seizures (Crespel et al. 2013, Landmark et al. 2019b).

1.4 Genetics of JME

The genetic contribution to the causation of JME has been well established. Up to 65% of all JME cases were found to be family history positive where both clinical manifestations and EEG patterns were inherited over generations (Panayiotopoulos et al. 1994, Renganathan and Delanty 2003, Jayalakshmi et al. 2006). First-degree family members have a 6% higher risk to present epilepsy, which is doubled in cases where absence seizures are also reported (Vijai et al. 2003a, Pal et al. 2006). In twin studies, concordance rates between monozygotic pairs were significantly higher than dizygotic pairs (Kjeldsen et al. 2005, Vadlamudi et al. 2014). Although these studies have indicated that a genetic component is involved in the causation

of JME patients, these have not been identified for majority of the cases due to both complex inheritance patterns and the non-availability of genomic technology. Linkage analysis, association studies, candidate gene sequencing, and next-generation sequencing studies have identified 29 JME associated loci, amongst which six genes have been identified. These are described in the following sections.

1.4.1 Genes identified by linkage studies

Linkage mapping using DNA polymorphisms, especially microsatellites as markers have been used to trace the transmission of chromosomal regions over generations in families. The identification of critical recombination events helps demarcate the critical region of interest. In general, higher the number of meiotic events, narrower the interval that denotes the smallest possible region associated with a trait. Large multi-affected families affected with JME have been utilized for genome-wide linkage analyses to identify causative genes such as *GABRA1*, *EFHC1*, *CASR*, and *CILK1*, which post-identification have been screened in JME cohorts to strengthen the evidence of the gene's involvement in the disorder. Linkage mapping uses an unbiased approach without any preconception. Hence, it does not require a prior knowledge of the biology or physiology of the trait to study its inheritance, contributing to its advantage over other methods. The discovery and role of the identified genes are described in the following section.

1.4.1.1 GABRA1

Genome-wide linkage analysis in a French-Canadian family afflicted with JME identified the marker with the highest LOD score of 3.1 at locus 5q34, where the mutation p.Ala322Asp in *GABRA1* was identified. This variation was absent in controls and in 83 sporadic GGE patients that included 21 JME cases (Cossette et al. 2002). *GABRA1* variations c.248+1G>T and p.Phe104Cys were identified in JME cases of European origin and p.Lys353delin18X and p.Asp219Asn in IGE cohort of French-Canadian background (Lachance-Touchette et al. 2011, Johannesen et al. 2016). *GABRA1* was the first gene to be identified to be associated with JME. Several other studies involving *GABRA1* gene sequencing identified no other mutations in JME patients (Kapoor et al. 2003, Ma et al. 2006b), while variations were identified in GGE patients, including CAE cases (May et al. 2018). Phenotypic differences have also been observed in monozygotic twins diagnosed with generalized epilepsy having the same *GABRA1* missense mutation p.Pro181Ser (Krenn et al. 2019). *GABRA1* variants have been identified in several other types of epilepsy, including Dravet syndrome, Epileptic

encephalopathy, Generalised epilepsy with febrile seizures, Infantile spasms, Ohtahara and West syndromes (Carvill et al. 2014, Johannesen et al. 2016, Kodera et al. 2016, Farnaes et al. 2017, Epi25 Collaborative 2019, Steudle et al. 2020).

γ -aminobutyric acid receptor subunit alpha-1 (*GABRA1*) is a ligand-gated chloride channel, a subunit of the heteropentameric GABA receptor, a major inhibitory neurotransmitter in the brain. GABRA1 is highly expressed in most brain regions, especially in the cerebral cortex, cerebellar sections, hippocampus, and synaptic regions. It binds to benzodiazepines or GABA neurotransmitters causing phasic inhibition (McKernan et al. 1991, Jacob, Moss and Jurd 2008). Homozygous *Gabra1* knockout mice exhibit abnormal behaviour and synaptic transmission, defective reactions to benzodiazepines, reduced life span, and essential tremors (Kralic et al. 2005). Heterozygous *Gabra1* knockout revealed alteration in the GABA receptor composition leading to a reduction in GABAergic synaptic current amplitude and an extended current rise and decay time (Zhou et al. 2013). Zebrafish *gabra1* knockouts exhibited seizures in juvenile stages that could also be induced by photic stimulation. They also exhibited defective and reduced inhibitory synaptic networks during development (Samarut et al. 2018). Zebrafish *gabra1* knockdown models exhibited reduced motility, which was partially rescued by introducing wild-type morpholinos and was not rescued in missense mutation p.Thr292Ile morpholino experiments (Reyes-Nava et al. 2020). Human epilepsy *GABRA1* mutations replicated in mammalian cells and mice presented defective postsynaptic currents affecting amplitude and decay time, absence like seizures at juvenile stages that continued and expanded to myoclonic seizures, increase in pyramidal cell spine's density, and defective GABA receptor composition (Cossette et al. 2002, Fisher 2004, Gallagher et al. 2004, Ding et al. 2010, Arain, Boyd and Gallagher 2012, Lachance-Touchette et al. 2014, Arain et al. 2015, Bai et al. 2019). The presence of altered synaptic functions and presentation of seizures in *Gabra1* mutants and null animal models explain the phenotype presented by JME patients with *GABRA1* mutations.

1.4.1.2 EFHC1

Mutations in *EFHC1* were first identified in JME patients in a Mexican family with the highest LOD scores at the 6p12-11 locus known to be linked to JME and named EJM1. Forty-four additional JME positive families from Belize and Los Angeles who also presented significant linkage at 6p12-11 were sequence analyzed for *EFHC1* identifying several other missense mutations (Liu et al. 1995, Liu et al. 1996, Suzuki et al. 2004, Bai et al. 2002).

Additional *EFHC1* mutations were also identified in other JME, IGE and TLE patients in populations from Mexico, Honduras, Japan, Austria, Italy, Turkey, and India (Stogmann et al. 2006, Annesi et al. 2007, Medina et al. 2008, Jara-Prado et al. 2012, Raju et al. 2017, Thounaojam et al. 2017, Şirinocak et al. 2019). In contrast, variants were absent or rare in Dutch, Swedish, German, and United Kingdom populations (Ma et al. 2006b, Pinto et al. 2006, Gilsoul et al. 2019). In certain Hispanic and African JME cohorts, *EFHC1* SNPs and certain previously ascertained pathogenic variations were found in a healthy control population that has questioned the variation's actual contribution to JME (Bai et al. 2009, Subaran et al. 2015). *EFHC1* missense mutations have also been identified in a SUDEP and a primary intractable epilepsy in infancy case of Moroccon-Jewish ancestry (Berger et al. 2012, Coll et al. 2016).

EF-hand domain containing 1 earlier known as Myoclonin 1 codes for a calcium-binding protein that contains three DM10 and one EF-hand domains. It localizes at the spindle poles and midbody in mitotic cells using its N- terminal microtubule-binding region, and its absence leads to cell division and migration defects (de Nijs et al. 2006, de Nijs et al. 2009). Monomeric EFHC1's C- terminal region interacts with Ca^{+2} or $Mg^{+2,}$ which is abolished on dimerization (Murai et al. 2008). *Efhc1* is expressed ubiquitously, strongly in fallopian tubes and testis. In the mouse adult brain, it is expressed in neuronal cells in the cortex, striatum, hippocampus, cerebellum, and a few glial cells in the cortex. It is also expressed in the cilia of ependymal cells lining the ventricles, sperm flagella, and tracheal cilia (Ikeda et al. 2005). During mouse development, at E16, *Efhc1* is expressed in brain ventricles, in choroid plexus, and radial glial fibres of piriform and neocortex regions (Léon et al. 2010, Suzuki et al. 2004, Suzuki et al. 2008). Homozygous and heterozygous *Efhc1* mice knockouts were fertile, but the former exhibited enlarged brain ventricles with reduced cilia motility and smaller hippocampus, whereas these abnormal anatomical features were absent in the latter. Both mice exhibited spontaneous seizures with higher frequency in homozygote knockouts, whose susceptibility increased on chemical induction (Suzuki et al. 2009). In *Xenopus laevis,* efhc1 is expressed in ciliary axonemes in epidermal, gastrocoel roof plate, and neural tube cells but is absent in the basal body. Its knockdown leads to defective CNS patterns and neural crest formation due to disruption in Wnt expression and its signalling pathway (Zhao et al. 2016). In *Drosophila Efhc1* knockouts, the number of synaptic boutons at the neuromuscular junction synapse, terminal branching of dendrites, and spontaneous neurotransmitter release were enhanced (Rossetto et al. 2011). In *C. elegans*, *efhc-1* is present in both nonmotile cilia

and motile cilia. It is also present at presynaptic regions of dopaminergic neurons wherein it regulates its mechanosensation (Loucks et al. 2019). Mutations identified in JME patients exhibited mitotic defects in mammalian cells, abnormal radial and tangential migration of neuroblasts, and altered apoptotic activity in primary mouse hippocampal neurons (Suzuki et al. 2004, de Nijs et al. 2009, Katano et al. 2012, de Nijs et al. 2012, Raju et al. 2017). These findings indicate that loss of EFHC1 functions leads to mild defects in neuronal migration and synapse formation resulting in abnormal neuronal circuitry during cortical development that probably induces seizures in postnatal timelines (Grisar et al. 2012).

1.4.1.3 CASR

Linkage analysis in a three-generation GGE affected family from India identified the locus at 3q13.3-q21 to be associated with the disorder, with most of the affected individuals having JME. Candidate gene sequencing identified rare missense mutation p.Arg898Gln in *CASR* that segregated in the family. Several other missense mutations were identified on screening *CASR* in additional GGE patients (Kapoor et al. 2008). A patient with both intractable generalized epilepsy in addition to intellectual disability and hypocalcemia with missense mutation p.Phe788Cys in *CASR* has been reported (Rossi et al. 2019). *CASR* mutation p.Leu123Ser was identified in a patient with hypocalcemia who also presented seizures, cognitive impairment, and neuropsychological disabilities during childhood, while his mother exhibited mosaicism for the variant and was normal (Regala et al. 2015). Mutations in *CASR* have also been identified in patients with hyperparathyroidism, hypocalcemia, and hypocalcemia associated with hypercalcemia or Bartter syndrome (Thakker 2004, Hendy, Guarnieri and Canaff 2009, Vahe et al. 2017, Nissen and Rejnmark 2019). Calcium-sensing receptor CASR is a seven-transmembrane G protein-coupled receptor that senses fluctuation in extracellular calcium ions using its extracellular domain, thereby maintaining homeostasis (Gama and Breitwieser 1998, Ray et al. 1997, Bai 2004, Huang et al. 2009, Brown 2013). The intracellular domain participates in regulating its cell surface expression and signal transduction via interaction with various intracellular factors (Ray et al. 1997, Ward 2004, Huang and Miller 2007). CASR is expressed highly in the parathyroid glands and kidney and is also present in various brain regions with the highest expression in the hypothalamus and corpus striatum. Its punctate localization in nerve terminals probably indicates its role in regulating neurotransmitter release in synaptic regions in response to calcium levels (Brown et al. 1993, Ruat et al. 1995, Kapoor et al. 2008). CASR was found to regulate voltage-gated sodium channels (VGSC), Ca^{2+} activated potassium (BK) channels currents, and non-

selective cation channel (NSCC) activity modulating neuronal excitability (Chen et al. 2010, Vysotskaya et al. 2014, Mattheisen, Tsintsadze and Smith 2018). *Casr* null mice exhibit parathyroid hyperplasia, reduced growth, osteomalacia, and premature death. They also have high serum concentrations of parathyroid hormone and calcium, which is also observed in heterozygous knockouts with partial effects (Ho et al. 1995, Tu et al. 2003). Homozygous *Casr* and *Gcm2*/*Pth* double knock out mice did not have the phenotypic abnormalities exhibited by *Casr* only null mice (Kos et al. 2003, Tu et al. 2003). Nuf mice model which exhibited cataract, ectopic calcification, hypocalcemia, hyperphosphatemia, hypoparathyroidism was identified to have a missense mutation p.Lys723Gln in *Casr*, while heterozygous Nuf mice exhibited a milder phenotype (Hough et al. 2004). *CASR* mutations identified in GGE patients disrupted its intracellular retention that is naturally brought about by phosphorylation of arginine motif, which led to an increase in plasma membrane expression in mammalian cells affecting its downstream signalling pathways (Stepanchick et al. 2010). CASR was also found to regulate dendritic and axonal growth of the prenatal peripheral nervous system and postnatal hippocampal pyramidal neurons (Vizard et al. 2008, Jones and Smith 2016). CASR's functions concerning neuronal excitability and its probable impact on seizures mirror the findings of *CASR* mutations amongst GGE/JME patients.

1.4.1.4 CILK1

Linkage analysis of a 37 -member JME afflicted family of European/Amerind ancestry identified significant linkage at locus 6p12.2 with a maximum LOD score of 3.35. Whole exome sequencing of 6 individuals from the family identified the only nonsynonymous variant p.Lys305Thr at this locus to be present in the *CILK1* gene that also segregated with the disease in the family. Additional 22 rare *CILK1* variants were identified among 310 JME patients who were mostly of Hispanic, European - American Indian, and Japanese ancestry (Bailey et al. 2018). Rare *CILK1* variants were not identified in 1149 GGE, of which 357 were JME in the North American population of European ancestry, indicating population-specific risk factors involvement (Lerche, Berkovic and Lowenstein 2019). Homozygous lethal missense variations in *CILK1* have been identified in patients with endocrine-cerebro-osteodysplasia initially diagnosed as Majewski-hydrolethalus phenotype in old order Amish and Turkish families (Lahiry et al. 2009, Oud et al. 2016).

Ciliogenesis associated kinase 1 (CILK1), earlier known as intestinal cell kinase (ICK), is a serine/threonine kinase related to mitogen-activating protein (MAP) kinases family. It is

ubiquitously expressed with the highest expression in the spinal cord, testis, ovary, and brain (Nagase et al. 1999). In adult mice, the *Cilk1* transcript is highly expressed in lung and colon tissues, especially in intestinal crypt compartments, and lower expression in other tissues, differing from human expression pattern (Togawa et al. 2000). In developing mouse brain, Cilk1 is expressed in ganglion cells, retinal progenitor cells, ependymal cells lining the walls of the lateral ventricles, the choroid plexus, pyramidal cells in the hippocampus, neocortical cells, Purkinje cells in the cerebellum, telencephalon and cortical plate, intermediate zone, and ventricular and subventricular zones of the cerebral cortex (Bailey et al. 2018). *Cilk1* null mice presented displayed cleft palate, peripheral edema, hydrocephalus, polydactyly, delayed skeletal development, and embryonic lethality with elongated cilia and reduced Shh signalling during limb digit patterning (Moon et al. 2014). Brain-specific *Cilk* null mice had smaller cerebellum, hippocampal dentate gyrus, and exhibited ciliary defects in neuronal progenitor cells with Hedgehog signal defects (Chaya et al. 2014). Isoflurane sleep-induced heterozygous *Cilk1* null mice exhibited tonic-clonic convulsions and myoclonia (Bailey et al. 2018). EGFP tagged CILK1 localizes to the nucleus and the centrosome in mammalian cells. It is absent in motile cilia while present at the basal body in non-motile primary cilia (Yang, Jiang and Chen 2002, Oud et al. 2016, Bailey et al. 2018, Wang et al. 2020). JME *CILK1* variants exhibited defective radial migration, mitotic progression, cell cycle exit and apoptosis in neural progenitor cells and unrestricted ciliary length, reduced ciliation, mislocalization along axoneme of primary cilia in mammalian culture cells (Bailey et al. 2018, Wang et al. 2020). *CILK1* role in ciliogenesis involves its interaction with KIF3A, Scythe, Raptor, and GSK3β that affects intraflagellar transport, autophagy, mTOR, and hedgehog signalling, respectively (Fu et al. 2019). Genetic and functional studies performed on the *CILK1* variants using cellular and mice models revealed a loss of function effect explaining their roles in JME.

1.4.2 Candidate gene sequencing studies

An alternative approach to identify a gene's link to the disease either through mutation or allelic association is by exploring the sequence of notable genes. The gene to be sequenced is decided based on their location in the genome, especially if at a linked locus, mutations identified in the gene exhibiting similar phenotype, expression in tissue of interest, and its contribution to the relevant disease mechanism. In JME patient cohorts, several genes have been chosen to be screened for variations by both Sanger sequencing technique and next-generation sequencing techniques that include targeted gene panel and exome sequencing.

Mutations in *SCN1A*, a voltage-gated ion channel expressed in the brain has been identified in patients with GEFS+, and Dravet Syndrome hence was selected to be screened in JME patients (Moulard et al. 1999, Baulac et al. 1999, Escayg et al. 2000a, Claes et al. 2001, Carranza Rojo et al. 2011). Missense mutations in *SCN1A* have been identified in European, Turkish and Malaysian populations in JME patients (Escayg et al. 2001, Lal et al. 2016, Chan et al. 2020). *CACNA1G*, a low voltage-activated Ca(v)3.1 T-type calcium channel, was selected to be screened in JME patients since mutations in its associate subunit *CACNA1H* have been identified in CAE patients (Chen et al. 2003, Heron et al. 2007). *CACNA1G* mutations in Japanese JME patients were identified (Singh et al. 2007). *CHD2* mutations are known to cause epileptic encephalopathy in children, among whom several patients exhibited photosensitivity (Carvill et al. 2013). Photosensitivity being a seizure trigger for several GGE patients, *CHD2* was screened among these that identified one JME patient with *CHD2* missense mutation in addition to mutations in other types of GGE patients (Galizia et al. 2015). Candidate genes extensively researched in JME patients include *GABRD* and *CACNB4*, which are described in the following section.

1.4.2.1 GABRD

GABRD gene was sequenced in IGE, GEFS+, and FS patients since mutations in $GABA_A$ receptor genes that GABRD forms channels with, have been detected in GGE patients. A homozygous p.Arg220His allele was identified in one JME patient. This variation was commonly found in controls and was absent as a homozygous allele in other JME and IGE cases screened, indicating it was not a common factor in JME. However, mammalian cell based electrophysiological experiments indicated that this variant exhibited reduced current amplitude in both homo- and hetero-zygous conditions. Variants p.Arg220His, p.Glu177Ala, and p.Arg220Cys in heterozygous conditions were identified in other GEFS+ family individuals that segregated in a polygenic manner that were primarily of European descent (Dibbens et al. 2004, Lenzen et al. 2005c). *GABRD* variations are not associated with temporal lobe epilepsy preceded by febrile seizures patients and in patients with neurodevelopmental disorders with epilepsy (Ma et al. 2006a, Heyne et al. 2019). *GABRD* ins-del variant leading to 2 missense variants, p.Met166Ile, p.Asp167Asn, was identified in a Rett syndrome patient (Okamoto et al. 2015). *GABRD* via gene-gene interaction with *GABRB3* contributes to increased risk for autism spectrum disorders in Argentinian populations while a *GABRD* SNP was identified to increase susceptibility to a childhood-onset mood disorders in the Hungarian population (Feng et al. 2010, Sesarini et al. 2014).

Gamma-aminobutyric acid receptor, delta (*GABRD*), is located at extra-synaptic and peri-synaptic locations and mediate tonic inhibition (Dibbens et al. 2004). It is highly expressed in various brain regions, especially in the cerebellum, cerebral cortex, putamen, occipital lobe, temporal lobe, and frontal lobe. It is expressed in low amounts in the kidney (Windpassinger et al. 2002). Homozygous *Gabrd* null mice exhibited postpartum depression leading to maternal neglect and reduced pup survival, impaired memory, enhanced fear acquisition, and impaired adult neurogenesis due to migration, maturation, and dendritic development defects (Whissell et al. 2013, Maguire and Mody 2008).

1.4.2.2 CACNB4

CACNB4 gene was sequenced in a cohort of GGE and Ataxia patients that included 49 JME cases since mouse mutant *lethargic* that contains truncating mutation in *Cacnb4* exhibited ataxia, hypomotor behaviour, focal motor seizures, and absence seizures. Heterozygous rare nonsense mutation p.Arg482Ter was identified in one JME case that co-segregates with the family's disease. Functional analysis of p.Arg482Ter in *Xenopus laevis* oocytes identified the absence of slowly inactivating inward Ba^{+2} currents, increased current density and altered inactivation kinetics and affected nuclear localization in HEK293, CHO, and hippocampal neurons at seven days *in vitro* cells due to its inability to interact with Ppp2r5 affecting neuronal excitability to gene expression coupling, but did not affect its nuclear localization in tsA-201 cells, skeletal myotubes, and in hippocampal neurons (Burgess et al. 1997, Escayg et al. 2000b, Tadmouri et al. 2012, Etemad et al. 2014). *CACNB4* heterozygous mutation p.Arg468Gln in *SCN1A* mutation-positive background in severe myoclonic epilepsy in infancy case exhibited increased Ba^{+2} current density in mammalian cell electrophysiological experiments. A homozygous mutation p.Leu126Pro in a patient with a neurodevelopmental disorder that included intellectual disability, psychomotor retardation, blindness, epilepsy, movement disorder, and cerebellar atrophy did not associate with a calcium channel complex, lost nuclear localization property in cultured myotubes and hippocampal neurons and interaction with TNIK impairing its function (Ohmori et al. 2008, Coste de Bagneaux et al. 2020).

Calcium voltage-gated channel auxiliary subunit beta 4 (*CACNB4*) codes for the auxiliary β_4 subunit of the P/Q-type calcium channels regulating the amplitude, kinetics, and voltage-dependence of calcium currents of high-voltage-gated calcium channels that in turn regulates rapid neurotransmitter release in the brain. CACNB4 is expressed in the kidney, testis, retina,

lymphocytes, and lymphoblasts with the highest expression in the cerebellum and lower expression in the hippocampus among brain regions (Day et al. 1998, Escayg et al. 1998). Its localization progresses from cytoplasm/plasma membrane to nucleus as *in vitro* neurons differentiate. In the plasma membrane, it affects calcium channel properties, while in the nucleus, it regulates gene expression of tyrosine hydroxylase by repressing it (Tadmouri et al. 2012).

1.4.3 Non-mendelian JME and association studies

A comparison of SNP allele frequencies between patients and healthy individuals would identify those SNPs that pose a greater risk for the disease's predisposition. This can be conducted either by selecting a previously known set of polymorphisms and screen them in a cohort or genome-wide association analysis that would pick up multiple loci that could be related to the disease. Mendelian and non-mendelian method of inheritance of alleles would be detected that would include incompletely penetrant alleles and risk alleles associated with oligo or polygenic inheritance models. Population stratification can pose problems in such studies, which increases the number of false positives. GWAS, case-control, and family-based association studies have been conducted in several JME populations and have identified several loci and genes to contribute to JME risk. While *BRD2*, *GJD2*, and *ME2* associations were replicated in more than one study, have been described below, other genes associated with JME have been listed in Table 2.

1.4.3.1 BRD2

Case-control SNP analysis in samples JME samples that had exhibited significant linkage at previously identified JME locus EJM1 at 6p21 identified a strong association of SNPs rs3918149 (c.-1765G>A) and rs206787 (c.-1900A>T) located in the promoter region of the *BRD2* gene (Pal et al. 2003). Studies have also reported no association between the *BRD2* gene and JME in Europeans and GGE with photosensitivity in the Turkish population (Cavalleri et al. 2007, Yavuz et al. 2012, Schulz et al. 2019). Opposing studies report *BRD2* promoter hypermethylation in JME patients of Caucasian origin (Pathak et al. 2018). The 5'UTR c.-198A/T polymorphism in the *BRD2* gene was found to be associated with JME patients, while c.-198A/A was overrepresented in controls when compared to JME patients (Mehndiratta et al. 2007).

Bromodomain-containing protein 2 (*BRD2*) is a transcription factor that belongs to the BET family of proteins containing two tandem bromodomains and an extra-terminal domain that are involved in cellular proliferation and differentiation processes (Taniguchi 2016). The balance between neuronal proliferation and differentiation is maintained by growth factor Pleiotrophin – Brd2 interaction by preventing the association of Brd2 to chromatin in the embryonic nervous system (Garcia-Gutierrez et al. 2014). *Brd2,* also known as female sterile homeotic related gene - 1, *Fsrg1* is ubiquitously expressed, especially in hormonally modulated epithelia, including the mammary gland, ovary, kidney, and uterus (Rhee et al. 1998, Trousdale and Wolgemuth 2004). Brd2 mRNA is expressed in the brain, both in the hippocampus and cerebellum, while the protein can only be detected in the cerebellar Purkinje cells and not in hippocampal cells. Human brain regions cerebellum, cerebral cortex, medulla, spinal cord, occipital cortex, frontal cortex, temporal cortex, and putamen expresses *BRD2* transcript (Shang et al. 2011). Homozygous *Brd2* null mice exhibit embryonic lethality at E13.5 when Brd2 is highly expressed in the developing brain and show severe nervous system development defects (Gyuris et al. 2009, Shang et al. 2009). Heterozygous *Brd2* null mice are seemingly phenotypically normal but exhibit spontaneous generalized seizures on flurothyl vapor inhalation whose threshold varied between sexes during adolescence. Behavioural tests identified increased aggression in both sexes while decreased anxiety in females. The number of GABAergic neurons was decreased in the neocortex, striatum, substantia nigra reticulata, superior colliculus, and basolateral amygdala in *Brd2* Het mice that probably leads to seizure susceptibility (Chachua et al. 2014, Velíšek et al. 2011). *Brd2* Zebrafish morpholino knockdowns exhibited defective patterning of segmental tissue, especially the CNS with reduced hindbrain and an ill-defined midbrain-hindbrain boundary (MHB) region; irregular notochord, neural tube, and somites; and abnormalities in the ventral trunk and ventral nerve cord interneuron positioning were observed with extensive cellular apoptosis (Murphy et al. 2017).

1.4.3.2 GJD2

GJD2 is a neuron-specific gap junction protein-coding gene that lies at the EJM2 locus at chromosome 15q14, contributing to genetic susceptibility to JME. The locus also has genes *CHRNA4,* $GABA_A$ α_5, β_3, and γ_3 subunit gene cluster (Elmslie et al. 1997, Sander et al. 1997). *GJD2* gene sequencing in EJM2+JME patients and ethnically matched controls identified homozygous SNP rs3743123, p.Ser196Ser, c.588C>T associated with the disease in the European population (Mas et al. 2004, Hempelmann, Heils and Sander 2006).

GJD2, also called connexin-36 (CX36), belongs to the connexin gene family that are radially organized to form intercellular channels. GJD2 is expressed in the inferior olivary complex, cerebellar cortex, olfactory bulbs, brainstem nuclei, reticular thalamic nuclei, cerebral cortex, and (GABA)ergic interneurons of the adult mice's hippocampus and in the developing mouse's forebrain at E9.5 (Belluardo et al. 1999). *Gjd2* null mice were deficient in electrical synapses, exhibited impaired γ-frequency oscillation between interneurons of hippocampus and cortex, and kainite- induced spontaneous generation of sharp wave/ripple activity in the interneuronal network (Deans et al. 2001, Hormuzdi et al. 2001, Pais et al. 2003).

1.4.3.3 ME2

Genome scan analysis of small IGE families that include JME families supports an oligogenic model with the strongest evidence for locus in chromosome 18 (LOD score - 4.4/5.2 multipoint/two-point) in addition to other loci (Durner et al. 2001). Case-control and family-based association studies identified strong evidence of several SNPs located in the promoter of the *ME2* gene association to adolescent-onset IGE cases that included several JME index patients. Homozygous nine-SNP haplotype in a recessive inheritance model was found to increase the risk for IGE populations of European origin (Greenberg et al. 2005). Analysis of *ME2* SNPs in a freshly recruited GGE cohort and reanalysis in previously collected data supported *ME2* haplotype's association with the disease while *ME2* SNP genotyping studies in a German IGE cohort indicated otherwise (Lenzen et al. 2005b, Wang, Greenberg and Stewart 2019). SNPs within the *ME2* gene have also been associated with psychosis and mania, as well as 5.6-fold lower expression of *ME2* in anterior cingulate tissue post-mortem brains of bipolar patients (Lee et al. 2007).

Malic enzyme 2 (ME2) is a nuclear gene that codes for a homotetrameric mitochondrial NAD(+)-dependent enzyme that is involved in glucose metabolism by reversibly catalyzing the oxidative decarboxylation of malate to pyruvate, which is required for the synthesis of γ-aminobutyric acid (GABA) in neurons and provide NADPH to maintain intramitochondrial glutathione required for survival of neurons (Greenberg et al. 2005). Malic enzyme kinetic activity was enhanced in mitochondria of cortical synaptic terminals when compared to primary cultures of cortical neurons, cerebellar granule cells, or astrocytes (McKenna et al. 2000).

Table 2. Gene SNPs associated with JME

Gene	Locus	SNP	Method	Population	Reference
EFHC2	Xp11.3	rs2208592 (p.Ser430Tyr)	CC	German	(Gu et al. 2005)
GRM4	6p21.31	rs937039, rs745501, rs2451334, rs2499697 rs2029461	CC CC, FB	German Indian	(Muhle et al. 2010) (Parihar et al. 2014)
TAP1	6p21.3	p.Ile333Val –p.Asp637Gly	CC	Tunisian, French	(Layouni et al. 2010a) (Layouni et al. 2010b)
CHRNA4	20q13.33	1674+11C>T	CC	Polish	(Rozycka, Steinborn and Trzeciak 2009)
KCNJ10	1q23.2	rs1130183 (p.Arg271Cys)	CC	European, German	(Buono et al. 2004) (Lenzen et al. 2005a)
CPA6	8q13.2	p.Arg36His, p.Asn271Ser	CC	French	(Sapio et al. 2015)
MAST4	5q12.3	rs775035801 (p.Val1496Met) rs752330863 (p.Thr347Met) rs39861	FB FB GWAS	Tunisia European	(Landoulsi et al. 2018) (Steffens et al. 2012)
CACNA1H	16p13.3	p.Arg564His, p.Leu1581Val, p.His2280Glnfs*30	FB	Tunisian	(Landoulsi et al. 2018)
KCNQ3	8q24.22	D8S558 (CA)20	FB	India	(Vijai et al. 2003b)
CHRM3	1q43	rs12059546	GWAS, CC	European	(Steffens et al. 2012)
HLA	6p21.32	DRB1*07 DR13 * 0603 and 0604	CC CC	Portuguese European	(Chaves et al. 2020) (Greenberg et al. 1996)
STX1B	16p11.2	rs1046276	GWAS	European	(ILAE et al. 2018)

CC- Case-control association study, FB – Family-based association, GWAS – genome-wide association study

1.4.4 Copy number variations in JME

Structural deletions or duplications, ranging from tandem to complex multisite chromosomal rearrangements larger than 1kb, are defined as copy number variants (CNV) but these can be several mega-bases in length. Certain genomic regions that are more prone to microchromosomal variations are labelled 'hotspots' although CNVs can be distributed throughout the genome impartially, which usually encompass several genes. As ascertained from population studies, normal individuals' susceptibility for CNV varies from 6% to 19% per chromosome, encompassing more nucleotide bases than SNPs in a genome (Redon et al. 2006). CNVs can be inherited or be of *de novo* origin and are associated with both simple

and complex disorders. The earliest methods to detect CNVs included karyotyping, Fluorescent In Situ Hybridization (FISH), Multiplex Amplifiable Probe Hybridization MAPH, and multiplex Ligation-dependent Probe Amplification (MLPA). With the advent of genomic technology, high-density microarrays such as comparative genomic hybridization (aCGH) and single nucleotide polymorphism (SNP)-based microarrays (SNP-arrays) are utilized. They are the current mainstay in identifying CNVs, although the next-generation sequencing technique's utility is expanding (Shaikh 2017).

Each recurrent CNVs that occur at hotspots are present in 0.5-1% of GGE patients but are absent or rare in healthy individuals, and 3% of GGE cases are carriers for each CNV. They are also associated with other neurodevelopmental disorders such as autism spectrum disorders (ASDs), intellectual disability (ID), and schizophrenia. These CNVs are present in 3% of GGE cases but were 10% in GGE with intellectual disability patients (Mullen et al. 2013). Three microdeletions at 15q13.3, 15q11.2, and 16p13.11 were recurrently occurring CNVs in GGE patients. Since these are also present in unaffected family members, they are considered more as risk factors than as the sole cause of the disease (Mullen and Berkovic 2018).

15q13.3 microdeletion was first reported in 6 Western European JME patients that comprised 3.8 Mb between breakpoints BP3-BP5 that contains six genes, including *CHRNA7*, which was considered a prime candidate JME gene (Helbig et al. 2009). Several other studies have supported 15q13.3 microdeletion that lies between breakpoints at BP4-BP5 that is about 1.5Mb long, susceptibility to JME and exhibited complex inheritance (Dibbens et al. 2009, de Kovel et al. 2010, Mefford et al. 2010, Jähn et al. 2014, Lal et al. 2015). 15q13.3 microdeletion mouse model (Df[h15q13]/+) recapitulated human phenotypes associated with it, including increased risk of generalized epilepsy while these were more strongly pronounced in homozygous null mice (Fejgin et al. 2014, Forsingdal et al. 2016).

15q11.2 microdeletion was first identified in JME patients 0.5 – 1.5 kb from the north-western European population and has been identified in several other studies. The breakpoint markers BP1-BP2 covers four genes that include *NIPA1, NIPA2* that code for Mg+ transporting ion channel, and *CYFIP1* gene, which are considered the prime candidate genes. This region overlaps with an extended microdeletion frequently found in Prader-Willi Syndrome and Angelman Syndrome patients who also exhibit seizures as one of the clinical manifestations (de Kovel et al. 2010, Mefford et al. 2010, Mullen et al. 2013, Lal et al. 2015).

iPSC neurons derived from patients containing this microdeletion exhibited altered dendritic morphology compared to wild-type and reduced expression of the genes that lie in the region (Das et al. 2015).

16p13.11 microdeletions that range between 0.9 Mb – 3.1Mb have been identified in JME patients of mostly west European ancestry covering seven genes, including *NDE1* (de Kovel et al. 2010, Mefford et al. 2010, Mullen et al. 2013). *NDE1* is involved in cell division and cortical development. *NDE1* null mice exhibit embryonic lethality or microcephaly that is recapitulated more severely in humans in addition to lissencephaly by *NDE1* homozygous truncating mutation (Feng and Walsh 2004, Alkuraya et al. 2011, Doobin et al. 2016). While heterozygous SNVs in *NDE1* are associated with neuropsychiatric disorders, indicating a dramatically different impact of *NDE1* on phenotype (Kimura et al. 2015).

Rare CNVs have been identified in JME patients from both European and non-European populations, including deletions and duplications. Phenotypic heterogeneity is also observed among some of these CNVs. Candidate genes are proposed based on their brain expression, the effect on the nervous system's functioning, and previously known genes associated with a similar phenotype. These have been listed in Table 3.

Table 3: Rare copy number variations in JME patients not associated with comorbidities

Location	Size	Change	Gene	Population/ Country	Ref
6p12.1	51.5 Kb	Duplication	*BMP*	Saudi arabia	(Naseer et al. 2015)
7q32.3	59.5 Kb	Deletion	*PODXL*	Saudi arabia	(Naseer et al. 2015)
22q11.22	254.6 Kb	Deletion	*TOP3B NA*	Tunisia European	(Daghsni et al. 2018) (Strehlow et al. 2016)
Xp22.31	1.7 Mb	Duplication	-	Italy	(Brinciotti et al. 2018)
2q21.1	0.49 Mb	Deletions	*RAB6C*	NW European	(Lal et al. 2015)
4p15.1	3.45 Mb	Deletions	*PCDH7*	NW European	(Lal et al. 2015)
11p15.4	1.09 Mb	Deletions	*C11orf40, TRIM68*	NW European	(Lal et al. 2015)
21q22.3	0.49 Mb	Deletions	*ADARB1, S100B*	NW European	(Lal et al. 2015)
16p11.2	700 Kb	Duplication	*SEZ6L2*	European	(Mefford et al. 2010)
6q12	1.06 Mb	Duplication	*EYS*	European	(Mefford et al. 2010)
7q11.22	78.7 Kb	Deletions	*AUTS2*	European	(Mefford et al. 2010)
9p21.3	427.5 Kb	Deletions	*KLHL9*	European	(Mefford et al. 2010)
13q31.1	671.8 Kb	Deletions	*SLITRK6*	European	(Mefford et al. 2010)
14q24.2	268.6 Kb	Deletions	*SIPA1L1*	European	(Mefford et al. 2010)

17p11.2	17.5 Kb	Deletions	*CYTSB*	European	(Mefford et al. 2010)
18q11.2	840.4 Kb	Duplication	-	European * 2ptn	(Mefford et al. 2010)
2q16.3	60 Kb	Deletions	*NRXN1*	European * 2ptn	(Møller et al. 2013)
16p13.3	100 Kb	Deletion	*RBFOX1*	NW European	(Lal et al. 2013)
14q23.3	129 Kb	Deletion	*GPHN*	Germany	(Dejanovic et al. 2014)

1.5 The relevance of JME genetics studies

As described in the above section, several independent studies have demonstrated that a genetic component underlies the aetiology of juvenile myoclonic epilepsy. These findings were initially made via family-based studies. Later, advances in genomic sequencing and large cohorts' study via consortia further supported the genetic basis for JME. A variety of genetic variants ranging from single nucleotide variants, including synonymous, missense and nonsense variants to copy number variants that are *de novo,* have been identified. Non-coding variants located in promoter or intronic regions have also been reported to be associated with JME. Both monogenic and oligo/polygenic inheritance models have been observed where the former mostly involved gene coding for ion channels. In contrast, the latter model has mapped non-ion channel genes that include transcription factors, enzymes, microtubule-associated proteins, etc. Genetic variant- risk to JME varies across populations. Despite best efforts, in the vast majority of JME cases, underlying genetic factors are yet to be elucidated, indicating genetic heterogeneity and complex genetic architecture.

The identification of genes associated with JME provides key insights into the mechanism underlying JME pathogenesis. Cellular models, including immortal cell lines, primary cell cultures, iPSC, and few animal models such as mice, zebrafish, Xenopus, and flies have been used to understand the effect of variants on gene's function leading to JME phenotype. The common feature among the JME genes identified is their contribution to neurogenesis and brain development, including cell proliferation, cell migration, neurite extension, neuronal connectivity, excitability, and synaptic transmission. Structural brain changes observed in JME patients reflect these subtle neurodevelopmental defects. The distinctive association of sleep with seizure triggers in JME cases also supports this rationale since its primary activity is neurobiological development. This also indicates that multiple pathways lead to the development of phenotype. The adolescent onset of the clinical presentation is yet to be elucidated, although sex differences in JME manifestation that has been observed in certain populations likely reflecting the neurobiology of seizure semiology. Certain JME associated

genes are also known to cause other disorders, and the presentation of various degrees of severity of the phenotype by the same gene complicates the genotype-phenotype correlation. Mechanistic understanding of JME shall help identify therapeutic targets or suggest appropriate treatment options.

1.6 Objectives of the current study

The aim of my research work was to explore genetic factors and their contribution to etiology of juvenile myoclonic epilepsy. I investigated a multigenerational JME family and a set of JME patients from the southern parts of India using next generation sequencing technology. This work involved examination of a family exhibiting Mendelian inheritance of JME, which has been previously subjected to a genome-wide linkage analysis identifying locus 5q34 for the disorder in the family. I also conducted a candidate gene sequence analysis of a neurodevelopmental disorder associated gene's possible role in JME. Main objectives of my work, discussed in this thesis are as follows:

i) Whole genome sequencing and focused 5q34 GABA gene cluster analysis in NIH34 family - identification and in vitro functional validation of rare segregating alleles.

ii) Functional evaluation of rare variants identified among a set of JME patients in the prospective gene, *SOX30* – a transcription factor, linked to JME using cell-based assays and identification of its potential genomic targets using chromatin immunoprecipitation followed by massive parallel sequencing.

iii) Identification of rare *CHD2* variants in a JME cohort and cell-based analysis of their potential effect on the gene.

Chapter 2

Whole-genome sequencing analysis of the 5q33-q35 region suggests a role for the SOX30 gene for juvenile myoclonic epilepsy.

2.1 Summary

In this chapter, I discuss analysis of the 5q33-35 locus linked to juvenile myoclonic epilepsy in an affected family and identification of a potentially causative 5'UTR variant (c.-57G>A) in *SOX30* in the family. The 5q33-35 linked region comprises 15Mb of the genome and contains 54 protein-coding genes. This region also includes the gamma aminobutyric acid receptor gene cluster containing four known epilepsy associated genes. Complete gene sequencing of GABA receptor genes in this locus including the 2kb promoter region identified no critical variants. Whole genome sequencing was conducted for the proband and her father, and this identified 985 rare, heterozygous variants at 5q33-35. No amino acid altering rare variants were identified. Bioinformatic analysis, family-segregation and in-house healthy individual sequencing analysis identified four UTR variants (*SOX30* c.-57G>A; *EBF1* c.*1437T>c; *CNOT8* c.*1058G>A; and *ZBED8* c.*322G>A) which satisfied the criteria for being potentially pathogenic variants. Reporter luciferase assays indicated that only the *SOX30* c.-57G>A variant altered activity when compared to the reference allele. From these findings, it was inferred that *SOX30*, a lineage-associated transcription factor, is the most likely pathogenic gene contributing to juvenile myoclonic epilepsy in the family.

2.2 Introduction

A previous genetic study of a four-generation south Indian family with several of its members affected with juvenile myoclonic epilepsy (JME) identified a genetic locus for the disorder. A whole genome-based linkage analysis found a region at chromosome 5q33-35, wherein the highest two-point LOD score of 3.66 was obtained for the microsatellite marker D5S415 at 90% penetrance value. The centromere-proximal and -distal limits of the region were defined by D5S2012 and D5S2075. This critical region corresponds to genetic interval of 19cM and encompasses 15Mb of the human genome sequence. While candidate gene analysis was conducted earlier for the critical region, a detailed whole genome-based analysis was not, and needed to be carried out. In the present study, the family and locus were analyzed employing next-generation sequencing technology wherein one can detect almost all the sequence variants present including in the non-coding regions and newly annotated genes, to draw a

more informed conclusion about a potentially causative gene as compared to a candidate gene approach.

2.3 Materials and methods

2.3.1 GABA gene cluster sequencing

Four genes, *GABRB2, GABRA6, GABRA1 and GABRG2,* situated in the 5q33-q35 locus encompass about 420kb of the genome. Overlapping primer sets covering this gene cluster were designed using Primer3 software and their propensity to form hairpin formation and self-complementarity were predicted by Oligocalc. Each amplicon overlapped with the adjacent amplicon by about 200bp. Thirty-two primer pairs were synthesized (Sigma Aldrich, St. Louis, USA) to examine *GABRA6*; 95 to examine *GABRA1*; 159 to examine *GABRG2*; and 362 to examine *GABRB2* (Table A2.2). Genomic DNA (50ng) PCR amplified using 1X standard reaction buffer (New England Biolabs, Ipswich, USA) containing 10mM Tris-HCl and 50mM KCl (pH 8.3), 0.25µM of forward and reverse primers, 800uM dNTPs, 1.5mM $MgCl_2$, and 1U of *Taq* Polymerase in a 20µl reaction volume. Amplification was carried out using a GeneAmp9700 thermal cycler (Applied Biosystems, Foster City, USA). The conditions for the PCR amplification were: initial denaturation at 94°C for 5minutes, followed by 40 cycles of denaturation at 94°C for 30 seconds, annealing at 55-65°C for 30 seconds, extension at 72°C for 30 seconds and a final extension at 72°C for 15 minutes. The amplified products were purified using MultiScreen® PCRµ96 Filter Plate (Millipore, Burlington, USA) using a vacuum manifold system. Cycle sequencing was performed on a GeneAmp 9700 thermal cycler using 100-200ng of purified PCR product, 4µL of 5X sequencing buffer 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 dideoxyterminators, 0.04µM of forward or reverse primers in a 20µl reaction volume, made up with deionized water on the following conditions: initial denaturation 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 in 96 well sequencing plate. Cycle sequencing products were precipitated in 95% ethanol, the DNA pellet washed with 70% ethanol followed by dissolution in 10µl HiDi Formamide (Applied biosystems), denaturing at 96°C for 5 minutes and snap chilling in ice. The samples were then electrophorized on a 3730 Genetic Analyzer (Applied Biosystems). Each amplicon was sequenced bi-directionally and

analyzed using Sequencing Analysis Software (Applied Biosystems) and SeqMan II 5.01 software (DNASTAR Inc., Madison, USA). Except for *GABRB2,* which was sequenced in the individual IV:1 due to sample limitation, other GABA genes were sequenced in the individual III:2.

2.3.2 Library preparation and sequencing

Whole-genome sequencing was carried out for an unaffected father (II:1) and his affected offspring (III:2). Genomic DNA concentration was measured using Nanodrop 1000 spectrophotometer (ThermoFisher scientific, Waltham, USA) and Qubit 2.0 fluorometer (ThermoFisher Scientific, USA). Quality of DNA was analyzed by 1.5% agarose gel electrophoresis to check for RNA contamination, if any, and signs of degraded genomic DNA. 50ng of genomic DNA was tagmented (transposon-mediated fragmentation followed by adapter incorporation at fragment ends) using the Nextera DNA library prep kit (Illumina, San Diego, USA). Limited-cycle PCR was used to amplify the DNA fragments, and these were purified using ZR-96 DNA clean and concentrator-deep well (Zymoresearch, Irvine, USA). Dual index approach was used, wherein two 8-base indexes are added to each sample using five-cycle PCR and purification done using size-selective AMPure XP beads (Beckman coulter, Brea, USA) (Table 1). Quality control analysis of the library was done using Agilent 2100 Bioanalyzer (Agilent, Santa Clara, USA) to confirm the expected fragment size and yield. The library generated was subjected to onboard cluster generation using TruSeq rapid paired-end cluster kit (Illumina) followed by sequencing by synthesis using Truseq rapid SBS kit (Illumina) in two batches for each sample, using two flow cells in the first batch and one in the second batch on Hi-Seq 2500 (Illumina) generating 2 X 100 paired end reads in rapid run mode according to manufacturer's instruction.

Table1:Index primers used to label samples for whole-genome sequencing

Sample	Index 1 (i7)	Index 2 (i5)
III:2	N701 - TAAGGCGA	N501 - TAGATCGC
II:1	N702 - CGTACTAG	N502 - CTCTCTAT

2.3.3 Whole-genome sequencing data analysis

Real-time analysis software (RTA)(Illumina) stores the base call data in BCL (Binary Base Call) format. For compatibility with the downstream analysis pipeline, BCL files are

converted to FASTQ files using BCL2FASTQ software (Illumina), which also de-multiplexes sample sequences based on cluster's index sequences. Raw FASTQ files were subjected to adapter trimming using Cutadapt. *Inhouse* scripts were used for primer and low quality base trimming. Reads with at least 70% of the bases having Phred score ≥ 20 were considered for alignment to the human reference genome (GRCh37/hg19 assembly) using Burrow wheelers aligner 0.7.5a, which helped generate SAM (Sequence Alignment Map) output files. Duplicate reads resulting from PCR artifacts were removed using SAMtools version 0.1.19-44428cd. SAM files were converted to BAM (Binary version of SAM) files that were *de facto* standard for storing large nucleotide sequence alignments using SAMTools for further processing. Local realignment around indels and base quality recalibration were performed using GATK v3.1-1-g07a4bf8 (Genome analysis toolkit). GATKHaplotype caller was used for calling single nucleotide variants (SNVs) and short insertions/deletions (InDels) generating VCF (variant call format) files. Minor allele frequencies of these variants were obtained from 1000 Genomes and dbSNP144 databases using *in house* scripts. SnpEFF was used to annotate and predict the effects of the variants on genes, categorizing them into nonsynonymous, synonymous, UTR, frameshift, splice site donor/acceptor, start/stop – loss/gain, intergenic and intronic subtypes. Read and alignment statistics were obtained using *in-house* scripts.

2.3.4 Sanger sequence validation of novel/rare variants

Genes harbouring variants of probable significance had their nucleotide sequence obtained from the GenBank sequence database. The variants in 5q33-35, which were present exclusively in the proband, and were novel or had a minor allele frequency less than 0.005, were taken up for validation by Sanger sequencing. Primers were designed across the exon/region carrying the variant (Table A2.7). Sanger sequencing was carried out as per the protocol mentioned in the previous section. Sanger sequencing confirmed variants were then checked for their segregation with epilepsy in the family. On confirmation, the variants were screened in 192 ethnically matched healthy controls. Genes which had remained uncovered in the previous candidate gene analysis were also examined by Sanger sequencing (Table A2.1)

2.3.5 Bioinformatic analysis of significant variants

In silico analysis was performed on rare and novel variants. Sequence conservation was obtained from Multiz alignment (UCSC Genome browser) for eight vertebrate species. The

deleteriousness of variants was scored using Combined Annotation Dependent Depletion (CADD) annotation, which is calculated using multiple genomic features such as conservation, epigenetic and regulatory features, allele frequency etc, wherein variants with CADD Phred-like scaled score greater than 10 were considered. Mutationtaster was used to predict the deleteriousness of the sequence variants. UCSC Genome browser track sets were used to analyze the presence of transcription factor binding sites (ChIP-Seq), histone modification (ChIP-Seq), Open chromatin regions (DNase-seq, FAIRE-seq), CpG island from ENCODE database. Repeatmaster was used to identify the presence of repetitive regions that includes transposon elements such as SINEs, LINEs, satellite DNA, simple and long terminal repeats. miRNA, ncRNA, and snoRNA binding sites in 3'UTR were analyzed using TargetScan, PolymiRTS, miRbase, miRDB, and RegRNA2.0. The splicing effect was determined by the Human splicing factor and Mutationtaster. Minor allele frequencies were obtained from Genome Aggregation Database (gnomAD), Trans-Omics for Precision Medicine program (TOPMED), 1000 Genomes, Exome variant server (EVS) databases, GenomeAsia100K, Indigene and INDEX-db databases.

2.3.6 Reporter constructs and site-directed mutagenesis

The 3'UTR of genes *ZBED8*, *CNOT8,* and *EBF1, which* were 558bp, 1353bp, and 3196bp, respectively, were amplified from human genomic DNA using primers sets containing sequences for restriction sites *SacI* and *HindIII* for *ZBED8;* and *SacI* and *MluI* for *CNOT8* and *EBF1*(Table A2.8). pMIR-REPORT Luciferase miRNA Expression Reporter vector was used to insert the 3'UTR PCR products in the multiple cloning sites that lies downstream of the luciferase gene. Restriction digestion was done for both vector and insert DNA, generating compatible restriction ends, followed by purification using the QIAquick Gel Extraction Kit (QIAGEN, Hilden, Germany). The vector to insert ratio of 1:3 was used for ligation at 16°C for 72hrs using T4 DNA ligase enzyme (NEB). XL10-gold $CaCl_2$ treated competent cells were used for the transformation of the ligation mix, and the bacterial colonies were screened by colony-PCR. Plasmids from the colonies positive for the insert were extracted using a QIAGEN Plasmid Mini Kit (QIAGEN). The insert sequences were validated by Sanger sequencing. Site-directed mutagenesis with specific primers was performed to introduce variants identified in the NGS study in their respective genes using QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies), which was confirmed by Sanger sequencing (Table A2.9). The 940bp fragment of *SOX30* 5'UTR was amplified from human wildtype and variant containing genomic DNA and inserted in between *NheI* and

HindIII restriction sites in pGL3 basic vector (Promega, Madison, USA) (Table A2.8). pCMVBeta-galactosidase vector was used for 5'UTR reporter assays, while pMIR-REPORT™ Beta-galactosidase Reporter Control vector for 3'UTR reporter assays for normalizing transfection efficiency.

2.3.7 Cell culture and transient transfection

HEK293 and SH-SY5Y cell lines were cultured in Dulbecco's modified eagle medium (DMEM) (Sigma Aldrich) containing 10% heat-inactivated fetal bovine serum (Sigma Aldrich), 2mM Glutamine and antibiotics (100U/ml penicillin and 0.1mg/ml streptomycin (Sigma Aldrich) in a humidified atmosphere of 5% CO_2 at 37°C. The cultured cells grown in 12- well dishes, at about 40% confluency were transfected with the 1µg of luciferase vector (empty, wildtype or mutant) along with a 100ng of control β-galactosidase vector using lipofectamine 2000 (Thermo Fisher) in media without serum. Six hours post-transfection, the media was replaced with the fresh media containing serum. Twenty-four hours later, cells were harvested and lysed in reporter lysis buffer (Promega) which were used for luciferase reporter assays.

2.3.8 Reporter luciferase assays for the variants

Cells collected post-transfection were washed once with 1X PBS and resuspended in 100µl of 1X reporter lysis buffer (Promega) and incubated on ice for 2 hours for cell lysis to occur. The sample was centrifuged for 15min at 4°C at 13000rpm. The supernatant lysate was collected and stored at -20°C. 2µl of the lysate was mixed with 10µl of luciferase assay reagent (Promega) and the firefly luciferase activity was measured by a luminometer (Berthold Detection Systems, Pforzheim, Germany). Differences in transfection efficiencies were corrected for by assaying for β-galactosidase enzyme activity. It was measured by mixing 20µl of lysate to 50µl of 2X assay buffer (Promega) which contains substrate ortho-Nitrophenyl-β-galactoside (ONPG) and incubation at 37°C for 45 minutes. 100µl of the stop solution containing 1M sodium bicarbonate was added to the reaction mixture to stop the reaction. The β galactosidase activity was measured using an ELISA plate reader (Molecular Devices, San Jose, USA) at 420nm. Relative luciferase values after normalizing the data were obtained from three biological replicates. Two-tailed Student's t-test was conducted using GraphPad Prism for statistical analysis.

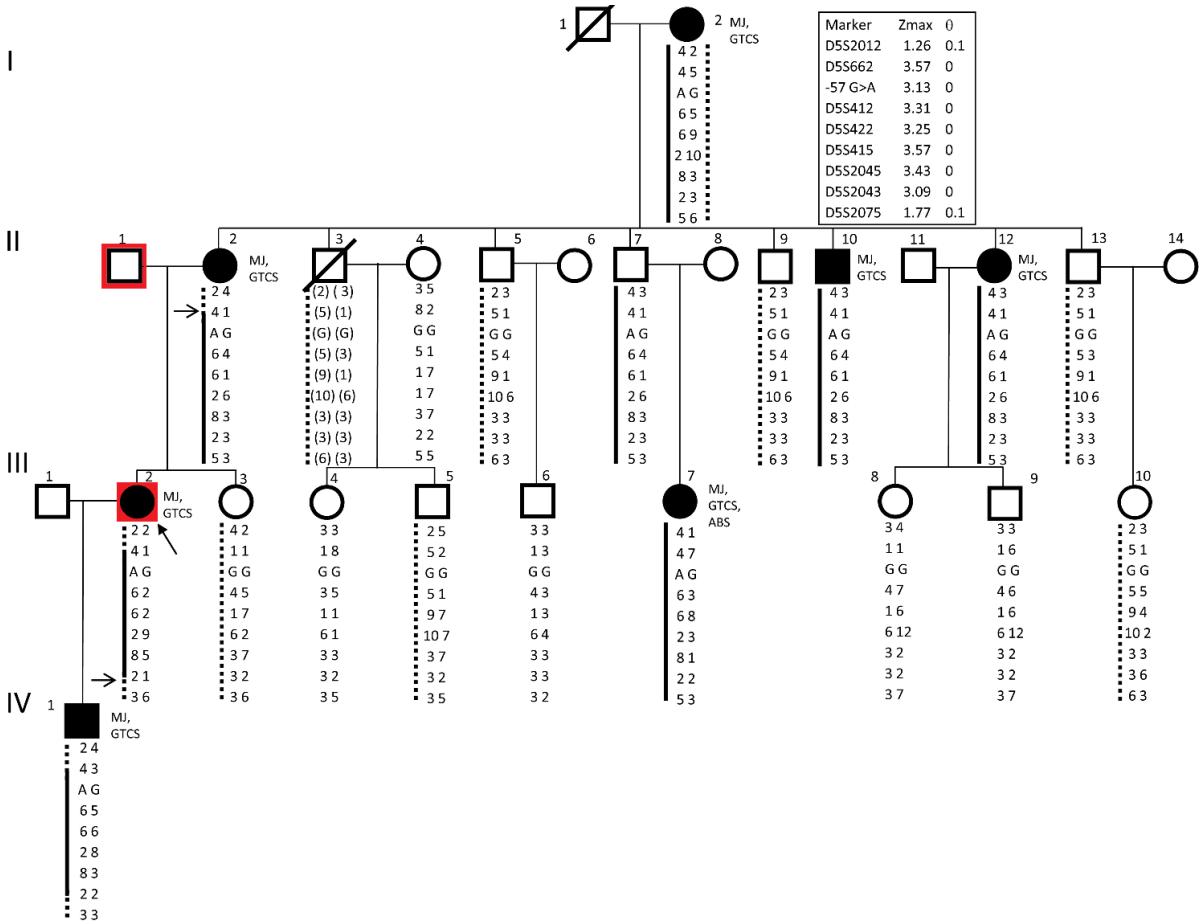


Figure 1: **Pedigree chart of the family, NIH34**: The squares represent males and circles, females. The filled symbols denote affected members, and the unaffected members are denoted by empty symbols. Clinical features of affected members are indicated beside their symbols (MJ: myoclonic jerks, GTCS: generalised tonic-clonic seizures, ABS: absence seizures. The nine-marker critical haplotype between D5S2012 and D5S2075 is denoted by a solid line below the individuals. Recombination sites are marked by arrows in individuals II:2 and III:2. Red-squared individuals were taken up for whole-genome sequencing (modified from Ratnapriya, R. PhD Thesis, 2009).

2.4 Results

2.4.1 JME 5q33-35 locus analysis

2.4.1.1 GABA gene cluster

Sequencing of the four 5q33-35, GABA receptor genes: *GABRA1*, *GABRA6, GABRG2* and *GABRB2* covered the exonic, intronic, and 2kb promoter regions upstream of the transcription start sites (TSS). We identified 194 variants, of which 47 were in *GABRA1*, 23 in GABRA6, 45 in *GABRG2* and 79 in *GABRB2* including both SNV and InDels (Table A2.3). No frameshift, nonsense or missense variants were present. Three synonymous variants, three 5'UTR and five 3'UTR variants were found which were commonly reported in the population with global minor allele frequency (MAF) greater than 0.01. Of these 194 variants, 94 were heterozygous of which two were novel variants in *GABRA6* and *GABRB2* and five were reported variants in *GABRB2* and *GABRG2* with MAF less than 0.005 (Table 2). No rare/novel heterozygous variants were identified in *GABRA1*, a well-established JME gene. *GABRG2* rare variants had MAF >0.005 in the Asian/ Indian population obtained from GenomeAsia100k and IndiGen databases. These seven rare/novel variants were in the deep intronic regions of the respective genes ranging from 983 to 22312 bases away from the closest exon, in poorly conserved regions, in no regulatory feature containing sites and were predicted to be polymorphisms by MutationTaster. They did not co-segregate with the epilepsy phenotype in the family indicating that these variants were unlikely to contribute to causation of JME.

Table 2: Novel/rare heterozygous GABA$_A$receptor subunit gene variants identified by Sanger-based sequencing.

Gene	Genomic variant GRCh37/hg19	cDNA change	rsID	Global MAF	GA100K/ GA100K India MAF	Indigen MAF	CADD score
GABRB2	g.161577753A>G	c.459-22312C>T			-	-	1.687
GABRB2	g.160881766G>A	c.458+4864C>T	rs146198359	0.002796	0.002012 / -	0.0044	1.415
GABRB2	g.160888862T>A	c.238-2012A>T	rs149477433	0.002796	0.002017 / -	0.0044	4.223
GABRB2	g.161513901G>A	c.238-10632A>G	rs532076409	0.0002	0.000575 / 0.0017	0.0015	2.544
GABRA6	g.160897482T>C	c.1086+4308C>T			-	-	6.12
GABRG2	g.161123514C>T	c.108-6933G>A	rs539352915	0.002196	0.003739 / 0.00829	0.0064	0.006
GABRG2	g.160860375G>A	c.1249-983A>G	rs145030721	0.004792	0.003223 / 0.005	0.0059	0.22

GA100K - GenomeAsia100K, IndiGen – IndiGenomes

2.4.1.2 Whole-genome sequencing and the 5q33-35 region

DNA samples of the proband (III:2) and her unaffected father (II:1) were subjected to rapid run whole-genome sequencing on an Illumina HiSeq2500 platform (Figure 1). About 181.5 Gb of data was generated for each sample, with an average of 654.96 million raw reads containing >90% high-quality reads (Phred score > 20) (Table 3). The processed reads were mapped to the GRCh37/hg19 human reference genome. An average genome coverage of 91.82% with a mean depth of 38.92 was obtained. The critical haplotype that lies between the markers D5S2012 and D5S2075 is a 15Mb sub-genomic region harbouring 54 protein-coding genes, 24 pseudogenes, 10 long noncoding RNA genes and 9 miRNA genes. The average coverage of the region of interest is 99.61%, with a mean read depth of 42.14. At least 96.12% of target regions with more than 20-fold read depth were successfully captured, with 97.48% of the exonic regions covered (Table 4). The uncovered exons in this region were exons 33, 34, and 35 of *SLIT3*. Exons with less than 5X read depth included first exons of genes *EBF1, HMMR*, *GABRB2* and *GABRA1*. Exons 8 and 9 of *SGCD*, exon 2 of *ADRA1B,* exon 16 of *EBF1* also had read depth less than 5. These 'missing exons' were examined by Sanger sequencing.

2.4.1.3 Variant filtration, prioritization, and bioinformatics

Variants common between the two individuals (II:1, III:2; Figure 1) were omitted from further evaluation because the unaffected individual (II:1) does not share the haplotype. Variants exclusively present in the proband were 10,48,722, of which 282 variants were heterozygous protein sequence-altering variants such as frameshift, nonsense, missense, codon deletion or insertion and splicing variants (Table A2.4). Among the proband-exclusive genomic variants, 5545 were present in the critical region. Considering an autosomal dominant mode of inheritance of JME in family NIH34, 5338 heterozygous variations were taken forward for further analysis. The variants were verified for their occurrence in dbSNP144 and 1000Genome databases. dbSNP144 was found to have annotated a higher number of variants in comparison to 1000 Genome project datasets. Of 5338 variants, 985 had minor allele frequencies less than 0.005 (Table A2.5, A2.6). Among these, seven variants were present in the mRNA coding region, while the rest (978) were present either in the intergenic or intronic regions or other RNA genes. All the variants were manually analyzed by viewing their read alignment in SAMTools to further filter out variants with possible strand-bias and variants that fall into homo- or hetero-polymeric regions (Figure 2).

Table 3: Summary of reads statistics of whole-genome sequencing for the individuals, III:2 and II:1

Read direction	**Total number of reads**	**% of HQ Reads**	**Read length**			**Total number of bases**	**Total number of HQ bases**	**% of HQ bases**	**Non-ATGC character in reads**		**A/T/G/C Content (in %)**			
			Min.	Max	Avg.				Count	%	Adenine	Thymine	Guanine	Cytosine
Raw reads sample III:2														
Forward	68,49,48,428	100.00	110	110	110	75,34,43,27,080	73,90,44,13,165	98.09	18,24,977	0.27	28.96	29.12	20.68	21.22
Reverse	68,49,48,428	99.97%	110	110	110	75,34,43,27,080	68,51,93,03,923	90.94	23,34,853	0.34	29.01	28.86	20.86	21.07
Processed reads sample III:2														
Forward	60,53,56,679	100.00	51	110	108.79	65,85,54,46,176	65,31,48,96,969	99.18	01,54,899	0.03	29.18	29.34	20.54	20.93
Reverse	60,53,56,679	100.00	51	110	108.56	65,71,59,22,651	64,00,55,41,863	97.40	04,22,861	0.07	29.09	29.43	20.52	20.96
Unpaired	07,64,17,197	100.00	51	110	108.95	08,32,53,54,390	07,91,43,00,176	95.06	06,85,790	0.9	27.66	29.06	21.01	22.26
Raw reads sample II:1														
Forward	70,45,66,553	100.00	110	110	110	77,50,23,20,830	75,72,96,64,049	97.71	10,13,873	0.14	28.85	29.01	20.7	21.43
Reverse	70,45,66,553	99.97%	110	110	110	77,50,23,20,830	71,80,58,48,133	92.65	15,86,277	0.23	28.9	28.85	20.97	21.16
Processed reads sample II:1														
Forward	63,82,28,325	100.00	51	110	108.46	69,22,07,72,245	68,45,78,42,676	98.90	02,05,825	0.03	29.04	29.25	20.59	21.12
Reverse	63,82,28,325	100.00	51	110	108.07	68,97,47,38,632	67,33,50,95,290	97.62	04,69,178	0.07	28.98	29.37	20.57	21.08
Unpaired	06,28,39,665	100.00	51	110	108.79	06,83,63,18,953	06,42,43,02,646	93.97	02,29,408	0.37	27.85	28.86	21.01	22.27

Table 4: Sequence alignment and coverage summary for whole genome sequencing of III:2 and II:1

Fields	III:2 (affected)				II:1 (unaffected)			
	Genome	Region	Region (all exons)	Region (coding exons)	Genome	Region	Region (all exons)	Region (coding exons)
Total number of reads	1,23,72,44,280				1,27,91,67,362			
Total number of reads aligned	1,12,26,03,906				1,12,10,52,885			
% of reads aligned	90.73				87.64			
Target length	3,13,71,61,264	1,50,32,743	2,33,154	92,423	3,13,71,61,264	1,50,32,743	2,33,154	92,423
Target covered	2,87,01,35,847	1,49,73,965	2,32,258	92010	2,89,06,78,875	1,49,73,836	2,32,272	92,010
% of target covered	91.49	99.61	99.61	99.55	92.14	99.61	99.62	99.55
% of target covered with at least 5X read depth	90.93	99.44	99.52	99.55	91.64	99.48	99.53	99.55
% of target covered with at least 10X read depth	90.03	99.06	99.35	99.41	90.58	99.12	99.28	99.47
% of target covered with at least 15X read depth	88.85	98.21	98.73	99.19	88.74	98.15	98.73	99.16
% of target covered with at least 20X read depth	86.72	96.36	97.51	98.49	85.37	96.01	97.44	98.27
Average read depth	38.99	42.41	43.24	43.97	38.85	41.88	42.55	43.47

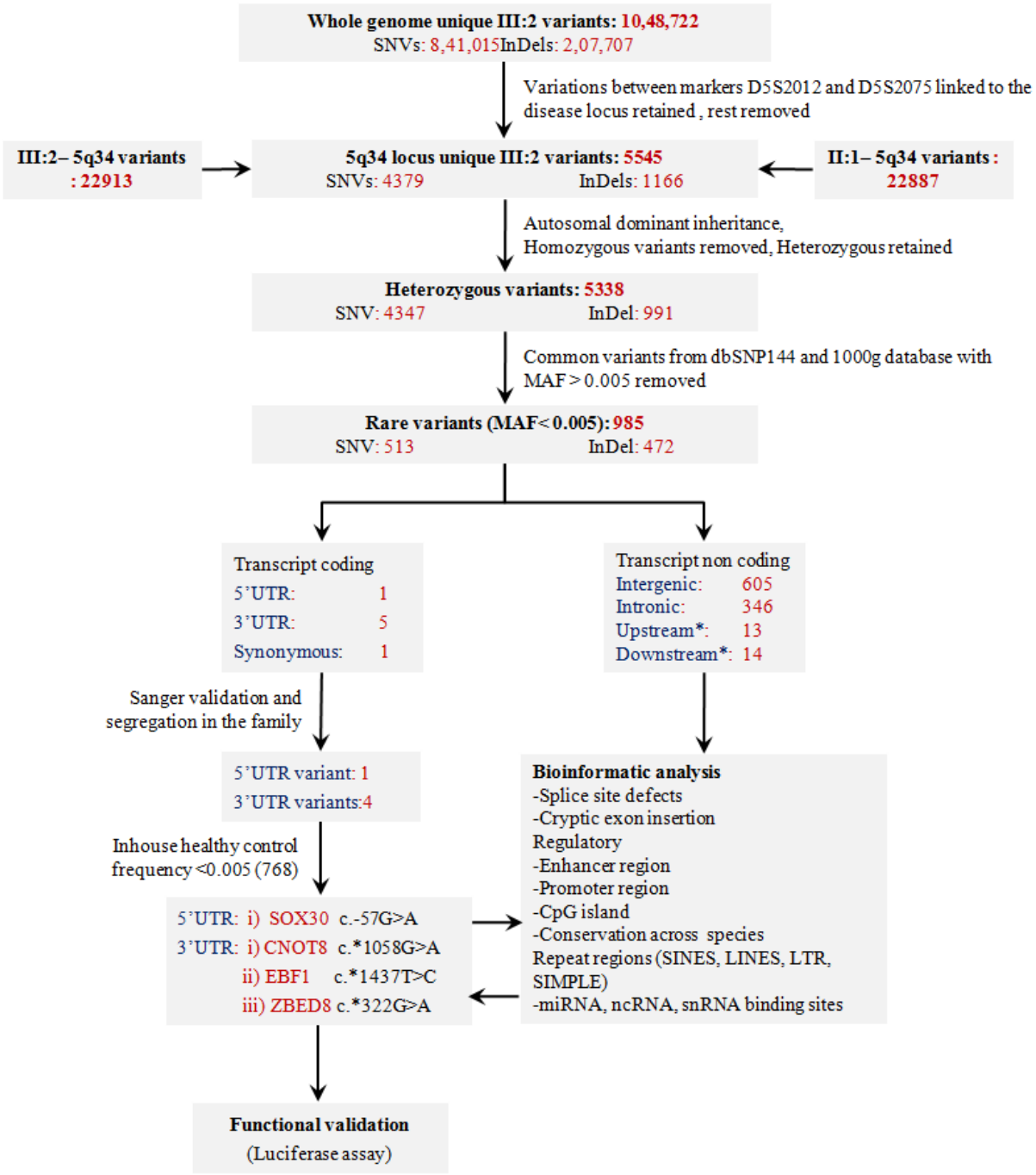


Figure 2: Flow chart showing filtering methods to prioritize variants obtained from whole-genome sequencing study. The variants were filtered based on their zygosity (heterozygous), minor allele frequency threshold of 0.005 obtained from dbSNP144 databases. *denotes bases up to 5kb.

GABA receptor genes were also separately analyzed and compared to the previously performed Sanger sequencing data in the proband sample. In addition to the variants identified by Sanger sequencing, several other variants were also called. Most of these lay in homo- and hetero-polymer regions or mostly in GC-rich polymer regions which are prone to errors in sequencing and were found to be false in Sanger sequencing datasets. None of the true variants were exclusively present in the affected members of the epilepsy family and were hence, discarded from further analysis.

No high or medium impact variants such as frameshift, nonsynonymous, or splice site variants were identified. Three missense variants present exclusively in the proband, were found in *RARS, FABP6,* and *C5orf52*. However, their allele frequencies were 0.01, 0.46, and 0.18, respectively - these were not considered further. The 978 non-coding variants were analyzed for their presence in regulatory regions: enhancers, promoters, silencers and repressors, DNase hypersensitivity regions, H3K27Ac epigenetic marks, transcription factor binding sites, miRNA and snoRNA binding regions, repeat regions such as LINES, SINES, LTR, etc., splice regulatory elements, cryptic or pseudo-exon inclusion, and their conservation, using bioinformatic tools and online datasets. CADD uses conservation, functional genomic, and expression data for calculation of scores. CADD scores for these non-coding variants were obtained, and variants with a value above 10, predicted to be 10% of most deleterious variants, were considered.

Updated minor allele frequency values were obtained from databases. Variants with MAF> 0.005 were discarded. In total, 30 rare non-coding variants with CADD phred score greater than 10 were present (Table 6). These were 29 SNVs and 1 Indel, and of which 5 were unreported variants. Of these, 18 variants were rare in the Indian population databases Indigene and GenomeAsia100k. Among these population specific rare variants, 10 were in repeat regions and hence were presumed to be inconsequential. Of the remaining 8 variants, 2 were in introns that were predicted to be polymorphisms by MutationTaster, and 6 were intergenic variants. No regulatory features were present at the location of these 8 variants. With this, we inferred that no critical variants were located in the non-exonic regions that could contribute to JME in this family.

Amongst the mRNA coding variants, five were present in 3'UTR, 1 in 5'UTR, and 1 was a synonymous variant. Sanger sequencing was done to confirm the variants. Two false positives - a synonymous change in *HAVCR1* and a 3'UTR variant in *CYFIP2* which were

in the polyA region were excluded. All the other Sanger confirmed variants segregated with the clinical phenotype in the family. A 3'UTR variant in *PWWP2A* was eliminated due to its presence in normal individuals at a minor allele frequency of 0.0075. Three of the remaining variants were present in the 3'UTR regions and were not reported in the databases, and one, in 5'UTR was rare with a MAF of 0.003 in our healthy control individuals and a MAF of 0.002 in global population databases (Table 5).

These variants were c.*322G>A in *ZBED8*, c.*1058G>A in *CNOT8*, c.*1437T>C in *EBF1*, and c.-57G>A in *SOX30* (Figure 3). They co-segregated with JME in the family; and as mentioned earlier, were absent or rare in both databases, and in- house controls of 384 individuals (Table 5).

Table 5: Novel or rare, mRNA variants identified at the 5q33-35 locus in III:2

Gene	Genomic variant GRCh37/hg19	mRNA change	Location/ Type	Protein effect	Variant ID	MAF Databases	Sanger validation	Segregation	MAF In-house normal individuals
CNOT8	g.154256057G>A	NM_00477: c.*1058G>A	3'UTR	-		-	TRUE	Yes	0.0013 (2/768)
HAVCR1	g.156482489T>G	NM_012206: c.102A>C	Synonymous	Leu34=	rs202050012	G=0.15867 (ExAC)	FALSE	-	-
CYFIP2	g.156820216delA	NM_014376: c.*208delA	3'UTR	-	rs33954943	—=0.3979 (TOPMED)	FALSE	-	-
SOX30	g.157079143C>T	NM_178424: c.-57G>A	5'UTR	-	rs369841660	T=0.0020 (1000G)	TRUE	Yes	0.003125 (3/960)
EBF1	g.158124682A>G	NM_024007: c.*1437T>C	3'UTR	-		-	TRUE	Yes	0 (0/768)
PWWP2A	g.159503427A>G	NM_052927: c.*1735T>C	3'UTR	-	rs569368848	G=0.0004 (1000G)	TRUE	Yes	0.0075 (3/192)
ZBED8	g.159820391C>T	NM_022090: c.*322G>A	3'UTR	-		-	TRUE	Yes	0 (0/768)

Table 6: Novel or rare, non-coding variants with CADD Phred score >10 identified at 5q33-35 in III:2.

Genomic change GRCh37/hg19	Gene	mRNA change	Location	rsID	MAF	Indigene	GA100K	GA100K India	CADD V1.6
chr5:g.153603947T>A	*GALNT10*	NM_198321:c.159+33362T>A	Intron	rs557958036	A=0.0026	0.0156	0.0046	0.0109	12.43
chr5:g.153605371T>C	*GALNT10*	NM_198321.3:c.159+34786T>C	Intron	rs1433617447		-	-	-	10.65
chr5:g.153606092A>G	*GALNT10*	NM_198321.4:c.159+35507A>G	Intron	rs578085750	G=0.0026	0.0146	0.005462	0.00134	15.64
chr5:g.153617822C>T	*GALNT10*	NM_198321.3:c.159+47237C>T	Intron	rs529537524	T=0.0026	0.0148	0.005175	0.0125	13.07
chr5:g.153648808T>C	*GALNT10*	NM_198321.4:c.160-25568T>C	Intron	rs559867144	C=0.0026	0.0147	0.00575	0.0142	12.52
chr5:g.153662288C>T	*GALNT10*	NM_198321.4:c.160-12088C>G	Intron	rs183237694	T=0.00014	0.0156	-	-	11.28

chr5:g.153666287G>C	*GALNT10*	NM_198321.4:c.160-8089G>C	Intron	rs573697057	C=0.0026	0.0146	0.005175	0.0125	11.14
chr5:g. 153667970_ 153667971GT (2)	*GALNT10*	NM_198321.4:c.160-6407TG[2]	Intron	rs1417464839	NA	-	-	-	15.55
chr5:g.154023501G>A			Intergenic	rs557860959	A=0.00274	0.0034	-	-	13.24
chr5:g.154023504C>T			Intergenic	rs577698296	T=0.0020	0.0034	-	-	13.35
chr5:g.154220113G>A	*FAXDC2*	NM_032385.4:c.1-2375C>T	Intron	rs933953537	NA	-	-	-	13.84
chr5:g.154505263C>T			Intergenic	rs913632939	T=0.00019	0.0024	0.000862	0.0008	15.72
chr5:g.154776009A>C			Intergenic	rs17117406	NA	0.1335	0.10782	0.1028	11.66
chr5:g.154836637T>C			Intergenic			0.001	0.000863	0.00088	14.22
chr5:g.157070731T>C	*SOX30*	NM_178424.1:c.1387+2914A>G	Intron	rs376901964	C=0.00009	0.002	0.001437	0.0033	10.1
chr5:g.157787457A>G	*LINC02227*	NR_109888.1:n.446+1953T>C	Intron			-	-	-	13.74
chr5:g.157905774C>T			Intergenic			-	-	-	16.26
chr5:g.158186924C>A	*EBF1*	NM_024007.4:c.1036+17497G>T	Intron				-	-	10.9
chr5:g.158328657A>T	*EBF1*	NM_024007.4:c.555-61539T>A	Intron	rs981292759	T=0.000016	-	-	-	10.69
chr5:g.159823728G>A	*ZBED8*	NM_022090.4:c.-49-1182C>T	Intron	rs949971500	NA	-	-	-	10.06
chr5:g.159823734C>A	*ZBED8*	NM_022090.4:c.-49-1188G>T	Intron	rs1009039398	NA	-	-	-	12.25
chr5:g.160267509T>A	*ATP10B*	NM_025153.2:c.-576+11439A>T	Intron			-	-	-	11.73
chr5:g.162341059C>G			Intergenic	rs369613283	NA		0.00489	0.0008	11.04
chr5:g.162976555A>G			Intergenic	rs542622633	G=0.00059	0.0015	0.000287	0.0008	15.57
chr5:g.163416845G>A			Intergenic	rs559156988	A=0.00259	0.0058	0.004032	0.01	12.08
chr5:g.164899684C>T			Intergenic	rs552373794	T=0.00030	0.001	-	-	10.11
chr5:g.165168773A>T			Intergenic	rs566767411	T=0.00256	0.0122	0.004315	0.0117	20.4
chr5:g.166929549T>C	*TENM2*	NM_001122679.1:c.502+127071T>C	Intron	rs528647855	C=0.0010	0.0103	0.0023	0.0059	15.42
chr5:g.167025527T>C	*TENM2*	NM_001122679.1:c.502+223049T>C	Intron	rs576861897	C=0.0014	0.0113	0.002587	0.0067	15.25
chr5:g.167171538A>G	*TENM2*	NM_001122679.1:c.503-131453A>G	Intron	rs370407114	G=0.0014	0.0020	0.0046	0.0109	16.79

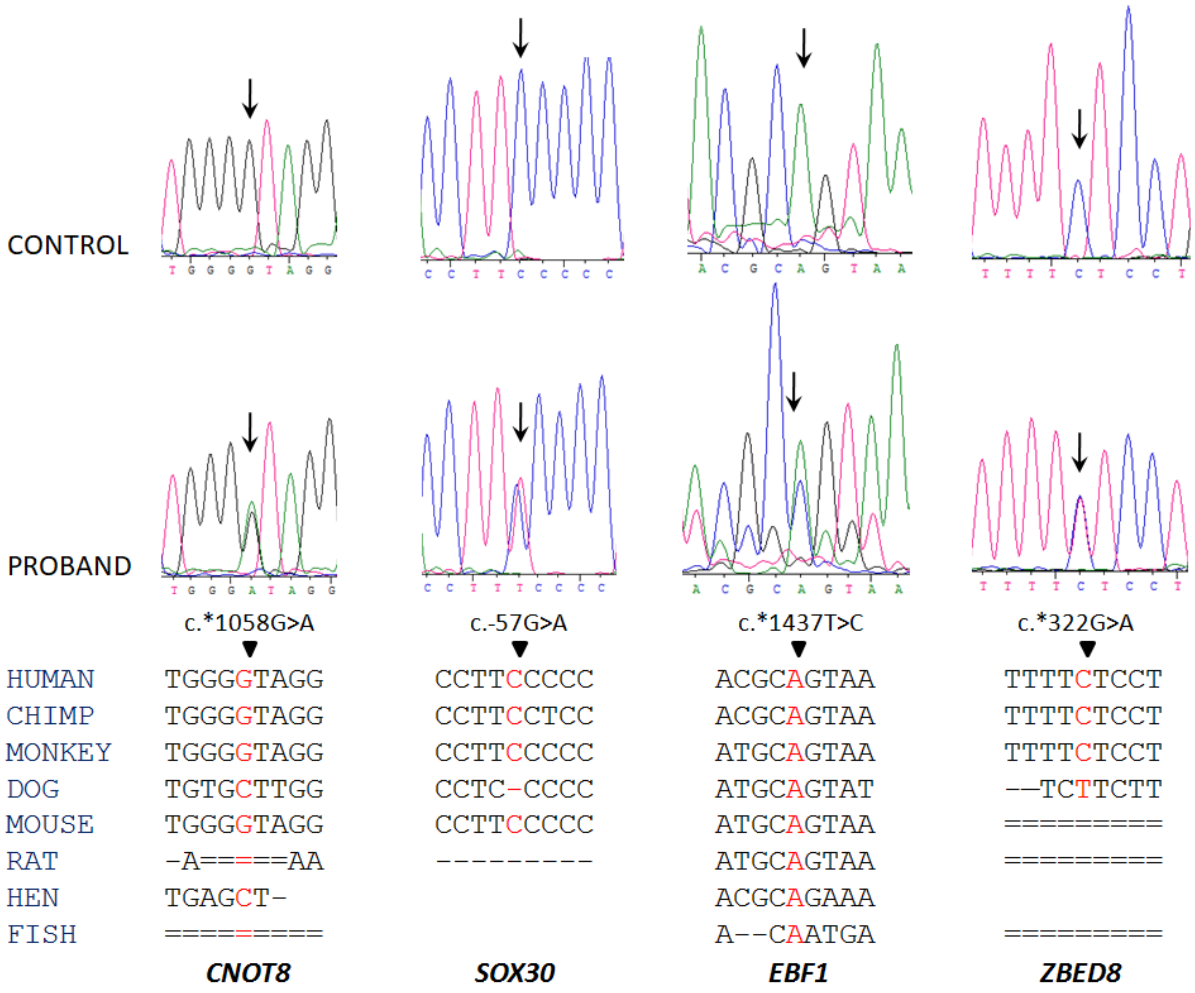


Figure 3: (A) Electropherograms of rare variants identified from the WGS study of the NIH34 in forward strand; (B) Conservation of the variant nucleotides across vertebrates obtained from the Multiz aligner.

2.4.2 Reporter assay of significant variants from WGS study

Bioinformatic analysis of the three 3'UTR variants did not identify the presence of any miRNA binding sites, H3K27Ac marks, ncRNA binding sites, UTR motifs, repeat regions at the variant locations. To find potential effects of the UTR variants, reporter luciferase assays was performed in cultured HEK293 and SH-SY5Y cells on the 5'UTR and 3'UTR variants. No significant difference was observed between the wild-type and 3'UTR variants in *ZBED8*, *CNOT8*, and *EBF1* (Figure 4). However, in the case of the *SOX30* 5'UTR, we found that the mutant c.-57G>A exhibited a higher level of luciferase expression as compared to the

wildtype. The increase in expression level was 2-fold in HEK293 and 3-fold in SH-SY5Y (Figure 4). This 5'UTR variant is present in a CpG island and DNaseI hypersensitive cluster indicating its regulatory function. No repeat sequences and epigenetic mark H3K27Ac was found. Transcription factors, SP1 and RBBP5 have been found to bind to this site, as determined from the ENCODE CHIP-Seq datasets. The 5' UTR variant (c.-57G>A) may affect the binding ability of these transcription factors leading to increased SOX30 expression. These findings point towards the possible role of *SOX30*, underlying JME in the family.

Gene	UTR	UTR length	Variant
EBF1	3'UTR	3196bp	+1437T>C
CNOT8	3'UTR	1353bp	+1058C>T
ZBED8	3'UTR	558bp	+322G>A
SOX30	5'UTR	940bp	-57G>A

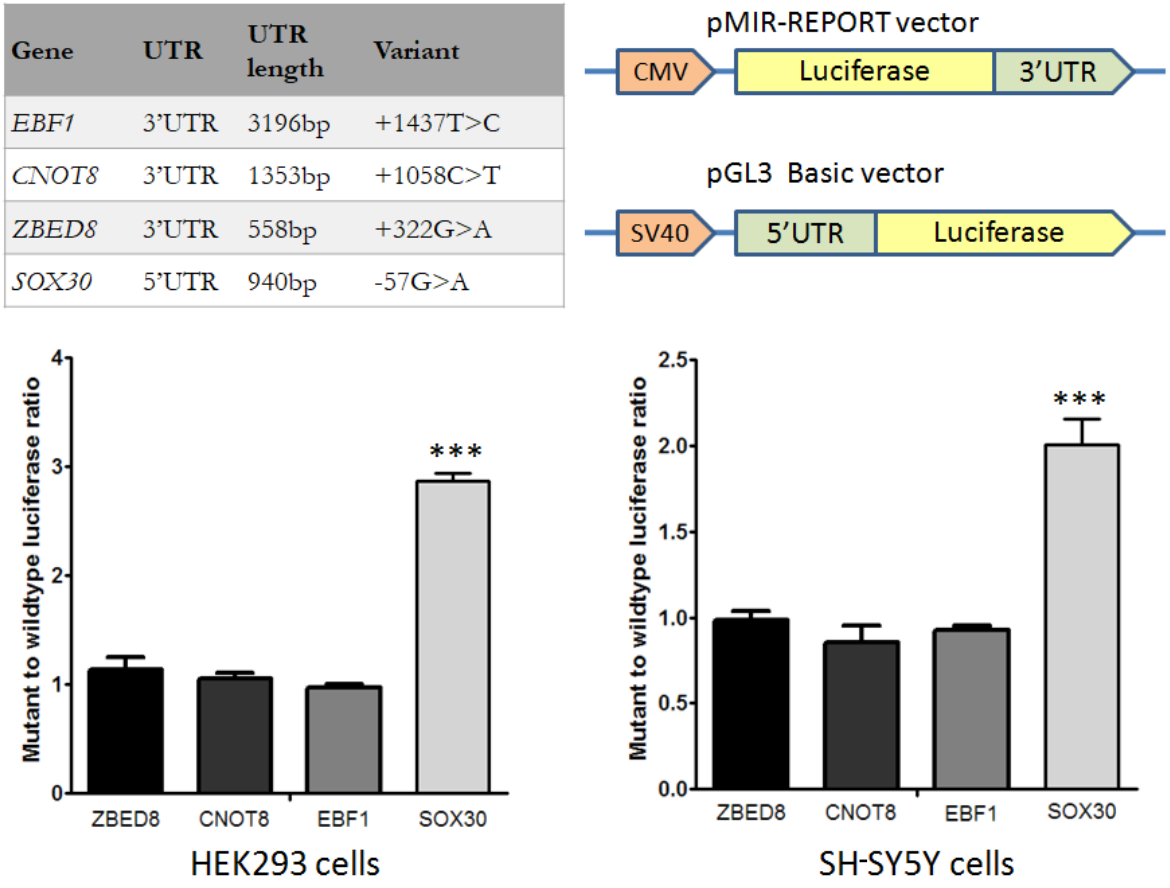


Figure 4: (A) Table lists the genes, length of the UTR cloned, and the variant (B) Schematic of vector used for reporter luciferase assays; (C) Luciferase reporter activity for 3'-UTR and 5'-UTR variants represented as normalized values of the variant to the wildtype. Statistical analyses were carried out on normalized values from three independent experiments using the one-way ANOVA ($p<0.0001$).

2.5 Discussion

JME is a relatively common class of genetic generalized epilepsy with substantial genetic basis to its aetiology. Family and twin studies have provided support to genetic predisposition underlying the disorder (Kjeldsen et al. 2005). There is high genetic heterogeneity, contributed in part due to the population-specific risk factors, due to which, in most JME cases, causative variants remain undetermined (Heinzen et al. 2012, Delgado-Escueta et al. 2013). GWAS employed to study non-mendelian factor for JME have identified mostly non-ion channel genes. Using linkage analysis in affected families, genes variants have been identified that co-segregated with Mendelian or near-Mendelian forms of JME.

Genome-wide linkage analysis of the Family NIH34 which consisted of seven JME affected individuals across four generations, inherited in an autosomal dominant pattern, identified significant linkage for the markers in the 5q33-35 region. The critical region was delimited by markers DS2012 and DS2075 comprising about 15Mb of the genome. Whole genome sequencing helped identify a previous unreported epilepsy gene, *SOX30* with a rare variant c.-57G>A segregating in the family, suggesting it as a potentially causative gene for JME in the family. The current study involved a comprehensive analysis of the disease-linked locus employing whole genome sequencing. Whole-genome sequencing on a father-daughter pair identified four rare variants in *EBF1, CNOT8, ZBED8* and *SOX30*. These variants were present in UTRs of the genes. All the four variants were rare in the ethnically matched control individuals and co-segregated with the disorder in the family.

The 5q33-35 locus harbours a set of the $GABA_A$ receptor genes which have been reported to be causative for epilepsy syndromes. Variants in *GABRA1, GABRA2, GABRA3, GABRB1, GABRB2, GABRB3, GABRD* and *GABRG2* have been identified in patients with epilepsy and epilepsy-associated disorders (Cossette et al. 2002, Harkin et al. 2002, Kananura et al. 2002,Dibbens et al. 2004, Srivastava et al. 2014, Epi4K Consortia 2016, Johannesen et al. 2016, Moller et al. 2017, Niturad et al. 2017,Shen et al. 2017, Zou et al. 2017,Orenstein et al. 2018, Baulac et al. 2020, Wallace et al. 2020). Mutations in $GABA_A$ receptors were enriched in a case-control exome study (May et al. 2018). GABR subunit coding genes are known to exhibit atypical chromosomal clustering due to their origin from a common ancestor by multi-layered gene duplications (Bailey et al. 1999, Simon et al. 2004). Among these, *GABRA1, GABRA6, GABRB2* and *GABRG2* are located at the 5q33-35 locus identified in this study. Hence were the first candidates to be analysed in the

family. These were examined by Sanger sequencing of the entire set of genes. Interestingly, no novel/rare variant was identified in these genes that segregated in the family. Whole-genome sequencing analysis carried out subsequently, also validated these observations. Whole-genome sequencing (WGS) was conducted for two members of the NIH34 family. WGS has a major benefit of representing the entire genome as compared to the exome sequencing. WGS has been reported to capture exonic variants better than WES due to unbiasedness, uniform coverage, read depth, data quality and lower false positives (Belkadi et al. 2015, Meienberg et al. 2016). WGS has also been reported to have identified disease-causing variants in patients previously undetected by the exome sequencing methods (Fresard et al. 2018, Shashi et al. 2018, Cho et al. 2020). WGS would also identify any structural and copy number variants that would have been overlooked previously. Given the identification of no amino acid coding nucleotide variant for this family in the previous study, it was important to look at the complete genome sequences. The proband and her unaffected father were chosen for the study since the marker haplotypes (Figure 1) indicated that the proband was likely to inherit the variant from the affected mother, and any variant inherited from the father would be inconsequential. The genes with important variants identified are discussed below.

EBF1 was first identified to be a regulator of *MB1* that codes for a membrane-bound antibody in early stages of B cell differentiation and ensures B cell lineage commitment by limiting GATA3 expression in lymphoid biased progenitors and therefore named Early B cell Factor (Hagman et al. 1991, Banerjee et al. 2013). Mice deficient in *Ebf1* have cells stalled at the pro-B cell stage, and no subsequent differentiated cells are present that are required for antibody production. Homozygous *Ebf1* knockout mice have severe growth defects that lead to prenatal lethality at around week four due to abnormal lipid metabolism (Hagman et al. 1993, Lin et al.1995). *EBF1* was earlier named Olfactory neuronal transcription factor (*OLF1*) since it drove the transcription of genes such as Olfactory marker protein (OMP) that is required for odour detection (Kudrycki et al. 1993, Wang et al.1993). *Ebf1* is involved in the development and transition of striatal cells from the subventricular zone to the mantle, differentiation of striatonigral neural projections and regulates axonal myelination (Garel et al. 1999, Lobo et al. 2006, Moruzzo et al. 2016). *EBF3*, a paralog of *EBF1,* is involved in nervous system development. Mutations in *EBF3* have been identified in individuals with epilepsy, intellectual disability, developmental delay, cerebellar ataxia, facial dysmorphism, and other co-morbidities (Chao et al. 2017, Harms et al. 2017, Sleven et al. 2017, Tanaka et

al. 2017). With these reports indicating *EBF1*'s involvement in neuronal development, it was considered a good candidate gene for JME in the family.

CNOT8 codes for 3'-5'exonuclease activity performing subunit of the CCR4-Not complex, which is involved in both transcription and mRNA processing that is required for gene expression (Miller et al. 2012, Collart 2016). It has overlapping functions with *CNOT7* in the complex with which it shares high sequence homology (Bianchin et al. 2005). Although ubiquitously expressed, the highest expression is observed in ovary, testis, spleen, and thymus, and its expression is down regulated during differentiation of neural stem cells (Chen et al. 2011). By deadenalysing, PMP22 and BTG/TOB proteins, it promotes cell proliferation in MCF breast cancer cells (Aslam et al. 2009, Doidge et al. 2012). CNOT8 is also required for the maintenance of the stability of rRNA clusters (Hosoyamada et al. 2019). A nonsense mutation in *cnot8* in the zebrafish model is known to cause increased differentiation of dopaminergic cells exclusively in the hypothalamus by affecting transcript expression of several FGF signalling pathway proteins. This indicated the involvement of *CNOT8* in the development of dopaminergic neurons (Koch et al. 2014). Therefore, *CNOT8* was also considered a candidate gene to cause epilepsy in this family.

ZBED8 is a relatively poorly studied gene. It shares homology with genes belonging to the hAT buster DNA transposon family. Unlike other members of the family, ZBED8 does not contain a zinc finger motif. ZBED family of proteins are involved in diverse processes (Hayward et al. 2013). Doxorubicin, an antineoplastic drug, was found to increase expression of ZBED8 in SH-SY5Y cells along with other Wnt signaling pathway proteins, demonstrating ZBED8 involvement in apoptosis (Suebsoonthron et al. 2017). *ZBED8* was also considered to be a likely candidate gene to cause JME in this family.

SOX30 belongs to the SRY-related HMG box containing family of proteins that are involved in the regulation of embryonic development and cell lineage determination. It is highly expressed in testes and very weakly expressed in other tissues in mice (Osaki et al. 1999). Our experiments suggested expression in various human brain regions (Ratnapriya 2009). Male mice lacking Sox30 were found to be infertile while the females were fertile. No other apparent phenotype was observed in these animals (Feng et al. 2017). Male germ cells were stalled at the round spermatid stage due to block in expression of post-meiotic spermatid specific genes (Bai et al. 2018, Feng et al. 2017, Zhang et al. 2018). *SOX30* promoter methylation caused non obstructive azoospermia due to impaired spermiogenesis (Han et al.

2014, Han et al. 2020). SOX30 was also found to autoregulate its expression and genes such as p53 affecting apoptosis, β catenin, and desmosomal genes affecting proliferation in lung cancer cells and tissues (Han et al. 2015, Han et al. 2018, Hao et al. 2018). To date, the role of *SOX30* in neuronal development has not been explored, while several other members of the SOX family proteins are known to be involved in neurogenesis. *SOX2* mutations in humans and *Sox1* in mice have been shown to cause epilepsy in addition to other brain pathological abnormalities (Malas et al. 2003, Sisodiya et al. 2006). The *SOX30* -57G>A variant was identified in 3 healthy individuals in from the southern parts of India and is present at low frequencies in databases (1000G, TOPMED, GnomAD). Reduced penetrance, genetic background and small biological effect size are possibly reasons as to why certain apparently healthy individuals harbour potentially pathogenic variants. *SOX30* is a variant tolerant gene hence the probability for the gene to accumulate variation is relatively high. Another example among GGE genes is *EFHC1* which is also a variant tolerant gene, where functionally defective missense variants are present in equal proportions in both patient and control cohort (Gonsales et al. 2020). It is interesting to note that all the four rare segregating variants in the family are in genes that are nucleic acid binding proteins, which directly or indirectly are involved in the regulation of other genes. The variants were present in the UTR regions of the genes. *Cis* and *trans*-regulatory sequences present in UTR regions are required for protein synthesis. Defects in sequence or structure of the UTR is known to cause several human disorders (Chatterjee et al. 2009). Mutations in miRNA binding sites located in 3'UTR of genes *GABRA3, GRM7, GABBR2, SOX11, MECP2, ADCY1,* and *ABCG2* are known for mesial temporal lobe epilepsy (Haenisch et al. 2015). 3'UTR variants in *SCN1A* in Dravet syndrome patients led to the formation of GAPDH binding sites impacting its mRNA stability (Zeng et al. 2014). None of these genes identified in this study were previously reported to be associated with any epilepsy syndromes; hence, it was important to identify the variant's contribution to disease. Since access to brain biopsy tissue is not feasible, as the first line of study conducted to interpret the functional consequence of the identified four variants were, luciferase reporter assays. The results indicated that variants in 3'UTR of genes, *ZBED8, CNOT8,* and *EBF1* did not affect their transcriptional capacity compared to the wildtype counterpart, while the *SOX30* variant situated in the 5'UTR of the gene exhibited enhanced transcriptional activity in cultured HEK293 and SH-SY5Y cells. The -57G>A variant led to a significant increased expression of the downstream luciferase gene suggesting this variant to be likely causative of JME in this family.

Chapter 3

Functional characterization of SOX30 rare alleles identified among JME patients

3.1 Summary

Identification of *SOX30* as a candidate juvenile myoclonic epilepsy gene led us to examine additional JME patients for variants in *SOX30*. Fifteen rare heterozygous coding variants were identified in *SOX30* of which 14 were missense and one was a nonsense variant. Two of these variants were in the HMG domain while the remaining were distributed in the rest of the protein. *In silico* prediction tools indicated, these variant's effect to vary from benign to damaging. To evaluate their potential impact on SOX30 function, *in vitro* overexpression experiments of the wild type and variant forms of *SOX30* were carried out in cultured cells, wherein they were examined for subcellular localization and in reporter luciferase assays. The missense variants did not affect the protein's exclusive nuclear localization while the nonsense mutation p.Trp404Ter exhibited both cytoplasmic and nuclear localization. SOX30 wildtype protein repressed downstream gene expression. The transcriptional activity of the variants was assessed by reporter gene activity and the variants p.Pro82Arg, p.Pro353Arg, and p.Trp404Ter resulted in a complete loss of transcriptional activity in both HEK293 and SH-SY5Y while variants p.Pro123Thr and p.Ser611Pro exhibited differences in transcriptional modulations. To investigate SOX30's regulatory functions in a neural cell line, I employed chromatin immunoprecipitation coupled with sequencing to identify genomic sites that are directly bound by SOX30 protein in cultured SH-SY5Y cells. Among the targets represented in two replicates, high ranking genes *ZSCAN9, STX16, BRWD1* and *BCL2L1* were validated to be SOX30 targets by ChIP-PCR and qRT-PCR in cells overexpressing SOX30. These genes have varied functions and are known to be involved in cell proliferation, differentiation, and apoptosis pathways. Any of the hitherto known epilepsy genes were not the direct targets of SOX30.

3.2 Introduction

Studies on the 5q34 locus presented in the previous chapter suggested a role of *SOX30* for JME in the family. SOX30 is an HMG box-containing transcription factor that belongs to SRY - related family of proteins, well known to be involved in embryogenesis and organogenesis. While mutations in the SOX family proteins have been implicated in neurological disorders, the role of SOX30 has not been reported for a clinical neurological manifestation so far. Reverse transcription PCR and western analysis has indicated the expression of SOX30 in

human brain tissues. We found 15 additional rare *SOX30* coding variants consisting of 14 missense, and one nonsense variant, among 480 JME patients screened. The amino acid at these variant locations were mostly conserved across species suggesting that they have a biological role. To examine possible functional effects, nuclear localization, and reporter assays for transcription regulatory capacities of the variants, were conducted. These initial studies may help guide further work to establish the genetic mechanism by which SOX30 and its variants lead to JME.

3.3 Materials and methods

3.3.1 Bioinformatic analysis of the SOX30 rare variants

Bioinformatic prediction tools available as web-based applications were employed to predict functional aspects of the *SOX30* variants. ClustalOmega was used to align protein sequences of all SOX family genes to examine conservation of the amino acids. Evolutionary conservation model-based tools used to predict the effect of the variant included Mutation assessor, FATHHM, Panther, PhD-SNP, and SIFT. SNP&GO, UMD predictor, ENVISION, and SNAP2, utilize protein structure, variant location, and biochemical function to predict the effect of the variant. Multiple sequence alignment-based tool, PROVEAN and tools that use a combination of sequence conservation and structure that included Mutationtaster, MutPred2, and PolyPhen2 were also used to determine effect of variants. CADD scores that is derived by contrasting natural variants with simulated variants were calculated.

3.3.2 Vector construction and site-directed mutagenesis

The human SOX30 full length clone (cDNA clone MGC:34408 IMAGE:5172577) in pCMV-SPORT6 vector was PCR amplified with *Taq* polymerase (NEB) and cloned between *NheI* and *HindIII* into pcDNA 3.1 (Invitrogen, Carlsbad, USA), *HindIII* and *EcoRI* into pCMV-3XFLAG-10 (Sigma-Aldrich) and *EcoRI* and *SalI* into pCMVTag4a (Stratagene, La Jolla, USA). The mCherry cDNA was inserted between between *NheI* and *EcoRI* at 5' end of SOX30 cDNA cloned into pCMVTag4a (Table S3.1).

Site-directed mutagenesis was performed on the pcDNA-SOX30 to introduce variants identified using the Quikchange SDM kit (Stratagene). Fifteen missense variants were incorporated individually in pcDNA-SOX30 (Table S3.2). The p.Trp404Ter variant was introduced into 3XFLAG tagged SOX30 encoding vector (Table S3.2). The constructs

generated, were transformed into DH5α or XL10-Gold competent *Escherichia coli* strains. Small-scale preparations and purifications of plasmid DNA were done using plasmid miniprep kit (QIAGEN) and were confirmed by Sanger sequencing (Table S3.3, S3.4).

For luciferase assays, an oligonucleotide containing four copies of the SOX30 binding site (5' GAGACAATGGGACAATGGCGAGACAATGGGACAAT 3') was cloned into *NheI* and *XhoI* restriction enzyme sites of pGL3 promoter vector (Promega, USA). The vector sequences were confirmed using primers RV3 5´-d(CTAGCAAAATAGGCTGTCCC)-3´and GL25′-d(CTTTATGTTTTTGGCGTCTTCCA)-3′. pCMV-β-Galactosidase was used as a control for transfection efficiency.

For Chip-Seq experiments, 3XFLAG-SOX30 cDNA was restriction digested using primers *NdeI* and *EcoRI* from its parent plasmid and inserted into pIRES2-EGFP generating pIRES-EGFP-3XFLAG-SOX30 vector.

3.3.3 Cell culture and transfection

HEK293 and SH-SY5Y cell lines obtained from ATCC were grown in Dulbecco's modified eagle medium (DMEM) supplemented with 10% heat-inactivated fetal bovine serum (USA Origin, Sigma Aldrich), 2mM L- glutamine, 100U/ml penicillin and 0.1mg/mg streptomycin (Sigma Aldrich) in a humidified chamber with 5% CO_2 at 37°C. Cells were grown to about 40% confluency before transfection was done. Lipofectamine 2000 (Thermo Fischer) was incubated with the respective plasmids to form Lipo-DNA complexes, which were then added to cells in serum-free media. Six hours post-transfection, the media was replaced with serum-supplemented media. The cells were collected 24 hours post transfection for immunocytochemistry, western analysis and reporter luciferase assays.

3.3.4 Immunocytochemistry

HEK293 cells were grown on poly-L-lysine coated coverslips. At about 40% confluency, cells were transfected with SOX30-wildtype or SOX30-variant construct. Twenty-four hours post-transfection, the coverslips were washed once with 1X PBS followed by fixation using 4% paraformaldehyde for 20 minutes. Cells were then washed in 1X PBS for 5 minutes twice, followed by permeabilization using 0.1% Triton X 100 for 15 minutes.

Blocking was performed to eliminate non-specific signals using 5% bovine serum albumin for 45 minutes. SOX30 -specific, rabbit -raised polyclonal antibody (Sox-30 (H-300): sc-20104, SCBT) at 1:100 dilution was used to study the localization of wildtype and missense mutant constructs. Anti-FLAG M2 monoclonal mouse raised antibody (F1804, Sigma-Aldrich) at 1:1000 dilution was used to study wildtype and nonsense p.Trp404Ter variant proteins. Cells were incubated with the appropriate antibody for 1 hour. Cells were then washed twice with 1X PBS and incubated with secondary IgG antibody raised in either in mouse or rabbit, conjugated to Alexafluor 568 (Molecular probes, Oregon, USA) at 1: 500 dilution for 1 hour. Cells were then washed twice in 1X PBS followed by nucleus staining using 1ug/ml DAPI for 15 minutes and then rewashed in 1X PBS for 10 minutes thrice to removed excess stain.

The above steps were conducted at room temperature. The solutions were made in 1X PBS, while the antibody dilutions were made in 1% BSA solution. Coverslips containing antibody labeled cells were mounted onto glass slides using 70% glycerol, and the edges of the coverslip were sealed using transparent nail polish. Confocal images of the labeled cells were then obtained using either LSM 510 meta or LSM 880 (Carl Zeiss, Oberkochen, Germany) at 63X magnification.

3.3.5 Nuclear cytoplasmic fractionation

HEK293 cells were grown in 6-well dishes. They were transfected with empty vector pCMV-3XFLAG-10, 3X FLAG-SOX30 wildtype or with SOX30 nonsense allele, p.Trp404Ter. Twenty-four hours post-transfection cells were collected by centrifugation at 1500rpm for 2 minutes and washed with ice-cold 1X PBS twice. Cells were then resuspended into 200µl hypotonic solution containing 20Mm Tris-HCl, pH 7.4, 10mM NaCl, and 3mM $MgCl_2$supplemented with 1mM PMSF and 1:1000 dilution of protease inhibitor cocktail (Sigma Aldrich) and incubated on ice for 30 minutes. The hypotonic solution and cell suspension mixture was supplemented with 25µl of 10% NP40 detergent and vortexed for 10 seconds followed by centrifugation at 3000rpm for 15 minutes at 4°C. The supernatant containing the cytoplasmic fraction was collected and stored separately at -20°C.

The nuclear pellet was washed twice in ice-cold hypotonic solution to remove residual cytoplasmic contents at 3000rpm for 5 minutes at 4°C. The pellet was then resuspended in 80µl of cell extraction buffer containing 10 mM Tris, pH 7.4, 2 mM Sodium orthovanadate (Na_3VO_4), 100 mM NaCl, 1% Triton X-100, 1 mM EDTA, 10% Glycerol, 1 mM EGTA, 0.1%

SDS, 1 mM NaF, 0.5% Sodium deoxycholate ($C_{24}H_{39}NaO_4$) and 20 mM Sodium pyrophosphate tetrabasic ($Na_4P_2O_7$) supplemented with 1mM PMSF and 1:1000 dilution of protease inhibitor cocktail (Sigma Aldrich) and incubated on ice for 1 hour with intermittent vortexing. The solution was centrifuged at 13000rpm at 4°C for 45 minutes. The supernatant containing the nuclear fraction was collected and stored at -20°C till immunoblotting was performed.

3.3.6 Western blotting

Whole-cell lysates were prepared from 6-well dishes by resuspending cells in 200µl RIPA buffer (50Mm Tris-HCl pH 7.4, 150mM NaCl, 1mM EDTA, 0.5% sodium deoxycholate, 1% NP-40 and 1% SDS) supplemented with protease inhibitor cocktail at 1:1000 dilution. It was incubated for 2 hours on ice with intermittent vortexing. The lysates were homogenized by passing through 1mL syringes ten times followed by incubation in ice for 15 minutes. This step was repeated thrice and then centrifuged for 45 minutes at 4°C at 13,000rpm. Lysate protein concentration was measured using bicinchoninic acid (BCA) assay (Sigma Aldrich).

20µg of whole cell lysate or 20µl of nuclear/cytoplasmic lysate was mixed with 3µl of 6X SDS gel loading dye and placed in boiling water for 10 minutes. The samples were then loaded into wells of 12% SDS containing polyacrylamide gel followed by vertical electrophoresis (BioRad, Hercules, USA). Protein separated in the gel were then transferred onto 0.2µm pore sized distilled water activated nitrocellulose membrane (PALL Corp, New York, USA) at 20V for 1 hour in transfer buffer (25mM Tris HCl, 192mM glycine, 20% methanol and 0.036% SDS) using semi-dry transfer system (BioRad).

The membrane was blocked in 5% skimmed milk for 1 hour at room temperature. Blocking was followed by incubation with 1:5000 dilution of Anti-FLAG antibody (F1804, Sigma Aldrich), 1:10,000 dilution of anti-histone H3 (tri methyl K4) antibody (ab8580, Abcam, Cambridge, UK) and 1:5000 dilution of anti-alpha tubulin antibody (Sigma Aldrich) in 1% BSA solution with 0.05% Tween20 4°C overnight. The membrane was then washed thrice in 1X PBS with 0.05% Tween20 for 10 minutes each at room temperature followed by secondary antibody staining using anti-mouse or anti-rabbit IgG conjugated to horseradish peroxidase (HRP) followed by a repeat of the wash step thrice. The membrane was then treated with the enhanced chemiluminescent substrate (ECL) for HRP (Pierce, Illinois, USA) for 5 minutes and exposed to either X-ray film or imaged using a Versadoc imaging system (BioRad)

3.3.7 Reporter luciferase assay

HEK293 and SH-SY5Y cells were plated in 12-well plates. 500ng of wildtype SOX30 or variant SOX30 expressing vector were co-transfected with 100ng of pGL3 promoter with SOX30 binding site and 50ng of β-galactosidase expressing vector. In parallel, only empty vector along with pGL3 promoter and SOX30 wildtype vector was also transfected in separate wells as controls. Twenty-four hours post-transfection, 100µl 1X reporter lysis buffer (E3971, Promega, USA) was added to the cells and was placed on ice for 30 minutes for lysis to occur. The lysate was collected and centrifuged at 13,000rpm at 4°C for 15 minutes. 2µl of the lysate was added to 10µl of luciferase assay reagent (E1500, Promega), vortexed immediately followed by measurement of luciferase activity in a luminometer (Berthold Detection Systems) with a 10-s premeasurement delay and repeated values were captured till saturated values are obtained.

Beta-galactosidase activity reading was measured by mixing 20µl of lysate to 50µl of 2X assay buffer (Promega) that contains ortho-Nitrophenyl-β-galactoside (ONPG) and is incubated at 37°C for 45 minutes. The reaction was halted by adding 100µl of the stop solution containing 1M sodium bicarbonate to the reaction mixture. The β -galactosidase activity was measured using an ELISA plate reader (Molecular Devices) at 420nm. All experiments were performed at least three times, producing qualitatively similar results. Data in each experiment are presented as the mean ± SEM of triplicates from a representative experiment.

3.3.8 Transient overexpression in cultured cells and fluorescent-activated cell sorting (FACS) for ChIP

SH-SY5Y cells grown in 100mm cell culture dishes were transfected with 15µg of pIRES-EGFP-3XFLAG-SOX30 vector when cells were at about 40% confluent in serum-free media. Six hours post-transfection, the media was replaced with 10% fetal bovine serum-containing media. Twenty-four hours post-transfection, cells were washed twice with plain DMEM media and collected in 2% FBS containing media in a FACS collection tube. GFP-positive cells were collected using FACS Aria III cytometer (Beckton Dickinson, New Jersey, UK), after gating the negative controls. Collected cells were then formaldehyde crosslinked, and stored at -80°C.

3.3.9 Antibody and magnetic bead conjugation

Anti-FLAG mouse-raised monoclonal antibody (F1804, Sigma-Aldrich) for FLAG-tagged SOX30 protein and mouse-raised IgG (kch-819, Diagenode, New Jersey, USA) control

antibody was used for chromatin immunoprecipitation. The antibody was validated using immunocytochemistry, western blotting of cells overexpressing FLAG-tagged SOX30 and western blotting of immunoprecipitated samples of cells overexpressing 3XFLAG-SOX30 for antibody magnetic bead conjugation (Figure 1 B, C).

50µl of Protein G-coated magnetic beads were washed with 100µl ice cold ChIP buffer C1 by mixing the solution gently and placing them in the magnetic rack for 1 minute till the supernatant is clear and is discarded. After washing, the beads were resuspended in 110µl of ChIP buffer C1. 10µg of FLAG antibody or 1µg of IgG was added to the beads and incubated in a rotating wheel for 4 hours at 4°C at 40rpm. Just before immunoprecipitation, the bead solution was placed in an ice-cold magnetic rack for 1 minute, and the antibody-conjugated beads were collected after discarding the supernatant.

3.3.10 Chromatin Immunoprecipitation (ChIP)

The Highcell# protein G ChIP kit (Diagenode) was used for chromatin immunoprecipitation with modifications to the protocol suggested by the manufacturer. 5 x 10^5 sorted 3XFLAG-SOX30 SH-SY5Y positive cells were collected in 1.5ml centrifugation tubes. The cell suspension was washed twice in 1 X PBS. The cells were resuspended in 500µl 1X PBS and subjected to chromatin crosslinking by adding 13.5µl of 36.5% formaldehyde (final concentration 1%) and incubated for 8 minutes at room temperature. The reaction was stopped by the addition of 57µl of 1.25M Glycine followed by incubation for 5 minutes with intermittent gentle vortexing. The fixed cells were centrifuged at 1500rpm for 5 minutes at 4°C to pellet the suspension. The cells were washed twice with 500µl ice-cold 1X PBS by centrifuging the cell suspension at 1600rpm for 5 minutes at 4°C. Cells were then lysed using 500µl ice-cold Lysis buffer L1 and resuspended by pipetting followed by incubation for 10 minutes at 4°C with gentle mixing. The nuclear pellet was then lysed using 500µl ice cold lysis buffer L2 for 10 minutes at 4°C followed by centrifugation at 1600 rpm for 5 minutes at 4°C and the supernatant containing cell debris was discarded. The chromatin pellet was resuspended in 200µl of Shearing buffer containing protease inhibitor at 1:200 dilution. The chromatin was sonicated using Biorupter Plus (Diagenode) with 30 seconds ON/OFF each cycle with a break every five cycles for 60 cycles in high power setting with vortexing and short spin every five cycles. The chromatin was consistently maintained below 4°C during sonication. The sonicated chromatin was diluted by the addition of 800µl of ChIP Buffer C1

and 5µl of protease inhibitor cocktail. Twenty microlitres of the diluted chromatin was kept aside as input.

For magnetic immunoprecipitation, the 980µl of diluted chromatin was spun at 12000rpm for 10 minutes. The supernatant was collected and transferred to antibody coated Dynabeads (Thermo Fisher) and incubated on constant rotation of 40rpm at 4°C overnight. The beads, now bound to chromatin were subjected to washes twice with ice cold ChIP Buffer C1 and once with Buffer W1 by adding 500µl of the respective solutions followed by incubation for 5 minutes at 4°C in a rotating wheel at a speed of 40rpm and then placed in a magnetic rack to remove the wash supernatant. The magnetic beads and input samples were then processed simultaneously by making up the volume to 200µl of Tris EDTA pH 8.0 containing 0.1% SDS. The samples were then incubated in a thermomixer at 65°C overnight to reverse the crosslinking. To degrade the residual RNA, 4µl of RNase H (Sigma Aldrich) was added and incubated at 37°C for 2 hours and 8µl of Proteinase K is added and incubated at 55°C for 2 hours to degrade the proteins. The remaining nucleic acid was purified using a PCR purification kit (QIAGEN) by mixing the sample with 1ml PB buffer and mixed gently and then transferred to a spin column and centrifuged at 13000rpm for 1 minute. 750µl of PE buffer was added to the column to perform alcohol wash on the column and spun at 13000rpm for 1 minute.

11µl of warm elution buffer (EB) was added to the column and spun at the same speed to collect the eluant. Immunoprecipitated DNA and its respective un-immunoprecipitated input control sample's quality was estimated by a Qubit fluorometer (Thermo Fisher). Sonication quality and size of DNA fragments were checked by agarose gel electrophoresis before immunoprecipitation. The samples were then processed further by next-generation sequencing.

For the ChIP – qPCR, post- IP washes, samples were resuspended in 100µl of 0.1% SDS containing TE buffer and incubated with 1µl Proteinase K at 55°C for 15 minutes followed by incubation at 99°C for 15 minutes. The samples are then collected at the bottom of the tube by a quick spin and then placed in the magnetic rack to collect the DNA.

3.3.11 ChIP-Seq

ChIP-seq libraries were generated for six samples (3 ChIP and 3 input controls) using NEBNext Ultra II DNA Library Prep Kit (NEB) from sheared immunoprecipitated DNA ranging from 7

to 35ng according to manufacturer's protocol for 2x150 pair-end sequencing on a HiSeq 2500 (Illumina).

3.3.12 ChIP-Seq bioinformatics pipeline

FASTQC and MultiQC (Ewels et al. 2016) software were used to check the quality of the sequence output and to obtain QC metrics including base call quality distribution, percent of bases above Phred score 20 and 30. Low-quality bases and adapter sequences were removed using fastp (Chen et al. 2018). The quality-filtered reads were mapped to the human genome GRCh38 reference assembly using bowtie2 (Langmead and Salzberg 2012). PCR duplicates were removed by eliminating multiple reads with the same start site using PICARD tools (http://broadinstitute.github.io/picard/). Samtools was used to discard multi-mapped reads, and only stringent single aligned sequences were retained. Non-redundant, uniquely mapped reads were then used for detecting ChIP enriched peaks using MACS2 (v2.1.2) with a mfold of [5, 50] and a p-value cutoff of 0.01 (Zhang et al. 2008). Peaks were filtered to remove ENCODE blacklisted genomic regions; non- human chromosome mapped peaks (Amemiya et al. 2019). Irreproducible discovery rate (IDR) with a global threshold of 0.05 was applied on the filtered peak data to identify reproducible peaks. The peaks were annotated using ChIP seeker (Yu et al. 2015)

3.3.13 qRT-PCR

For validation of potential ChIP targets obtained, ChIP-qPCR primers were designed spanning the peak region located in 5' UTR or upstream regions of genes of interest (Table S3.5). For measuring ChIP target transcript (cDNA) expression differences between cells expressing wildtype SOX30 and p.Trp404Ter variant, qRT-PCR primers were designed to overlap an exon-exon junction with amplicon size ranging between 75-200 bp (Table S3.5). Each qPCR reaction contained 10µl of SYBR Green Master, 0.25µl of 25µM forward and reverse primer each, 0.5µl of cDNA or 2µl of ChIP or input DNA and the volume is made up to 20µl with double distilled water. The conditions of PCR were as follows, initial denaturation at 95°C for 10 minutes, 40 cycles of denaturation at 95°C for 15 seconds and annealing at 60°C for 1 minute. The melt curve protocol followed with 5 seconds at 65°C and then 5 seconds each at 0.5°C increments between 65°C and 95°C.qPCR was done using FastStart Universal SYBR Green Master (Rox) (Roche, Basel, Switzerland) using the primers listed in Table (Table S3.5). All qPCR reactions were performed in triplicates on CFX96 Touch Real-Time PCR Detection System (BioRad). Data were analyzed using the comparative Ct method.

3.4 Results

3.4.1 Bioinformatic analysis of SOX30 and its rare variants

The longest SOX30 isoform codes for a 753 amino acid protein, while the smaller isoform codes for a 501 amino acids long protein. Between amino acids, 337 – 405 lies the conserved DNA-binding domain identified as the high mobility group (HMG) box, which SOX30 shares with its family of proteins. SOX30 contains proline-rich regions at the N- and C- terminal ends of the proteins, between 6 – 41 and 564 – 646 amino acids that are also found in several transcription factors, wherein they are usually involved in interaction and trans-regulation of the protein (Figure 1A). Multiple sequence alignment of the SOX family proteins identified several highly conserved amino acids in SOX30 that are involved in nuclear localization, DNA binding, DNA bending and protein interaction (Figure 4).

While JME rare coding variants were all novel at the time of their identification, now with the availability of several variant databases, all except two have been reported. All the variants have their minor allele frequency less than 0.005 in global databases, with the highest MAF being 0.00219. In south Asian population databases, except for p.Asn667Ser, which had a MAF of 0.00571, all others have their minor allele frequency less than 0.005. Among the in-house controls, all the variants had their MAF below 0.005, with the highest MAF being 0.004167 for the missense variant p.Met645Ile. In the Epi25K database, variations p.Pro564Ala, p.Tyr638Cys, and p.Asp667Ser had comparable alleles frequencies in cases and controls. Variants p.Val571Phe and p.Met645Ile were found in epilepsy cases only, while rest were not listed in the database (Table 1). *In silico* prediction of deleteriousness of the variants ranged from being polymorphic to pathogenic. PROVEAN predicted all the variants to be neutral, while FATHHM predicted all to be damaging. No single variant was unanimously predicted to be pathogenic. p.Ala228Thr variant was predicted to be mostly benign. Variants p.Asp654Asn and p.Tyr638Cys had the highest number of bioinformatic tools predict as being damaging, while variants p.Pro353Arg, p.Val571Phe, and p.Ser611Pro variants were next in rank of being pathogenic. Variants present in the C-terminal region of the protein were mostly predicted to be damaging. Nonsense mutation p.Trp404Ter was predicted to be deleterious by the tools applicable to it (Table 2.3)

Table 1: Minor allele frequencies of *SOX30* rare coding variants

	Global population database					Asian subpopulation database			Disease database
Protein change	**dbSNP**	**GnomAD**	**TOPMED**	**1000g**	**EVS**	**In-house control (960)**	**GenomeAsia 100k**	**INDEX-DB**	**EPI25K**
p.Pro82Arg	rs772412777	0.000015	absent	absent	absent	0.001042	absent	absent	absent
p.Pro123Thr	rs182220520	0.000722	0.000366	absent	absent	0	0.000575	absent	absent
p.Pro123Ser	rs182220520	0.000011	0.000032	0.001	absent	0	absent	absent	absent
p.Ala228Thr	rs747852816	0.000132	absent	absent	absent	0	absent	absent	absent
p.Ala231Pro	rs779659102	0.000016	absent	absent	absent	0.001042	absent	absent	absent
p.Pro353Arg	rs772733963	0.000016	absent	absent	absent	0	absent	absent	absent
p.Trp404Ter	absent	absent	absent	absent	absent	0	absent	absent	absent
p.Pro564Ala	rs138390114	0.000633	0.000319	0.001	0.000615	0	0.000287	absent	Case-0.000218, Control-0.000119
p.Val571Phe	rs370749800	0.000014	0.000048	0.0002	absent	0	absent	absent	Case-0.0000545
p.Ser611Pro	rs1202430003	absent	0.000008	absent	absent	0	absent	absent	Absent
p.Tyr638Cys	rs147847969	0.000086	0.000064	absent	0.000077	0	absent	absent	Case-0.000109, Control-0.000119
p.Met645Ile	rs368081998	0.000172	absent	0.0002	absent	0.004167	0.00115	absent	Case-0.0000545
p.Asp654Asn	absent	absent	absent	absent	absent	0.001042	absent	absent	absent
p.Asn667Asp	rs748365237	0.000064	absent	absent	absent	0.002083	0.000575	0.00286	absent
p.Asn667Ser	rs142156325	0.002190	0.002158	0.002	0.002076	0	0.000575	0.00571	Case-0.00463, Control-0.00468

Table:2. *In silico* analysis of *SOX30* rare JME variants - part 1

Protein change	Polyphen2	PROVEAN	SIFT	MutationTaster	PANTHER (preservation time)	FATHMM	Mutation Assessor FIS
p.Pro82Arg	Probably damaging 0.998	Neutral 1.13	Damaging 0	polymorphism 0.00017	probably benign 176	Damaging 4.95	low 1.79
p.Pro123Thr	Benign 0.244	Neutral 0.74	Damaging 0.001	polymorphism 0.00015	probably benign 176	Damaging 4.89	low 1.79
p.Pro123Ser	Benign 0.012	Neutral 0.5	Damaging 0.002	polymorphism 0.00017	probably benign 176	Damaging 4.87	low 1.79
p.Ala228Thr	Benign 0.001	Neutral 0.06	Tolerated 0.993	polymorphism 0.00000000000221	probably benign 91	Damaging 4.1	neutral 0.55
p.Ala231Pro	Possibly damaging 0.876	Neutral 0.88	Damaging 0.001	polymorphism 0.00000000013	possibly damaging 324	Damaging 4.44	low 1.895
p.Pro353Arg	Possibly damaging 0.895	Neutral 1.58	Tolerated 0.274	disease causing 0.99986	possibly damaging 325	Damaging 5.56	neutral 0.495
p.Trp404Ter	NA	NA	NA	disease causing 1	possibly damaging 325	NA	NA
p.Pro564Ala	Probably damaging 0.988	Neutral 2.13	Damaging 0.009	polymorphism 0.02208	possibly damaging 325	Damaging 4.25	medium 2.25
p.Val571Phe	Probably damaging 0.971	Neutral 1.11	Damaging 0	polymorphism 0.06990	possibly damaging 324	Damaging 4.5	medium 2.14
p.Ser611Pro	Benign 0.029	Neutral 1.35	Damaging 0	polymorphism 0.49092	possibly damaging 325	Damaging 4.63	neutral 0.55
p.Tyr638Cys	Possibly damaging 0.653	Neutral 1.56	Tolerated 0.197	polymorphism 0.37703	possibly damaging 325	Damaging 4.68	low 0.975
p.Met645Ile	Benign 0.278	Neutral 0.88	Damaging 0.003	polymorphism 0.23509	possibly damaging 325	Damaging 4.31	low 0.895
p.Asp654Asn	Probably damaging 1.000	Neutral 1.13	Damaging 0	disease causing 0.99575	possibly damaging 325	Damaging 4.78	low 0.975
p.Asn667Asp	Benign 0.188	Neutral 1.01	Tolerated 0.209	polymorphism 0.05408	possibly damaging 325	Damaging 4.36	low 0.975
p.Asn667Ser	Probably damaging 0.988	Neutral 0.98	Tolerated 0.052	polymorphism 0.06165	possibly damaging 325	Damaging 4.54	low 0.975

Table 3: *In silico* analysis of *SOX30* rare JME variants- part 2

Protein change	PhD-SNP	CADD score	UMD predictor	SNP&GO	MutPred2	Envision	SNAP2
p.Pro82Arg	Disease 3	22.8	Polymorphism 5	Neutral 6	0.192	0.766963	effect 55
p.Pro123Thr	Disease 1	14.21	Polymorphism 45	Neutral 8	0.148	0.920551	effect 14
p.Pro123Ser	Disease 0	12.24	Probable polymorphism 60	Neutral 9	0.138	0.97443	effect 7
p.Ala228Thr	Neutral 8	0.001	Polymorphism 39	Neutral 10	0.039	0.993951	neutral 67
p.Ala231Pro	Disease 6	8.092	Polymorphism 36	Neutral 4	0.311	0.869338	effect 7
p.Pro353Arg	Neutral 2	23.5	Pathogenic 100	Disease 2	0.767	0.756688	neutral 73
p.Trp404Ter	NA	39	NA	NA	0.56186	NA	NA
p.Pro564Ala	Neutral 1	22.8	Polymorphism 38	Neutral 5	0.116	0.83626	effect 19
p.Val571Phe	Disease 4	20.1	Probable polymorphism 57	Neutral 1	0.173	0.8463	effect 32
p.Ser611Pro	Disease 7	20.1	Pathogenic 75	Disease 3	0.455	0.946159	effect 27
p.Tyr638Cys	Disease 4	22.7	Pathogenic 93	Disease 3	0.662	0.800717	effect 15
p.Met645Ile	Disease 6	22.6	Probable polymorphism 60	Disease 1	0.417	0.944517	effect 37
p.Asp654Asn	Disease 8	28.1	Probable polymorphism 63	Disease 2	0.38	0.814897	effect 44
p.Asn667Asp	Disease 4	21	Probably pathogenic 66	Neutral 2	0.186	0.935386	effect 16
p.Asn667Ser	Disease 7	23	Polymorphism 44	Disease 1	0.149	0.831676	effect 14

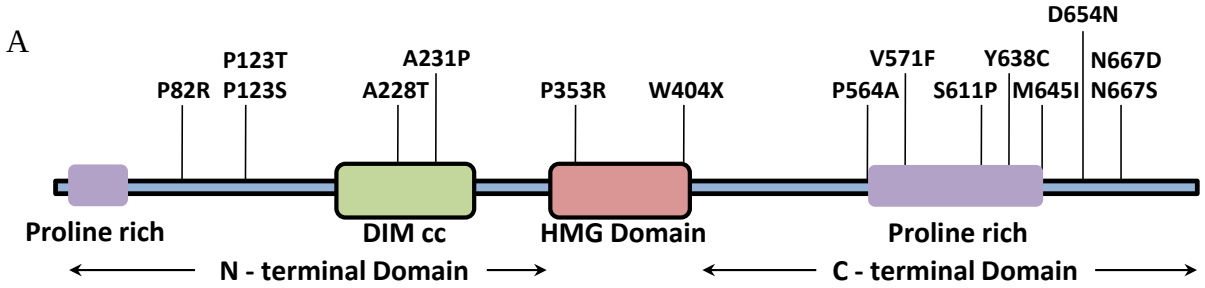
3.4.2 Subcellular localization of **SOX30** *wildtype and variant proteins*

Endogenous SOX30 was detected using the SOX30 N-terminal specific antibody in HEK293 cells, where it was found in the cytoplasm at very low intensities (Figure 1C). This cytoplasmic staining could also be a result of background noise since a high concentration of antibody was used for the staining as recommended by the manufacturer. Lower concentration of antibody reduced the signal intensity. A similar observation was also made in cells transiently transfected with an empty vector. The no-primary antibody control experiment in endogenous and SOX30 overexpressing cells did not detect any signal (Figure 1C). HEK293 cells transiently transfected with wildtype SOX30 cDNA encoding vector led to localization of SOX30 protein exclusively to the nucleus with homogenous distribution except in the nucleolus (Figure 1B, 2A). Nuclear localization was also recapitulated in cells transfected with FLAG-tagged SOX30 wildtype protein, where anti-FLAG antibody was used. Mitotic stages observed indicated that SOX30 was present all over the cell and did not bind to the condensed chromatin (Figure S3.2A).

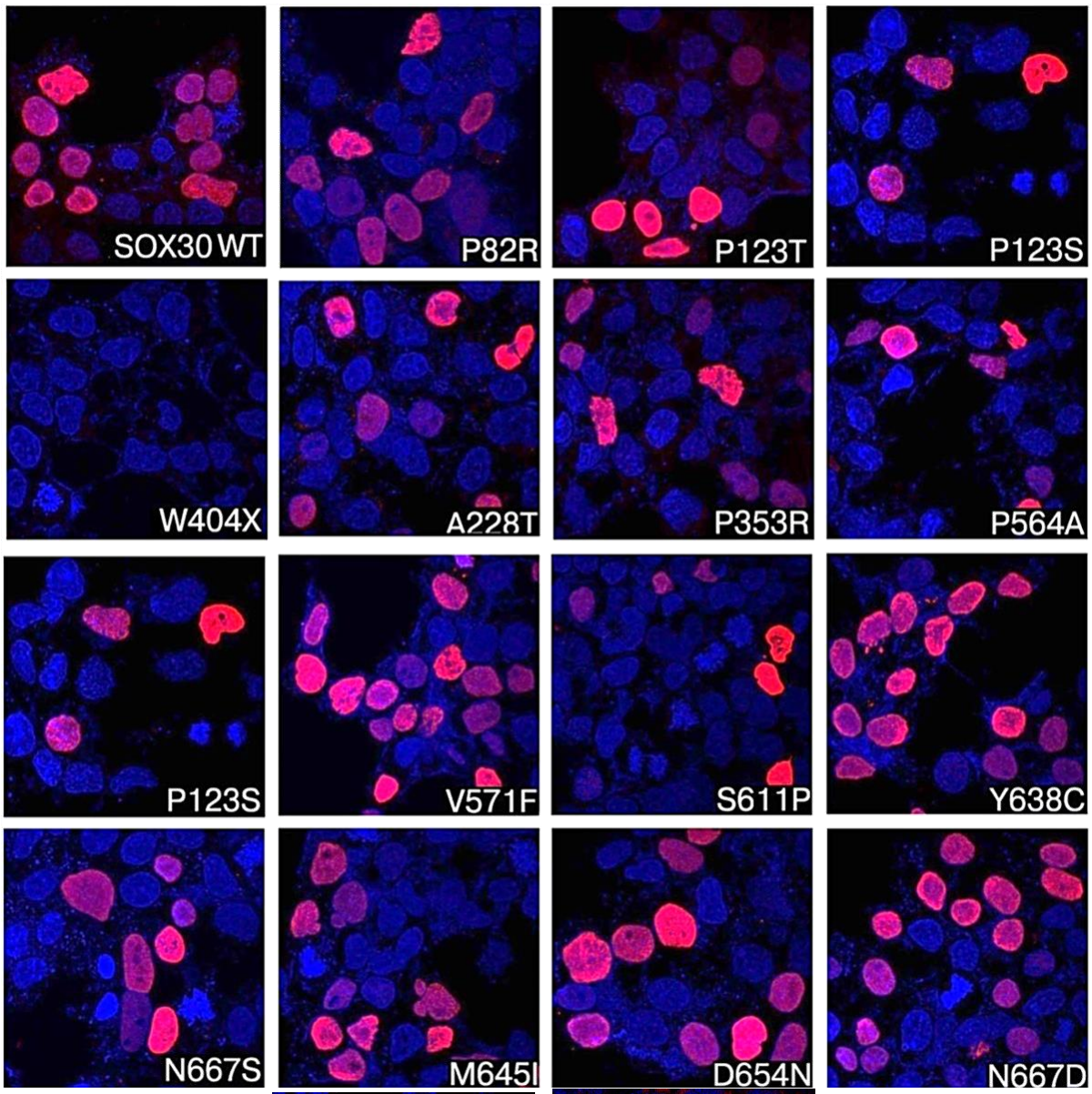
Fourteen missense SOX30 variants also displayed similar nuclear localization like the one observed for wildtype SOX30. No cytoplasmic signal was observed. Nuclear localization signal intensity varied to a small extent across the variants. These variants also exhibited staining at the periphery of the nucleus, which was also observed in wildtype protein perhaps reflecting reduced antibody penetration into the nucleus (Figure 1B). mCherry tagged SOX30 protein did not show intense peripheral staining (Figure S3.2B).

Anti SOX30 antibody did not pick up the nonsense p.Trp404Ter variant, which could be due to the degradation of the partial protein. Hence, this variation was studied in an N-terminal FLAG tag background using an anti-FLAG antibody. FLAG-tagged variant p.Trp404Ter exhibited both nuclear and cytoplasmic localization with predominant cytoplasmic expression (Figure 2B). This suggested that the protein is not degraded but does not entirely get transported to the nucleus, which could be due to the loss of nuclear localization signal. Western blot analysis of p.Trp404Ter protein from nuclear and cytoplasmic fractions of cells overexpressing the variant indicated its presence in the cytoplasm and the nucleus, while the wildtype protein was expressed only in the nucleus (Figure 2C).

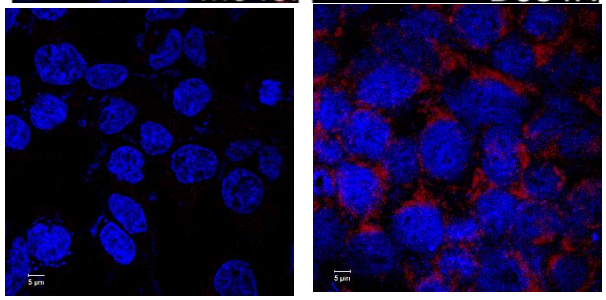
A



B



C



No primary antibody Untransfected

Figure 1: (A) Schematic of SOX30 rare coding mutations identified in JME patients, DIM cc: Homodimerization coiled-coil, HMG: High mobility group. (B) Immunofluorescence to detect subcellular localization for SOX30 wildtype and mutant proteins, performed in transiently transfected HEK293 cells. SOX30 is detected using the anti-SOX30 antibody, and the nucleus was stained using DAPI. (C) Control images representing endogenous protein in untransfected cells and no primary antibody in cells overexpressing wildtype SOX30. Merged images are represented.

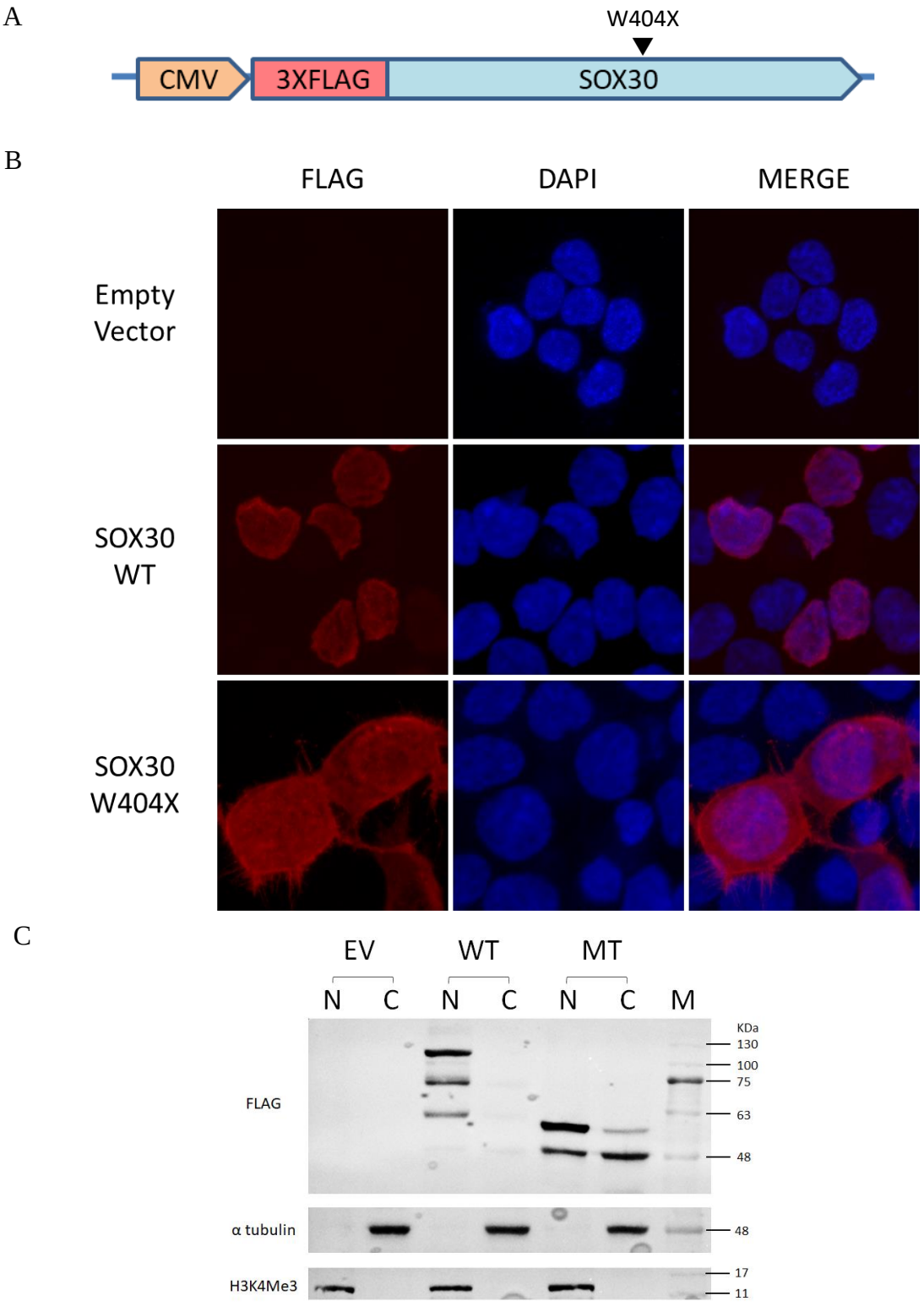


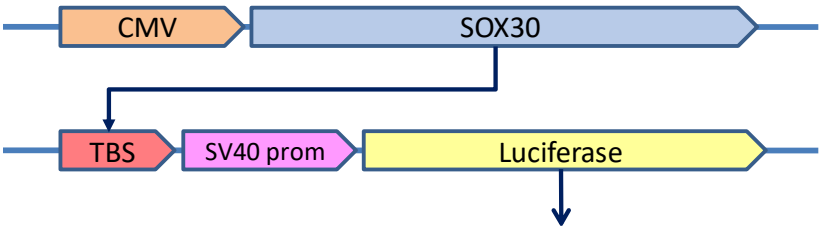
Figure 2: Subcellular localization of Flag-tagged SOX30 wildtype and W404X mutation in cultured HEK293 cells. (A) Schematic of SOX30 flag-tagged vector used for the study. (B) Using immunofluorescence, detecting monoclonal mouse anti-FLAG antibody using Alexa Fluor 568 conjugated anti mouse secondary antibody. The nucleus is stained using DAPI. (C) Western blot analysis using nuclear and cytoplasmic fraction of cells overexpressing wildtype or mutant protein. EV – Empty vector, WT – Wildtype, MT – Mutant, N- nuclear fraction, C- cytoplasmic fraction. Histone H3K4me3 was used as a nuclear and α tubulin as a cytoplasmic protein marker. M – Marker.

3.4.3 Luciferase assay

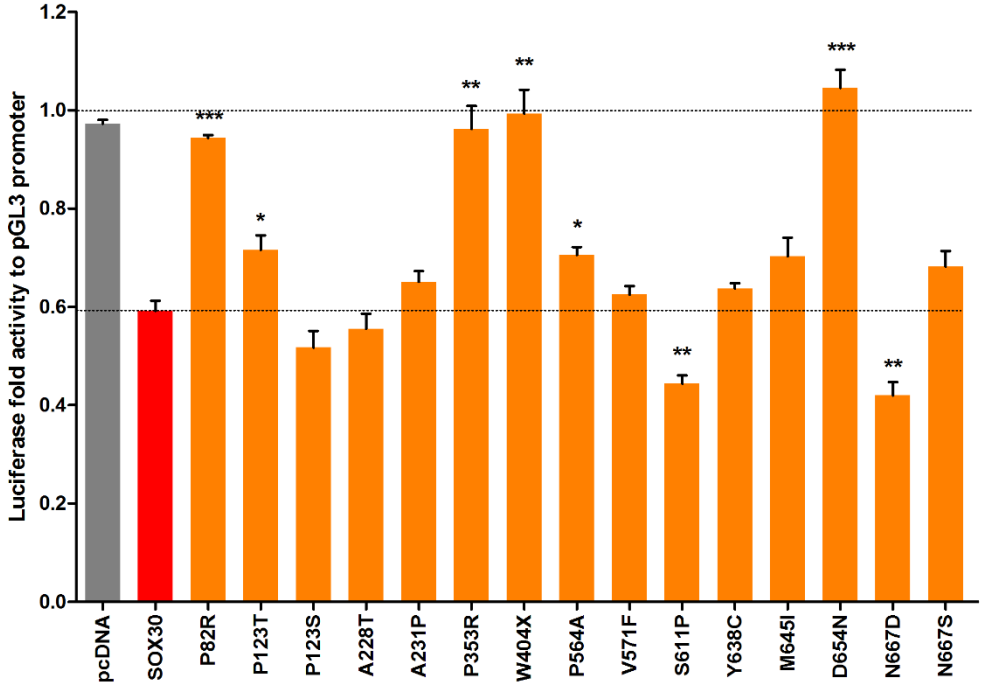
The transcriptional activity of SOX30 mutants was evaluated by a luciferase-based reporter assay. Plasmids expressing either the wildtype or a mutant SOX30 were co-transfected into cultured HEK293 and SH-SY5Y cells along with a luciferase reporter plasmid containing four copies of high-affinity SOX30 binding sites upstream of the minimal SV40 promoter. After normalizing the luciferase values for transfection efficiency differences using β galactosidase activity, values were compared.

Compared to the empty vector control, wildtype SOX30 displayed significant repressed reporter gene transcription. Wildtype SOX30 downregulated luciferase expression by about 40% in HEK293 cells and about 48% in SH-SY5Y cells (Figure 3 B, C). The mutants p.Pro82Arg, p.Pro353Arg, and p.Trp404Ter were found to have lost their ability to drive transcription and behaved like an empty vector in both the cell lines. Statistical analysis indicated mutant p.Pro123Thr also exhibited significant loss of function, where transcriptional repression of 29% in both the cell lines was observed, which is about 11% less than wildtype but not to the extent of exhibited by the empty vector (Figure 3B, C). Mutants p.Asp654Asn in HEK293 cells and p.Pro564Ala, p.Val571Phe, p.Ser611Pro and p.Asn667Ser in SH-SY5Y cells were also found to behave like empty vector indicating loss of trans-repression activity. Variants p.Ser611Pro and p.Asn667Asp in HEK293 cells exhibited enhanced luciferase repression when compared to wildtype, which was found to be significant. They demonstrated nearly 58% repression of luciferase transcription, which is about 18% greater than wildtype transcriptional activity (Figure 3B). This indicated that these two were gain-of-function variants. However, this was not observed in SH-SY5Y cell lines. The rest of the mutants did not indicate any loss- or gain-of-function and were not significantly different from wildtype. The differences in the behaviour of the variants between the cell lines could be due to an inherent difference in regulation and expression of proteins because of cell line specific differences probably as a result of their lineage (Figure 3 B, C).

A


CMV
SOX30
TBS
SV40 prom
Luciferase

B


Luciferase fold activity to pGL3 promoter
1.2
1.0
0.8
0.6
0.4
0.2
0.0
pcDNA
SOX30
P82R
P123T
P123S
A228T
A231P
P353R
W404X
P564A
V571F
S611P
Y638C
M645I
D654N
N667D
N667S

C

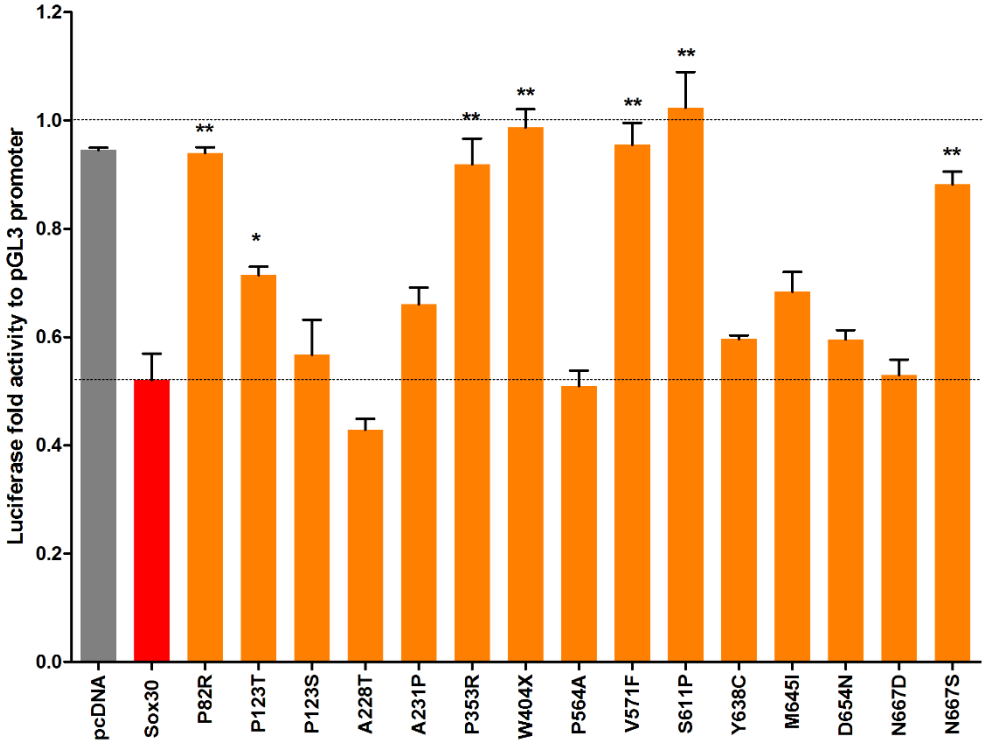

Luciferase fold activity to pGL3 promoter
1.2
1.0
0.8
0.6
0.4
0.2
0.0
pcDNA
Sox30
P82R
P123T
P123S
A228T
A231P
P353R
W404X
P564A
V571F
S611P
Y638C
M645I
D654N
N667D
N667S

Figure 3: (A) Schematic of the SOX30 protein expressing vector and reporter luciferase vector and their interaction. (B) Reporter luciferase analysis in HEK293, of wildtype and mutant SOX30 variants. (C) Reporter luciferase analysis in SH-SY5Y, of wildtype and mutant SOX30 variants. Values are represented as a ratio of normalized luciferase values obtained in presence to the absence of the SOX30 binding site. Values are represented as mean of triplicates, and the error bars represent standard error of the mean. Student's t-test was performed, comparing each mutant with the wildtype, *$p<0.05$,**$p<0.01$, and ***$p<0.001$.

3.4.4 The ChIP-Seq expression vector design and antibody validation

Commercially available, anti-SOX30 antibodies available were not validated for ChIP-Seq experiments (Feng et al. 2017). These were therefore, not selected for this study. SH-SY5Y cell lines selected to conduct ChIP-Seq did not express SOX30 endogenously; hence, a vector encoding N terminal FLAG-tagged SOX30 protein was expressed in these cells and anti-FLAG antibody that targets the N-terminal tag was selected for ChIP-Seq. According to ENCODE guidelines, antibody validation was performed to check for its specificity, reactivity, and degree of enrichment. These were analyzed by immunofluorescence, immunoblotting, and immunoprecipitation experiments by FLAG antibody on cells or cell lysates expressing FLAG-tagged SOX30. The antibody correctly recognized FLAG-tagged SOX30 in cultured cells and nothing in empty vector transfected cells (Figure 4B). Antibody specificity was also checked by western analysis of untransfected and FLAG-tagged SOX30 expressing cells where anti-FLAG antibody picked up a specific band at ~100kDa in for the tagged-SOX30. No band was present at the expected size of the protein in the empty vector lane (Figure 4C). The antibody also effectively immunoprecipitated the protein from the FLAG-tagged SOX30 expressing human cell lysates. These experiments validated both the expression vector and antibody for use in ChIP-Seq (Figure 4D).

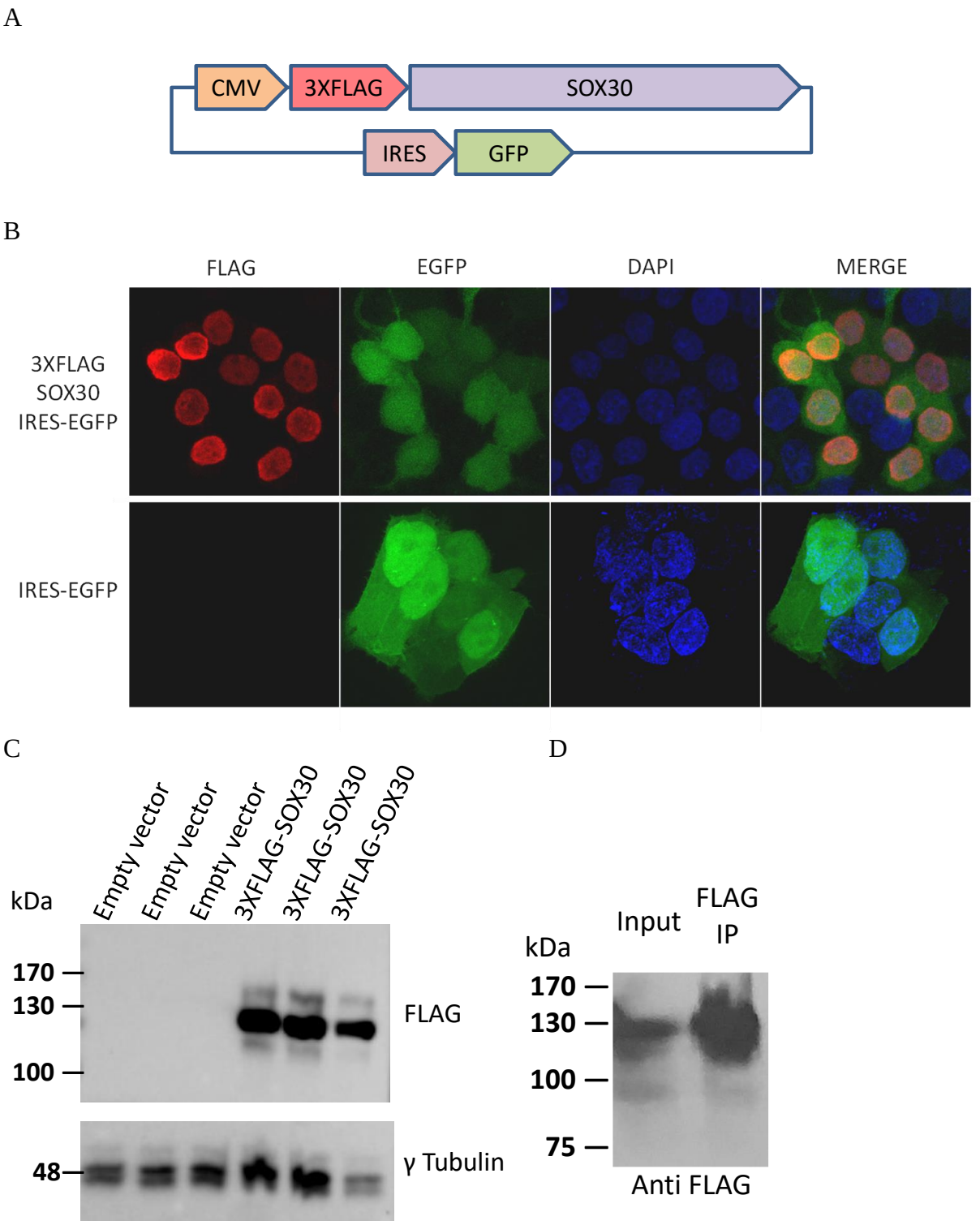


Figure 4. (A) Schematic of vector used to express FLAG-tagged SOX30. **(B)** FLAG antibody validation by immunocytochemistry of HEK293 cells overexpressing N- terminal FLAG-tagged SOX30. Empty vector was used as a negative control. **(C)** Western blot analysis of HEK293 cells overexpressing N-terminal FLAG-tagged SOX30. Empty vector was used as a negative control. γ tubulin was used as a loading control. **(D)** Immunoprecipitation validation of the FLAG antibody in HEK293 cells overexpressing FLAG-tagged SOX30.

3.4.5 SOX30 chromatin immunoprecipitation and massive parallel sequencing analysis

ChIP-sequencing (2x150 paired-end) of three independent replicates of ChIP - input pairs were carried out on HiSeq2500. On an average, each sample produced 3,30,99,274 sequence reads with 99% of the reads having phred score greater than 20. The percentage of reads with phred score which greater than 30 was 93.28%. Post alignment, the average number of reads across samples were 31859293. On average, 92.24% of the sequences were mapped to the human genome (Table 4). 64120 ChIP enriched regions were determined through peaks, obtained by MACS2 analysis. Post peak quality filtration, the average number of peaks across the replicates was 4232. About 60% of chromatin-bound SOX30 was localized over protein-coding gene sequences within 3 kb upstream of the TSS (transcription start site) and 1 kb downstream of the TTS (transcription termination site). Each replicate analyzed for distribution of peaks throughout the genome revealed that promoter regions up to 3 kb were occupied by 17%, 19%, and 4.5% of the peaks (Figure 5). Among these, 11%, 15% and 1.23% of the peaks were present within 1kb upstream of the transcriptional start site. Closer inspection of the peaks located near transcription start sites (TSSs) revealed that they were concentrated in the 5′ untranslated regions (5′ UTRs) and upstream of the TSSs except in the third replicate (Figure 5). No clear-cut enrichment in the third replicate E was observed wherein no substantial difference in amplitude between the input and ChIP peaks was present. This indicated that the third replicate had not performed optimally. No peaks overlapped completely across all three replicates with a cutoff q-value <0.05, while peaks that overlapped between replicate C and D were 53; D and E were 2; and C and E were 2 (Table S3.6). With a reduced threshold setting, common peaks across all three samples were 229 in number. Due to the poor reproducibility rate, the data could not be confidently considered to represent the binding sites.

Table 4: Raw reads and alignment statistics for the ChIP-seq data set.

	Raw sequences				**Aligned sequences**			
Sample	**Number of reads**	**GC %**	**% Bases >Q20**	**% Bases >Q30**	**Total reads**	**Mean read length**	**GC %**	**% mapped to the genome**
Input-C	3,13,94,752	44	99.4	92.74	2,99,89,896	142.2	43	89.61
ChIP-C	3,51,42,598	45.5	99.32	91.72	3,33,56,980	145.38	45	92.18
Input-D	2,75,76,484	43	99.49	94.34	2,67,35,504	141.82	43	88.68
ChIP-D	3,98,39,766	45.5	99.17	92.02	3,79,33,754	146.29	45	92.72
Input-E	2,97,87,132	43	99.75	95.63	2,92,11,650	137.45	43	94.26
ChIP-E	3,48,54,914	46	99.66	94.43	3,39,27,974	144.54	46	95.99

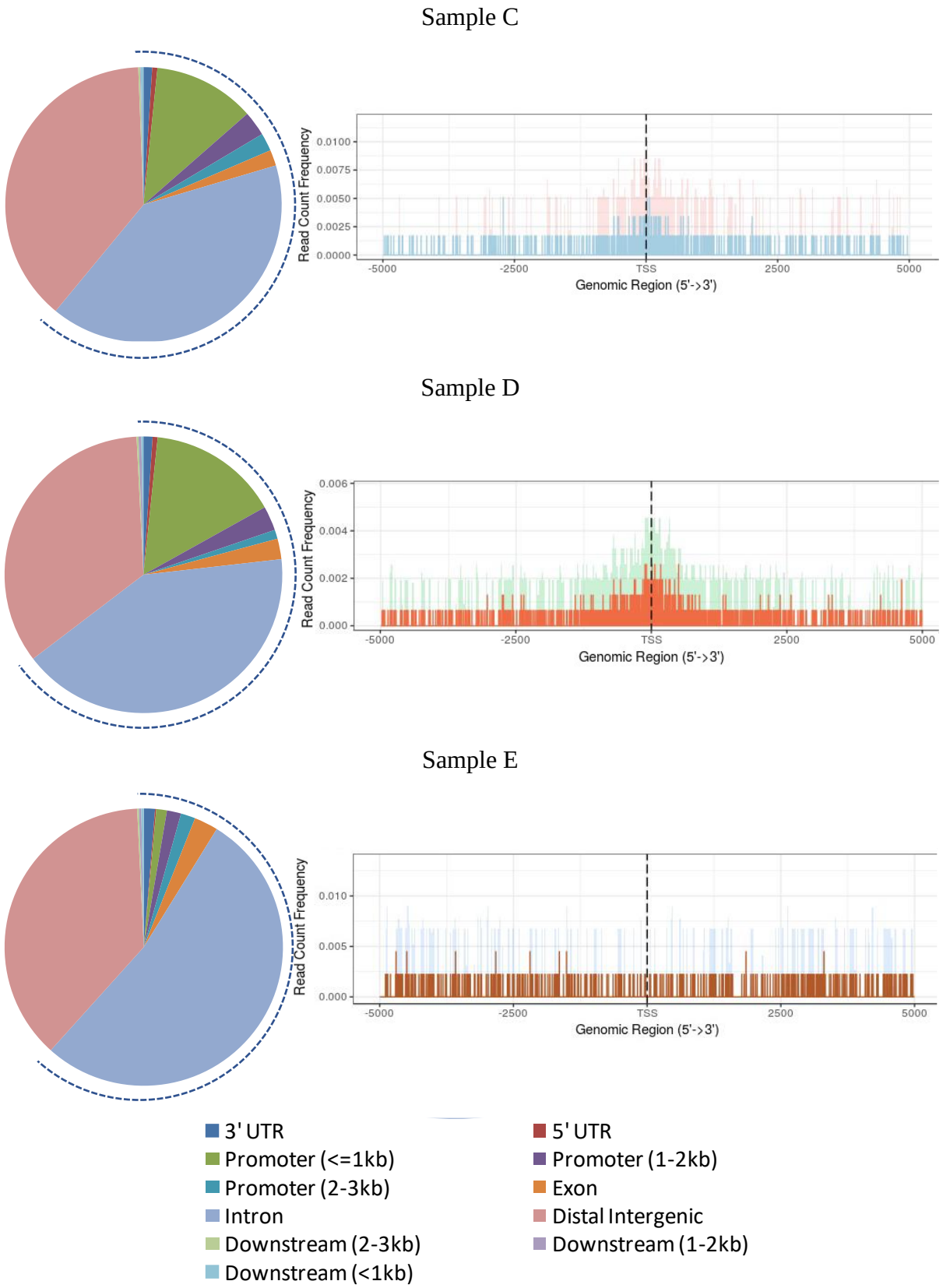


Figure 5. *Left panel*: Pie chart showing the distribution of SOX30 ChIP-seq peaks over various genomic regions. A dashed curved line indicates the region from 3 kb upstream of TSS to 1 kb downstream of TTS (or polyadenylation site). *Right panel*: Average profile of peak reads found ±5000 bp around TSSs. The confidence interval was estimated by the bootstrap method(resample=1000). The darker color is

input, and the lighter color is the ChIP ped sample. (Acknowledgement: Clevergene Pvt. Ltd, Bengaluru)

3.4.6 Validation of SOX30 occupancy on genomic regions

To validate SOX30 occupancy at some of these sites, peaks were filtered for the following criteria for each replicate: exhibiting fold enrichment beyond 5, is located in the promoter or 5' UTR or early exonic regions and is present in both the replicates. The highest-ranking peak of the first two replicates was in the first exon of *SOX30* indicative of probable autoregulation. *ZSCAN9*, *STX16*, *BRWD1* and *BCL2L1* were selected for validation by ChIP-qPCR with IgG as the control. Independent anti-SOX30 ChIP experiments confirmed SOX30 recruitment, with a 5- to 25-fold enrichment relative to the control IgG ChIP-ed samples represented as a percentage of input control (Figure 6). As an additional negative control, primers recognizing genomic regions in the promoter of MyoD that is regulated only in muscle lineage cells were used for ChIP-qPCR. No peaks were identified in the MyoD promoter region in the ChIP-Seq datasets, and no enrichment was observed in ChIP sample over IgG sample in ChIP-qPCR validation, thus confirming the lack of SOX30 recruitment at non-target sites and the specificity of the anti-SOX30 ChIP (Figure 6). The amount of enrichment for these targets also was reflected in the peak scores obtained from ChIP-Seq datasets except for *BCL2L1*, which, although it had exhibited high enrichment, ranked the lowest.

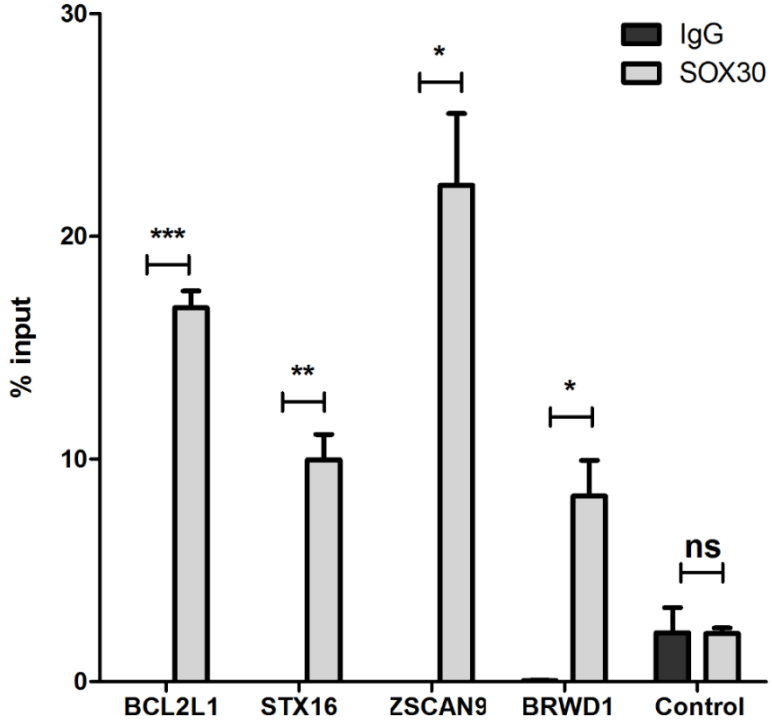


Figure 6: ChIP with anti-IgG and anti-FLAG antibody on SH-SY5Y cells expressing FLAG-tagged SOX30. Enrichment was determined with qPCR and for each locus normalized against the input.

Control is the MyoD promoter region. Error bars are SEM of two biological replicates. Asterisks indicate significant difference from IgG ChIP (P < 0.05=*, <0.01=**, <0.005=***, Student's t-test).

To identify whether the association of SOX30 at the promoter of these genes led to regulation of its expression, cells overexpressing wildtype or nonsense mutant p.Trp404Ter were analyzed for transcript expression of these target genes in three independent replicates. In the presence of wildtype SOX30, genes *BCL2L1, STX16* and *BRWD1* were repressed while in the presence of mutant SOX30, these were upregulated (Figure 7). ZSCAN9 transcript expression was enhanced in the presence of wildtype SOX30, while in the presence of mutant SOX30, the expression was downregulated (Figure 7). The repression or activation of gene expression was with reference to cells not expressing the SOX30 protein. Statistical analysis by student's t-test did not indicate any significant difference between the wildtype and mutant SOX30 expressing cells of target genes, while p values for all samples laid between 0.4-0.09.

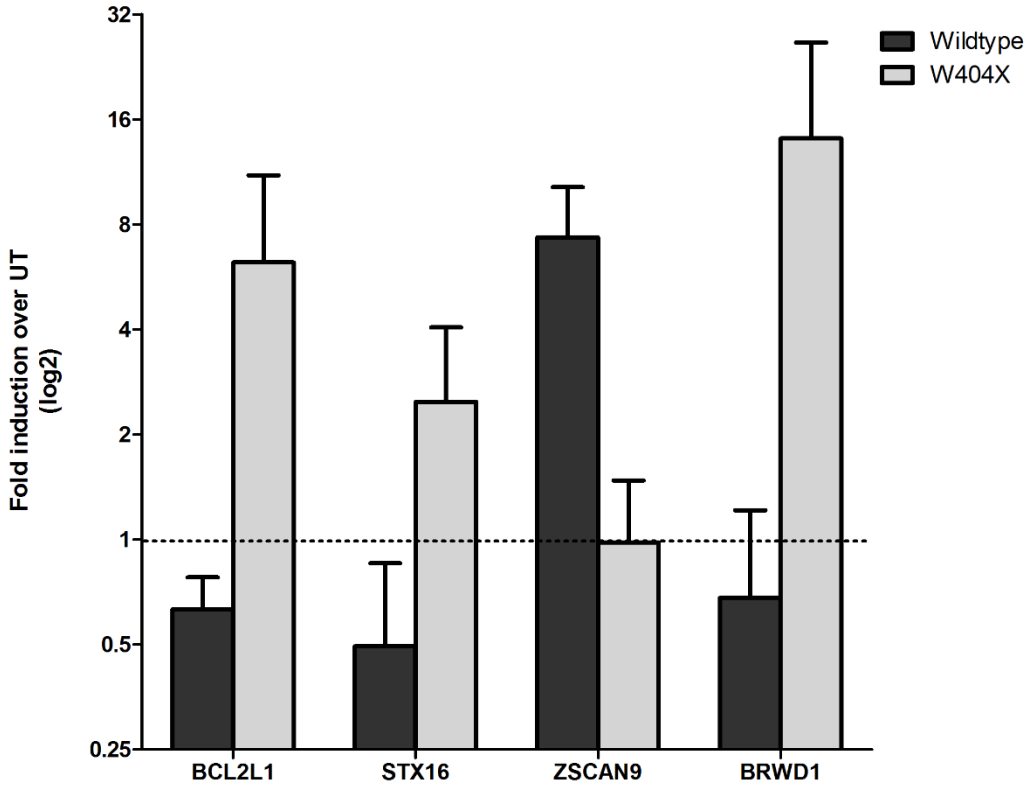


Figure 7: Overexpression of SOX30 wildtype or p.W404X mutant modulates gene expression in SH-SY5Ycells. Quantitative RT-PCR analysis of the relative levels of mRNA transcripts of a few genes targeted by SOX30. The values were normalized with GAPDH expression levels. Expression values were calculated applying the $-2\Delta\Delta CT$ algorithm. Data are expressed as the mean ± SEM of each group's level of mRNA transcripts. Data are representative of three independent experiments. Unpaired student t test was used to determine statistical significance.

3.5 Discussion

The SOX family of proteins currently consists of 19 members that have originated through multiple rounds of genomic duplications and divergence. They share an evolutionarily conserved high mobility group (HMG) box, DNA-binding domain of about 50% identity to SRY's HMG domain (Bowles et al. 2000, Stros et al. 2007). SRY was the first member of the family and is involved in primary male sex determination (Gubbay et al. 1990, Sinclair et al. 1990). HMG box of SOX proteins bind to a consensus sequence (A/T)(A/T)CAA(A/T)G but with varying degrees of efficiency. They also contain two nuclear localization signals (NLS) at N-terminal and C-terminal of HMG box and one nuclear export signal (NES) that control its nuclear-cytoplasmic shuttling (Malki et al. 2010, Wegner et al. 2010). SOX proteins are involved in regulatory processes from embryonic to adult stages of development in humans, mice, and other vertebrates and invertebrates (Kamachi et al. 2000, Phochanukul and Russell 2010, Kamachi and Kondoh 2013, Sarkar and Hochedlinger 2013). They regulate gene expression by either trans-activating or trans-repressing their targets through their interactions with other transcription factors and cofactors (Wegner 2010). Based on phylogeny, protein structure, and function, the SOX proteins are divided into nine subgroups, labelled A-J, in humans (Bowles et al. 2000, Schepers et al. 2002). Mutations in SOX proteins have been identified in several congenital disabilities, usually due to imperfect developmental processes. These are now collectively named as SOXopathies (Angelozzi and Lefebvre 2019). SOX30 is the most recently identified member of the SOX family. It was found as a WT1 (Wilm's tumor suppressor gene) -associated protein in mouse testis using yeast two-hybrid system, and further screening of human testis cDNA (Osaki et al. 1999). The SOX30 protein domain structural organization is similar to SOXD subgroup proteins consisting of SOX5, SOX6 and SOX13, which contain HMG domain and homodimerizing coiled-coil domains (Heenan et al. 2016, Angelozzi and Lefebvre 2019). Phylogenetic analysis using the neighbour-joining tree method indicated SOX30 to share the closest identity to SRY (Bowles et al. 2000, Schepers et al. 2002, She and Yang 2015). Functionally SOX30 is like SRY, SOX3, SOX8, and SOX9 – known to play key roles in the development of the reproductive system. Mutations in these genes have been identified in patients with disorders of sexual development and intellectual disability (Heenan et al. 2016, Angelozzi and Lefebvre 2019). *SOX30* codes for two major isoforms, the 753 residue-long is the canonical isoform containing the C terminal domain required for transactivation and the other is a 501-residue long isoform lacking the transactivating domain (Osaki et al. 1999). Sox30 transcripts in mice are expressed in the brain, lung, heart, stomach, pancreas,

adrenal glands, kidney, ovary and testis at the embryonic stage 13.5 dpc (Osaki et al. 1999, Lioubinski et al. 2003, Petit et al. 2015, Feng et al. 2017). In adult mice, Sox30 is expressed highly in testes and in lower quantities in muscle and lungs (Osaki et al. 1999, Han et al. 2014, Feng et al. 2017). Sox30 protein is synthesized only in testicular tissues, which increases during postnatal development from P1 to P21 and after that the expression remains apparently the same. In contrast, very low levels of expression have been detected in the brain and uterus at 30 dpc (Han et al. 2014, Petit et al. 2015, Bai et al. 2018). Among the germ cells, Sox30 is expressed in pachytene spermatocytes, round spermatids, and Leydig cells (Bai et al. 2018, Roumaud et al. 2018). In humans, SOX30 transcripts are detected in various regions of the adult brain, such as hippocampus, hypothalamus, cerebral cortex, and cerebellum. SOX30 protein was found to be present in the hippocampus, frontal, parietal, temporal, and occipital lobes of the cortex, basal ganglion, and amygdala in the adult brain (Ratnapriya 2009). Sox30 null male mice are sterile. The testis size of these mice is smaller when compared to that of the wildtype. Histological analysis has indicated the arrest of meiosis at the round spermatid stage with large multinucleated cells and disorganized tissue morphology. Female Sox30 null mice are fertile. Heterozygous mice, both males and females are also normal with no apparent phenotypic and behavioural abnormality (Feng et al. 2017, Bai et al. 2018, Zhang et al. 2018, Han et al. 2020).

Here, I have presented *in silico* and functional characterization of previously unreported *SOX30* variations identified among JME patients. Fourteen missense variants and one nonsense variant were identified in JME patients which were either rare or absent in an ethnically matched control population. The Residual Variation Intolerance RVIS score for SOX30 is 1.29 (93.85%) indicating that it is a tolerant gene, hence, common functional variants are likely to be present in this gene. On the contrary, the probability of being loss-of-function Intolerant (pLI) score for SOX30 is 1.0 indicating that the gene is unlikely to tolerate loss of function variants. Haploinsufficiency Score (HI index) for SOX30 is 45.16 indicating that the gene has about 50% probability of exhibiting haploinsufficiency (Petrovski et al. 2013, Lek et al. 2016). The subRVIS for SOX30, based on protein domain regions indicated that p.Pro353Arg and p.Trp404Ter, located in the HMG box, have percentiles less than 35% (Figure S3.4). The subRVIS based on exonic regions indicated variations located in exons 1 and 2 have exon subRVIS percentiles less than 35%, which include p.Pro82Arg, p.Pro123Thr, p.Pro123Ser, p.Ala228Thr, p.Ala231Pro, and p.Pro353Arg suggesting that these variations are probably disease-causing (Figure S3.5) (Gussow et al. 2016). Metadome server analysis of variants

revealed that p.Pro353Arg and p.Ser611Pro were intolerant to changes while the rest was slightly intolerant to missense variations (Wiel et al. 2019). Although the RVIS score supports our finding of several missense variants in a relatively small JME cohort, other intolerance scores indicated that a number of these variants could have pathogenic effects. Bioinformatic analysis of these variants indicated varying predictions from benign to highly pathogenic. Proline is the most frequently mutated residue with five variants changing proline to polar to less-polar amino acids only. No variations were identified at residues that were positively charged and in special residues cysteine and glycine. Although the variants were scattered across the protein sequence, a closer look reveals that 9 of 15 of them are concentrated in structural motifs (Figure 1A). Protein mis-localization has been the cause of several neurological disorders due to mutations in the NLS sequence and in residues that undergo posttranslational modifications (Hung and Link 2011). Experimentally validated posttranslational modifications from reported literature for SOX30 lists phosphorylation of Serines at positions 203, 607, 676, 682, and 685 (Hornbeck et al. 2012). The effect of variants identified on subcellular localization was determined using immunocytochemistry with an anti-SOX30 antibody whose epitope lies in the HMG domain. None of the missense variants exhibited any apparent defect in their nuclear localization. The nonsense mutant was not detected by the antibody possibly due to possible incomplete access to its epitope because of truncation of the protein or due to absence of protein owing to nonsense mediated decay (NMD) of the nonsense variant containing transcript (Figure 1B). Although the p.Pro353Arg variant is located in a conserved calmodulin-binding bipartite N-terminal NLS that has overlapping motif for nuclear export using exportin 4 interaction, no localization defects were observed (Malki et al. 2010, Angelozzi and Lefebvre 2019). Since p.Trp404Ter was not detected by the anti-SOX30 antibody, a FLAG -tagged vector was used to analyze subcellular localization using an anti-FLAG antibody. It exhibited abnormal cytoplasmic localization in additional to nuclear localization that was also recapitulated by western blotting of cellular sub-fractions (Figure 2B, C). This partial mislocalization to the cytoplasm is probably due to loss of the C-terminal NLS sequence that performs nuclear translocation using the importin β interaction pathway. This also indicated that the transcript did not undergo nonsense-mediated decay perhaps because this mutant resembles the SOX30 isoform b, that contains 501 amino acid residues.

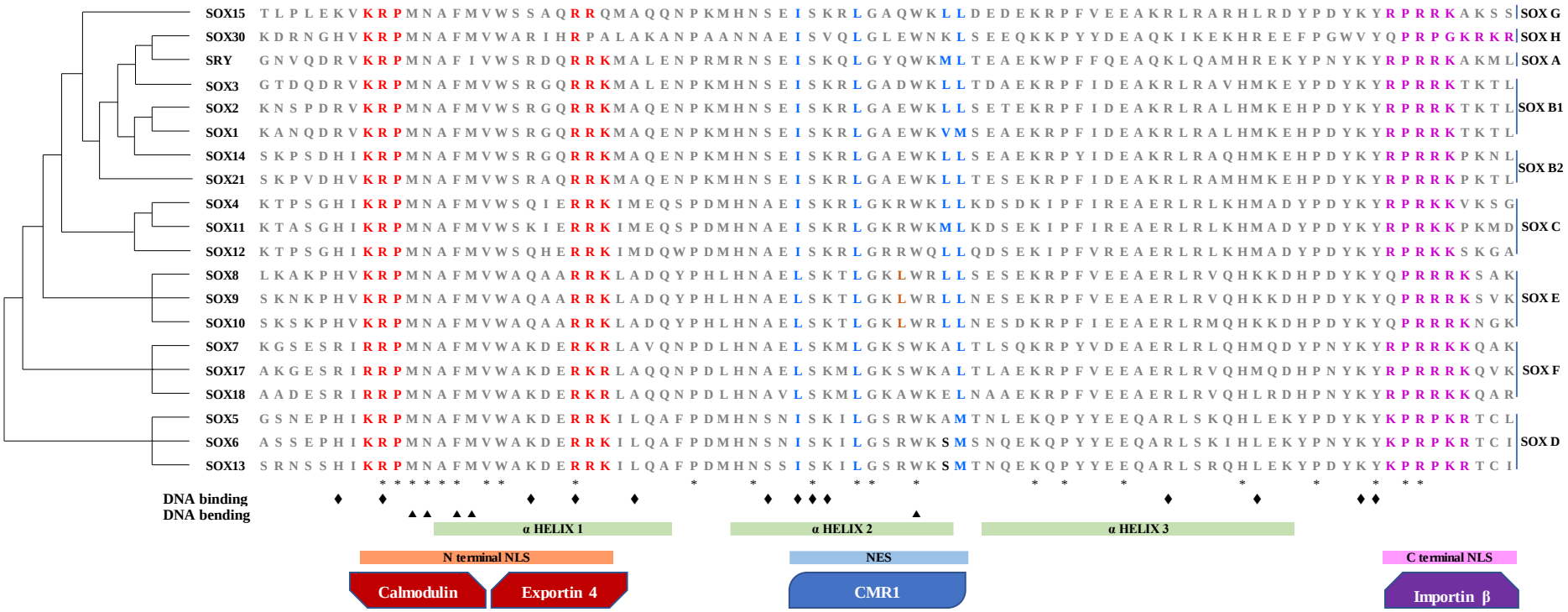


Figure 4. Alignment of HMG domain amino acid residues of the SOX family members. ClustalOmega was used to align the sequence and for visualization. The phylogenetic tree was developed using phylogeny.fr using ClustalW for alignment, GBlocks for curation, Bayesian inference of phylogeny model was developed using MrBayes (right). SOX subgroups are labelled on the left of the aligned sequence. * below the residues at the bottom of the alignment indicates a 100% consensus of amino acids. Residues involved in DNA binding are marked by ♦, in DNA bending are marked by ▲. Green bar represents alpha helices that form an L shape to bind to the minor groove of DNA, resulting in bending of the nucleic acid. N terminal NLS residues are marked in red, blue residues code for Nuclear export signal (NES), pink residues represent C terminal NLS. Dark red residues lying in NES represent additional residues required for nuclear export in SOX subgroup E. Factor involved in nuclear export and import are labelled below the bars representing their respective motifs (Malki et al. 2010, Angelozzi and Lefebvre 2019)

To assess the ability of the *SOX30* variants to regulate transcription, reporter luciferase assays were conducted using the SOX30 binding site that contains four copies of the 'ACAAT' motif (Osaki et al. 1999) located upstream to SV40 weak promoter in the cultured HEK293 and SH-SY5Y cells. HEK293 was selected for this study since they are easy to culture, manipulate and express large amounts of functional recombinant proteins (Ooi et al. 2016). SH-SY5Y were chosen since they are frequently used model cells to study neurological conditions. They are of human origin, have neuronal properties, including expression of tyrosine hydroxylase and dopamine-beta-hydroxylase, as well as the dopamine transporter and can be differentiated into functional neuronal cells using chemical agents (Xie et al. 2010, Xicoy et al. 2017). In our study, wildtype SOX30 repressed the expression of luciferase which was contrary to the reported findings where SOX30 has been found to predominantly activate its downstream gene targets such as tumor suppressor *TP53* and desmosomal genes *DSP*, *JUP* and *DSC3* in lung cancer cell lines, spermatogenesis genes such as *Ccdc54* and *Spata19*, haploid cell development-related genes such as *Tnp1*, *Hils1*, and *Tsksm* in mice testicular tissues (Osaki et al. 1999, Han et al. 2015, Bai et al. 2018, Hao et al. 2018, Zhang et al. 2018). SOX30 has been reported to repress β catenin expression in addition to interacting with it, leading to inhibition of the Wnt signaling pathway (Han et al. 2018a, Han et al. 2018b). This is in line with the findings of other SOX family members that are mostly involved in repression of the Wnt signaling pathway (Kormish, Sinner, and Zorn 2010). SOX30 was also found to autoregulate its expression via a positive feedback loop (Bai et al. 2018, Zhang et al. 2018).

As expected, p.Trp404Ter exhibited the loss of repression activity in the assay employed. since the localization assay indicated partial cytoplasmic retention of the protein (Figure 3B, C). N-terminal variant p.Pro82Arg and HMG domain variant p.Pro353Arg exhibited loss of repressible activity comparable to empty vector in both the cell lines. p.Pro564Ala in HEK293 and p.Val571Phe and p.Asn667Ser in SH-SY5Y also exhibited complete loss of transregulation. p.Pro123Thr exhibited a partial loss of transrepression in both the cell lines. p.Ser611Pro and p.Asn667Asp and p.Ser611Pro exhibited significant repressive function in HEK293 while p.Ser611Pro had complete loss of function in SH-SY5Y (Figure 3B, C). The rest of the variants showed similar activity like the wildtype protein; hence, these could be incidental variants or low penetrant variants which are not enough to cause the disorder on their own but, in combination with another variant, may increase the susceptibility. p.Pro353Arg is located in the HMG domain and also in the N terminal NLS motif, a highly conserved region, hence this mutation could have led to a disturbance in its ability to drive transcription by

affecting the binding of the protein to the target DNA sequence (Figure 1A, 4). Variants located in C- terminal proline-rich regions that included p.Pro564Ala, p.Val571Phe, and p.Ser611Pro exhibiting loss of function, could be due to the residues being important for interaction with other cofactors (Figure 1A, 4). No consistent feature was observed among the remaining variants that exhibited loss or gain of transcriptional activity. Multiple sequence alignments using ClustalW and MAFFT of human SOX proteins revealed that none of the *SOX30* variants present in the corresponding residue of its orthologs were reported to have a pathogenic variation (Figure 4). Subtle differences may not have been picked up due to the redundancy of SOX molecules, which in the context of SOX30 is not known yet. *SOX30* variants have not been reported to be associated with any human disorders. Hence, functional characterization of variants was important to gain initial evidence of their functionality. Other experiments such as interaction studies by co-immunoprecipitation and efficiency of binding to DNA by EMSA would shed more light on the mechanism by which these variants affect the protein function. This study provides evidence that certain rare *SOX30* variants identified among the JME patients, demonstrate altered localization or transcriptional regulatory activity.

To understand the mechanism by which SOX30 regulates expression of other genes, RNA-seq and ChIP-Seq studies have been reported. Knockout of *SOX30* led to downregulation of several genes involved in spermatid development, differentiation, and spermiogenesis -related genes in testes of P21 mice. A few upregulated genes belong to the class of proteins highly expressed in either spermatogonial stem cells or round spermatids. Several of these dysregulated genes had their promoters bound to by SOX30 (Bai et al. 2018). Round spermatids in Sox30 null mice also revealed downregulation of testis-specific genes. These genes belonged to various protein classes such as enzymes, DNA- binding proteins, structural and transmembrane proteins. Key spermatogenesis genes, *Crem, Rfx2, Trf2, Ddx25, Boule, Miwi,* and *Tpap* were found to be independent of Sox30 regulation (Zhang et al. 2018). The absence of *Sox30* in mice analyzed for postnatal developing stages revealed that at 21 dpp, abnormal germ cell begins to be formed. At 23dpp, several spermatogenesis related genes, including *Fhls, Tnp1, Tnp2, Prm1, Prm2, Prm3, Odf1, Castsper1,* and *Smcp*, begin to get downregulated (Feng et al. 2017).

Publicly available databases indicate that Sox30 is not highly expressed in murine brain. Allen brain atlas ISH datasets present its transcript expression in various brain subregions with highest being in cortex, olfactory areas and cerebellum in adult mouse brain. The expression is unreported in developing mouse brain at both pre- and postnatal stages except at P28 stage.

At P28, highest expression was detected in the telencephalonic vesicle and, to lesser extent in the hindbrain. The HPA Mouse brain RNA-Seq and BrainStars datasets indicate highest Sox30 transcript expression in retina with lower expression in cerebral cortex, hippocampal formation, amygdala, basal ganglia, midbrain, corpus callosum and pituitary gland. Spatio-temporal mouse brain expression by BrainTx states constant cerebellar expression from E18-P56 and hippocampal expression at P7 or P21. SOX30 mRNA has also been reported to be present in small quantities in the lungs in adult mice (Han et al. 2014).

In patients diagnosed with lung adenocarsinoma, increased SOX30 expression in lung tissues was corelated with better chances of survival of the patient (Han et al. 2015b). In patients with reduced SOX30 expression was to be due to methylation of *SOX30* leading to suppression of its expression (Han et al. 2015a). SOX30 is also a prognostic marker for advanced stage ovarian cancer patients wherein higher expression of SOX30 in ovarian cancer tissues predicted better survival of the patient (Han et al. 2019).

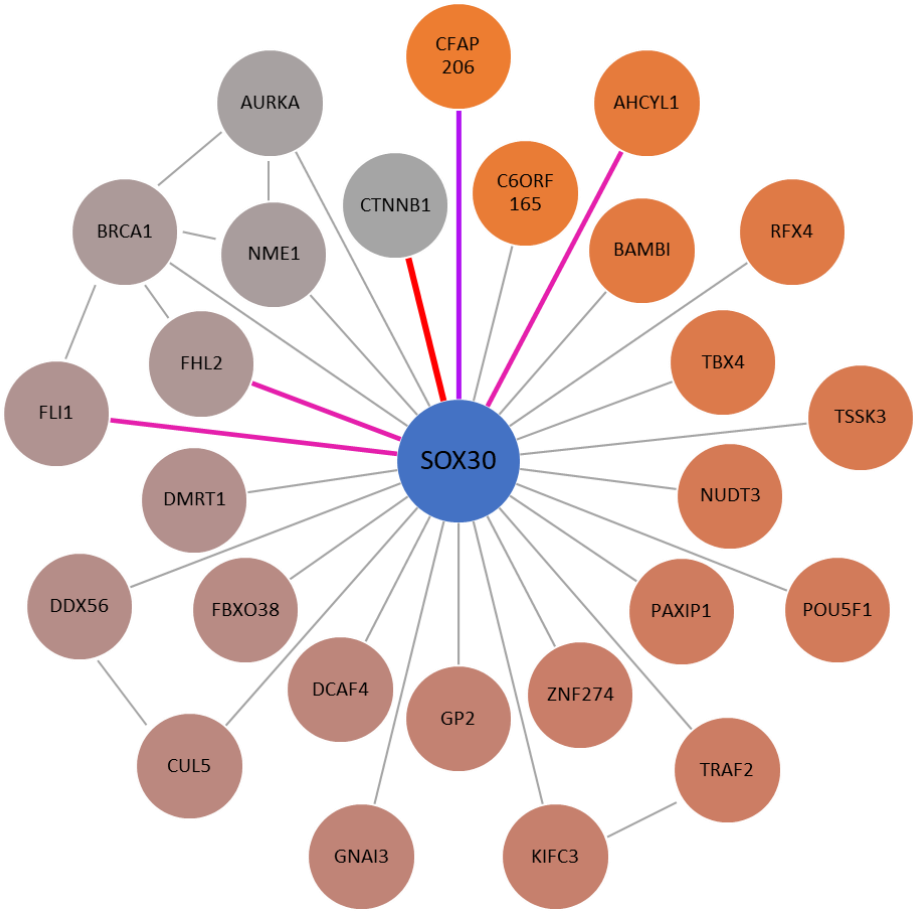


Figure 5. Protein interaction network of human SOX30 protein determined from databases. The nodes (colored circles) are proteins, (lines) represent protein-protein interaction among the nodes, (red>purple>pink>grey line) decreasing order of confidence of interaction obtained from String-db, MINT, IntAct, Biogrid databases and literature.

SOX30 interactors obtained from databases and literature were predominately determined by two-hybrid assays. Transcription factors, signal transduction proteins were overrepresented in the interactors list. Network analysis of these genes highlighted cellular and organism developmental process, cell differentiation and reproductive process clusters to be enriched. CTNNB1 physical association with SOX30 was validated via immunoprecipitation experiments. Microarray and RNA array studies in SOX30 overexpressing human lung cancer cell line A549 revealed that genes involved in cell proliferation, migration, invasion, and apoptosis were dysregulated. *TP53* and *CTNNB1* were identified as direct targets of SOX30 hence responsible for directly regulating apoptosis via p53 pathway and cellular propagation via *Wnt* signalling pathway, respectively. Here, TP53 was upregulated in presence of SOX30 while β catenin expression was repressed (Han et al. 2015a, Han et al. 2018a).

SOX30 targets that have been identified to be misregulated in lung cancer cell lines were not found to be dysregulated in mice testicular tissues. Sox30 was found to be bound to the *CTNNB1* promoter but did not affect the expression levels in testicular tissues. These differences between the transcriptome studies in mice testis and lung cancer cells indicated that SOX30 regulates genes differently based on the type of cells it is expressed in. Hence, to study the mechanism behind *SOX30* as an epilepsy-associated gene as identified in our genetic study, it would be ideal to examine SOX30 targets in a neuronal cell line or neuronal tissue background. Given that the acquisition of the human brain tissue has been difficult, and the adult mouse does not seem to express SOX30, an overexpression model in the neuroblastoma cell line was used to study SOX30 genomic targets.

Commercially available antibodies were not suitable for chromatin immunoprecipitation; hence, an attempt was made to generate anti SOX30 antibody against the first 150 amino acids. Both Protein G and affinity purified antibody on western blotting analysis identified a band at around 100kDa, the expected size of SOX30, in cells transiently overexpressing SOX30 but immunocytochemistry analysis indicated signals at centrosomes in interphase HEK293 cells and was also observed at spindle poles through the cell cycle overlapping with tubulin staining in both endogenous and overexpression system (Figure S3.1). Endogenous SOX30 transcripts were not found in HEK293 cells. Due to this unexpected staining and discrepancy with existing data, this antibody was not selected. Instead, a FLAG antibody was used to immune-precipitate FLAG-tagged SOX30 overexpressed in SH-SY5Y, which did not express SOX30 endogenously.

SOX30 ChIP- Seq was performed in triplicates. Data from one of the replicates was of poor quality. The overlap between the first two replicates was weak, although the quality of individual experiments surpassed the standard threshold. Motif enrichment analysis did not provide accurate results due to poor overlap between the replicates. Selected candidate targets ranking high in peak scores were validated for SOX30 binding to its promoter region, indicating that the obtained ChIP-Seq data does represent SOX30 occupancy at a genomic level in these cells. These targets were *ZSCAN9*, *BCL2L1*, *BRWD1* and *STX16* (Figure S3.3).

ZSCAN9 alias *ZNF193* is a C2H2 zinc finger transcription factor that is expressed ubiquitously in all organs and tissues with the highest transcript expression in testis followed by the brain (GTex, FANTOM5). SCAN domain was first identified in this protein, which was initially identified from a fetal cDNA library when screening for a protein containing domains present in WT1 that are involved in transcriptional repression of growth factors in cells that do not express WT1 protein. *ZNF193* represses the expression of *PDNF* and *TGF-β* by binding to their promoter regions (Williams et al. 1995). It was identified as the second highest-ranking *SOX30* target where SOX30 binds within 1000bp upstream in the current study. In this region, about 500bps upstream of TSS, two copies of SOX30 binding motif 'ACAAT' in inverted tandem separated by 5bp is present, and two additional copies at 750bp and 970 bp upstream was present. Wildtype SOX30 resulted in increased expression of ZSCAN9, while a nonsense SOX30 mutant resulted in no change in expression when compared to untransfected cells indicating that *ZSCAN9* is a target of SOX30.

STX16 belongs to the syntaxin or t-SNARE (target-SNAP receptor) family that is located on cell membranes that interact and fuse with V-SNARES (vesicle-SNAP receptors) in certain synapses. STX16 isoforms A and B are found to localize in Golgi stack regions while isoform C was present in the cytosol (Simonsen et al. 1998, Tang et al. 1998). STX16 is ubiquitously expressed across various tissues with the highest protein expression in the brain (Simonsen et al. 1998, Chua and Tang 2009). In brain sections, STX16 is expressed in neurons and is localized to perinuclear regions, but not in oligodendrocytes and astrocytes, and is required for neurite outgrowth (Chua and Tang 2009). STX16 is required for retrograde transport of three exogenous cargoes, Shiga B-subunit toxin (STxB), cholera toxin B-subunit (CTxB), and ricin, and endogenous cargo protein mannose 6-phosphate receptor (MPR) in Hela cells and Glut4 in adipocytes between endosomes and the TGN/Golgi. It also regulates cell surface expression of cystic fibrosis transmembrane conductance regulator (*CFTR*) in intestinal epithelial cells (Proctor et al. 2006, Amessou et al. 2007, Gee et al. 2010). *STX16* is required for delivering

Cep55, *PDCD6IP*, and exocyst complexes to the intracellular bridges for abscission in telophase cells hence affecting cytokinesis in its absence (Neto et al. 2013). Deletions in chromosome 20 covering some exons of *STX16* is associated with Pseudohypoparathyroidism 1b in human while mice knocked out for *STX16* do not exhibit the phenotype (de Lange et al. 2016, Linglart et al. 2005, Sbrocchi et al. 2011, Turan et al. 2012, Nagasaki et al. 2013, Sano et al. 2015). STX16's expression in neurons, the requirement for dendritic outgrowth, and association with Pseudohypoparathyroidism 1b whose symptoms include convulsions indicate that its dysregulation could lead to epilepsy. In the current ChIP-Seq study, *STX16's* promoter is the target for SOX30 and is the third-highest ranked peak. On inspection of this region, no canonical SOX30 binding motif was identified. Overexpression of SOX30 wildtype repressed the expression of STX16 while that of the mutant enhanced its expression.

BRWD1 is a bromodomain and WD repeat domain-containing protein. It belongs to the WD-repeat protein family that is involved in several cellular functions. It is located in Down syndrome critical region 2 (DCR-2) on chromosome 21, which is the minimum region required in partial trisomy that causes down syndrome. BRWD1 functions as both a chromatin remodeler by interacting with BRG1, a component of the ATP-dependent SWI/SNF chromatin remodeling complex, and a transcriptional regulator (Huang et al. 2003). It is expressed ubiquitously in various adult human and mice tissues. During embryonic development, BRWD1 begins its expression at 8.5dpc, peaks at 11.5dpc after which it reduces (Ramos et al. 2002, Huang et al. 2003). BRWD1 is reported to regulate enhancers in pre B cells leading to lineage-specific transcription factor recruitment for B cell development. *Brwd1*$^{-/-}$ exhibits similar phenotypes as Hypogammaglobulinemia patients that have reduced antibody count, leading to decreased immunity, and increasing the risk of infections (Mandal et al. 2018). The current ChIP-Seq study suggested that SOX30 bound to the promoter region of *BRWD1* and the peak rested upstream to the BRWD1-IT2 (intronic transcript) that itself lies upstream of *BRWD1*. Overexpression of SOX30 wildtype resulted in suppression of BRWD1, while mutant SOX30 led to the upregulation of BRWD1.

BCL2L1, also called Bcl-X, is a Bcl-2-like protein 1 that belongs to the BCL-2 protein family involved in regulating apoptosis in various tissues and organs. It codes for two isoforms Bcl2-X(L) that codes for the longer isoform containing 4 Bcl-2 Homology domains (BH1, BH2, BH3, and BH4) and is antiapoptotic in its role. The shorter isoform Bcl2-X(S) does not have the BH1 domain and contains a partial BH2 domain and is proapoptotic. They localize on the outer membrane of mitochondria, thereby interacting with other Bcl 2 family of proteins

preventing pore opening leading to apoptosis (González-García et al. 1994, Lindenboim et al. 2001, Michelset al. 2013). Homozygous null mutants exhibit embryonic lethality due to extensive apoptosis in postmitotic neurons and hematopoietic cells (Motoyama et al. 1995, Fogarty et al. 2016). They are highly expressed during embryonic development, especially by regulating neurogenesis and haematogenesis (González-García et al. 1994, Fogarty et al. 2016, Opferman and Kothari 2017). BCL2L1 is known to interact with TP53, which is disintegrated when *PUMA*, a TP53 target, is recruited displacing BCL2L1, leading to apoptosis (Chipuk et al. 2005). Whereas SOX30 is known to regulate TP53 transcription and expression in lung cancer cell lines, thereby controlling apoptosis (Han et al. 2015a). The presence of SOX30 binding motif 'ACAAT' in the promoter of BCL2L1 led us to validate DNA binding of SOX30 at the promoter of BCL2L1, which was affirmative. Overexpression of SOX30 led to the suppression of BCL2L1, whereas SOX30 nonsense mutant expression led to upregulations of BCL2L1 protein expression.

Twenty percent of peaks identified as SOX30 targets in P28 mice testis had their expression levels differ between wildtype and knockout mice (Bai et al. 2018). Three percent of genes whose promoter was predicted to have SOX30 binding sites were found to be downregulated in round spermatids of *SOX30* knockout mice (Zhang et al. 2018). On comparing our ChIP targets where peaks which were in promoters or 5'UTR region to the predicted SOX30 targets identified in the Zhang et al. 2018 study, 163 genes were shared, which was about 8% of the predicted sites. Of the targets identified in promoter and 5'UTR of genes in our study, 31 were downregulated in P23 *SOX30* knocked out mouse testis consisting of mostly round spermatids and 5 in pachytene stage spermatocytes and 1 in round spermatids in P21 testis.

Recently, the Epi25 consortium has analyzed exomes of over 13000 epilepsy patients of various subtypes of European ancestry to identify variants that are ultra-rare, deleterious, and overrepresented in cases over 8000 ethnically matched controls. This examination highlighted the mutation burden in ion channels, mainly in inhibitory GABAergic receptor genes (Epi25 Collaborative 2019). *SOX30* targets that were identified in our study when matched with the Epi25 ultra-rare variation containing-gene list highlighted genes *DDX3X*, *JUN* and *EMP2A*- a previously reported epilepsy-associated gene that causes Lafora's disease was also a *SOX30* target. The current targets identified that were downregulated in round spermatids and pachytene cells of *SOX30* null mice were *LRRC8B* and *GSK3B.* On cross-examination of *SOX30* target genes dysregulated in *SOX30* null mice testis and genes that are known to have ultrarare variation in epilepsy patients identified genes *CCDC50*, *SOX5* and *ADCY10*.

CTNNB1 was also identified as a *SOX30* target, which has been reported to be regulated by *SOX30* in adenocarcinoma and acute myeloid leukemia cell lines (Han et al. 2018a, Han et al. 2018b, Liu et al. 2020). Autoregulation of *Sox30* has been reported in mice testis (Zhang et al. 2018, Bai et al. 2018). *SOX30* autoregulation was also found in our study where the top-ranking peak was located in the *SOX30* promoter region in all replicates. No single gene was identified to be shared across published *SOX30* targets, dysregulated genes in testicular tissues of *Sox30* null mice, and *SOX30* targets identified in our study.

Massive parallel sequencing of *SOX30* bound genomic sites in this study has led to the identification of probable *SOX30* regulating genes in SH-SY5Y cells. These targets differed from the reported targets identified in mice testicular tissues. No single-family or pathway-related proteins were identified to be enriched in the obtained dataset. Cell differentiation, proliferation, and apoptosis pathway genes had the highest-ranking peaks. There is a possibility that *SOX30* may not be directly regulating genes involved with epilepsy etiology. Compounding transcriptome analysis with ChIP-Seq analysis would provide a more confident dataset. Further studies using brain tissues expressing SOX30 could lead to more accurate identification of direct gene targets.

Mouse models can be used to explore Sox30's function in the brain. With brain expression restricted during embryonic stages, developmental defects would be interesting to look at neuronal migration and differentiation (Sox30 is required for post meiotic male germ cell differentiation), proliferation and apoptosis (Sox30 is a tumor suppressor), and electrophysiological defects in neurons of various brain subregions. In adult mice, spontaneous seizures are not reported yet hence, seizure susceptability using chemical convulsants, neuroanatomical examination, behavourial trait assesement and electrophysiological studies can be perfromed to understand SOX30's contribution to epilepsy in adult knockout model.

Chapter 4

CHD2 variants in juvenile myoclonic epilepsy

4.1 Summary

In this chapter, I discuss the contribution of *CHD2* rare variants to JME aetiology. *CHD2* is an ATP -dependent chromatin remodeller involved in genome reorganization and gene expression involved in developmental processes. Pathogenic de novo heterozygous mutations in *CHD2* have been implicated in various neurodevelopment disorders associated with early onset epilepsy. Photosensitivity is common in patients with *CHD2*-related epilepsy phenotypes. The overlapping clinical features between patients with JME and *CHD2* related neurodevelopmental disorders, mainly photosensitivity and seizure types, led to investigation of *CHD2* in aetiology of JME. Massive parallel targeted sequencing was conducted on 189 JME cases of south Indian ancestry, covering all the 39 exons of *CHD2*. 251 ethnically matched healthy individuals were sequenced for the rare variants detected. Twelve rare variants were identified of which three were heterozygous missense variants and two were located at splice site regions. These variants have not been previously reported in patients with *CHD2*-related neurodevelopment disorder. Consequence of two splice variants analysed by exon trap assays indicated that variant c.4692+1G>C induced insertion of five nucleotides from the adjacent intron into the cDNA. At the protein level, the variant is predicted to lead to early truncation of the protein at amino acid residue 1574 leading to loss of the DUF1777 domain. Identification of rare variants in JME patients and presentation of splicing defect by a canonical splice site variant suggests a role of *CHD2* as a genetic risk factor for JME.

4.2 Introduction

CHD2-linked neurodevelopmental disorders include phenotypes such as developmental delay, intellectual disability, autism, and epileptic encephalopathy with convulsions primarily being myoclonic and absence seizures, being a consistent feature across the phenotypic spectrum (Carvill et al. 2013, Carvill et al. 2015, Lamar et al. 2018). *CHD2* variants are overrepresented in patients with photosensitive epilepsies, over those with photoparoxysmal response without seizures (Galizia et al. 2015). A large fraction of patients with juvenile myoclonic epilepsy exhibit photosensitivity with early-onset seizures as well as myoclonic and absence seizures. With these apparently common features, we considered screening *CHD2* in a set of juvenile myoclonic epilepsy patients. Massive parallel sequencing of *CHD2* after target enrichment of 39 exons was carried out in 189 unrelated JME patients. Ethnically matched control samples

and NGS identified variants were validated by Sanger sequencing. Here, I describe outcome of the study.

4.3 Materials and methods

4.3.1 Patients

One hundred and ninety-two unrelated patients affected with juvenile myoclonic epilepsy were subjected to the *CHD2* targeted resequencing study. Ethnically matched, 192 unaffected individuals were considered as controls. Additionally, 59 non-JME individuals from the same population whose whole-exome sequencing data were available, were also considered as controls for this study. Study participants were primarily of south Indian descent. Written informed consent was obtained from all individuals participating in the study. Ten millilitres of venous blood was collected, and genomic DNA was extracted using phenol-chloroform method. The quality of genomic DNA was checked on a 1.5% agarose gel and quantified using Nanodrop (Thermo Fisher) and Qubit double stranded DNA assay (Thermo Fisher).

4.3.2 Primer design

Sequence coordinates and exon-intron boundaries of *CHD2* were obtained from the human genome build GRCh38/hg38. Primers were designed to amplify exons and 50bp flanking intron sequences. Primer pairs for 55 amplicons covering 39 exons of the gene were synthesized. The average size of the amplicons was 400bp. The primer pool was validated by sequencing a human DNA sample, and 97.6% of the target region was covered.

4.3.3 Library preparation

Multiplex PCR was performed for each sample using two pools of primers sets: Pool A contained odd while the Pool B contained even sets of primer pairs. The PCR reaction mix included 25ul of 2X Cleversense reaction mix, 5ul of primer pool and 20ul of genomic DNA (7ng/ul). The PCR conditions were 95°C for 5 min initial denaturation and 25 cycles of 95°C for 30s, 55°C for 2 mins and 72°C for 1 min and final extension at 72°C for 4 min. The PCR product was checked on a 2% agarose gel to confirm amplification and the left-over PCR product was purified using AMPure XP (Beckman coulter). 5ul of each PCR pool was mixed, followed by Illumina barcoded adapters and index primer addition using limited cycle PCR whose conditions were 95°C for 5min, 25 cycles of 95°C for 30s, 55°C for 45s and 72°C for

1min and final extension at 72°C for 4min. Barcoded libraries were purified using AMPure XP magnetic beads (Beckman coulter).

4.3.4 Targeted resequencing

The barcoded libraries were pooled in equimolar amounts and sequenced on Illumina MiSeq platform to generate 2 x 300 paired-end reads using MiSeq sequencing reagents v3 as per manufactures protocol (Illumina, USA). One flow cell was used for 192 samples.

4.3.5 NGS data analysis

Raw data quality was checked using FastQC. TrimGalore and Trimmomatic was used to remove adapter sequences and low-quality bases, respectively. Reads below 35bp length were excluded from further analysis due to spurious alignment issues. Quality filtered reads were mapped to the human reference genome (GRCh38 build, hg38) using Burrow-Wheelers aligner's algorithm with default parameters. Post alignment processing was performed on BAM files that included sorting and indexing using SAMtools package. Picard tools was used to remove PCR duplicates. Variant calling was performed using Genome analysis tool kit (GATK). Only variants with quality score greater that 20 were called. Vcftools v.0.1.14 was used to calculate allele frequency of the variants called.

4.3.6 Variant filtration and Sanger confirmation

Candidate variants were manually checked on Integrative Genomics Viewer using BAM files. Synonymous and intronic variants beyond 50bp from splice site and poor-quality variants were excluded from further analysis. Primers were designed using Primer3 software, and features were checked using OligoCalc. Sanger sequencing was performed on all the filtered variants in patients and 192 control samples (Table A4.3). Regions uncovered by targeted resequencing were also Sanger sequenced. Sanger sequencing confirmed variants were then filtered to exclude variants whose minor allele frequency was less than 0.005 as obtained from dbSNP database that contains evidence from 1000Genome, INDEX-db, TOPMED, GnomAD, ExAc, EVS, TWINSUK, PAGESTUDY, Go-ESP, ALSPAC, GenomeAsia100k and Indigenomes directory.

4.3.7 Bioinformatic analysis

Prediction tools were used to forecast the probable effect of the variants. Conservation of nucleotide across species was checked using Consurf and CADD. Missense mutations were

analyzed using SIFT, PolyPhen, FATHMM, PANTHER, Mutationtaster, Provean, PhD-SNP, MutationAsssessor, MutPred2, SNP&GO and AlignGVGD. Splice site variants were analysed using Human splicing finder and MutationTaster. 5'UTR variants were checked for their location in binding sites of other proteins using UCSC genome browser tracks and 3' UTR variants were checked for miRNA coding sequence and binding sites from miRDB, TargetScan and mirWalk.

4.3.8 Minigene reporter construction for splice site variants

Vector pSPL3 was kindly provided by Dr. Stuart Tompson, University of Wisconsin, USA (Tompson et al. 2017). Exons containing the splice variant with a minimum of 200bp of flanking intronic were amplified using primers containing restriction sites SacI and BamHI at 5' and 3' end of the amplicon, respectively (Table A4.4). The amplicon was inserted in the multiple cloning site that is located in between two artificial exons in the vector backbone using restriction digestion-ligation technique. Side directed mutagenesis using Quikchange® SDM kit (Stratagene) was done to introduce the splice variants.

4.3.9 RNA extraction

Splicing reporter plasmids with wildtype minigene or variant minigene [c.1720-3T>C (Exon 15), c.4692+1G>C (Exon36)] constructs and the empty reporter vector were transfected into HEK293 cells seeded in 6-well dishes at about 40% confluency using Lipofectamine (Thermo fisher). Twenty-four hours after transfection of 1ug of the reporter plasmid, cells from each well were washed with ice-cold 1X PBS and resuspended in 500ul of TRIzol™ reagent (Invitrogen, USA) and incubated for 15 minutes. A set of experiments were also performed by treating cells with 300ug/ml Cycloheximide for 5 hours just 24hours post-transfection. All the glassware used to handle RNA was DEPC-treated, and plasticware was autoclaved twice. Solutions were made using DEPC-treated autoclaved double distilled water. 100ul of chloroform was added to the TRIzol™-treated cells and was incubated for 5 minutes, followed by centrifugation at 13000rpm for 15min at 4°C. The upper aqueous layer was collected carefully, and an equal volume of isopropanol was added, mixed gently and then incubated for 10min. The sample was centrifuged for 10min at 13000rpm, 4°C to pellet the RNA. The supernatant was carefully removed, and the pellet was washed with 75% alcohol by displacing the pellet by vortexing the sample and then centrifuging at 13000rpm for 5min at 4°C. The supernatant was removed, and the pellet was allowed to dry in thermomixer at 60°C for 30min. 30ul of DEPC treated water was added to the pellet and was allowed to dissolve in it for 30min

at 60°C. The RNA quality was checked on a 1% agarose gel and quantified using Nanodrop (Thermo fisher) and was stored in -80°C. Samples that had good RNA integrity and 260/280 ratios between 1.8-2.1 were used for cDNA synthesis.

4.3.10 cDNA synthesis

cDNA was synthesized from 2ug of RNA using SuperScript™ III Firststrand synthesis system (Thermo Fisher) in a 10ul reaction. A 5ul mix containing RNA ,0.5ul of OligoDT primer and 0.5ul of 10mM dNTP mix and water was incubated at 65°C for 5min and then placed on ice for 1min. Another 5ul mix containing 1ul of 10X RT buffer, 2ul of 25mM $MgCl_2$, 1ul of 0.1M DTT and 0.5ul of RNaseOUT™ and Superscript® III each was added to the RNA mix followed by incubation at 42°C for 50 minutes and then at 70°C for 15 minutes followed by snap chilling on ice for at least 1 min. RNase H (0.5ul) was added to degrade residual RNA and, cDNA was stored at -20°C.

4.3.11 PCR and qPCR amplification of spliced product

PCR was performed using 0.5ul of cDNA using primers V1-F (5' TCTGAGTCACCTGGACAACC 3') and V2-R (5' ATCTCAGTGGTATTTGTGAGC 3') located at 5' end of exon V1 and 3' end of exon V2 respectively using Taq polymerase in a 20ul reaction. PCR was performed on GeneAmp PCR system 9700 (Applied Biosystems) with the conditions: 94°C for 5min, 40 cycles of 94°C for 30s, 55°C for 30s and 72° for 3min followed by a final extension at 72°C for 15min. The amplicon from the empty vector generates a 262bp product. qPCR was performed using Faststart universal SYBR master mix (Roche) on CFX96 Touch Real-Time PCR System (Biorad). A 20ul reaction with 0.5ul of cDNA with 0.5ul of primers V1F and V2R, 10ul of 2X SYBR master mix and 8.5 ul of distilled water was prepared. The reaction was run using the following program: 95°C for 10min, 40 cycles of 95°C for 15sec and 55°C for 1min followed by melt curve beginning at 65°C for 5sec with 0.5°C increment till 95°C. Cq values were noted post-run. The values were normalized to GAPDH expression using primers GAPDH – F: 5'TCACCACCATGGAGAAGGCT 3' and GAPDH – R: 5' AAGCAGTTGGTGGTGCAGGA 3' to amplify a region of GAPDH cDNA. Student unpaired t-test was used for statistical analysis. Both PCR and qPCR products were run on 1.5% agarose gel to check the product's size.

4.4 Results

4.4.1 Clinical features of JME patients

Of the 192 samples sequenced, 9 were IGE, while the rest 183 were JME patients. Among these, 88 were males, and 104 were females. Electroencephalograms were normal in 28 members, generalized spikes and waves were present in 120 members, 4 samples had 3hz spike and wave discharges while EEG information was unavailable for 40 individuals. Photo paroxysmal response was present in 57 of these patients, absent in 107, and was unavailable for 28 samples.

4.4.2 Targeted sequencing average coverage and read depth

Of the 192 samples taken up for targeted sequencing for *CHD2*, three samples under-performed where the average read depth was below 15. The average read depth other 189 samples ranged between 81 to 27903, with the mean being 2557X. Exon 35 did not get sequenced efficiently- an average read depth of 1 with the highest read depth of 10. Hence, this exon was sequenced using Sanger-based method for the 189 samples. The average read depth of exon 20 across samples was 28, with values ranging between 0-243, which was comparatively lower when compared to other exons (Figure A4.1, A.4.2). (Acknowledgement: Clevergene Pvt. Ltd)

4.4.3 Targeted sequencing analysis results and variant filtration

In total, 49 variants were identified from NGS analysis (Table A4.1). Variants with their minor allele frequencies greater than 0.005 were 19 that included, 2 synonymous variants, 9 intronic variants that lay beyond 30 bases from the exon-intron boundary and 2 low quality 3'UTR variants. These were not considered for further downstream analysis. On Sanger sequencing, two 3' UTR variation and one frameshift variation were found to be false. On sequencing ethnically matched control samples, one missense variation was found to be common and was excluded from further analysis. Exon 35 Sanger sequencing identified three variants whose MAF was less than 0.005 and these were absent in 251 healthy control individuals. Among these, one intronic and one synonymous variant were not taken up for further analysis (Table A4.2).

4.4.4 Variant interpretations using prediction tools, published literature and datasets

In summary, 12 rare variants were identified in *CHD2* among 189 JME/IGE patient samples. Two were in the 5'UTR region, 5 were in 3'UTR region, 2 were located at splice site regions

and 3 were missense variants (Table 1; Figure 1,2). All these variants were heterozygous except for 3'UTR variant c.*248C>T that was homozygous. The variant nucleotide was present in more than 25% of the reads. Clinical features of the individuals with *CHD2* variation are detailed in Table S4.5. We do not see photoparoxysmal response to be present in all patients with rare CHD2 variations in this cohort. The number of females with *CHD2* rare variation was higher when compared to males.

Variants in the 5'UTR were in binding sites of the RBL2 (Retinoblastoma-like protein 2) and SMARCA4 (SWI/SNF related, matrix-associated, actin-dependent regulator of chromatin subfamily a, member 4) proteins as inferred from ORefAnno datasets. Bioinformatic analysis (Table 2) indicated that the missense variant, p.Gln153Glu results in amino acid change with a small physiochemical difference. It is currently not reported in patients with *CHD2*-related neurodevelopmental disorders and is predicted to be tolerated by most prediction tools. ClinVar reports the variation to be of uncertain significance due to the clashing interpretation of pathogenicity. p.Glu1500Asp missense variant's prediction ranges from tolerated to strongly damaging. Although it is a highly conserved amino acid, the variation does not seem to drastically alter the physiochemical properties. It is also not reported in patients with a *CHD2*- related neurodevelopmental disorder, and ClinVar has currently classified it as variation with uncertain significance. Missense variant p.Pro1699Ser is predicted with conflicting effects of being tolerated or damaging. Proline 1699 is also a highly conserved residue, and the change may introduce a difference in the secondary structure of the protein. This change is reported in the Epi25K database in a GGE patient and in a matched control sample, but this variant is not listed in ClinVar. The five 3' UTR variants were spread throughout the length of the UTR. Of these, two were novel. All these 3'UTR variants were poorly conserved and were predicted to be polymorphisms. None of these are in known miRNA coding or binding sites. Splice variant c.1720-3T>C is at the splice acceptor site. The nucleotide is moderately conserved, while in some species, the variant nucleotide is found instead. Prediction tools have conflicting interpretations of its effect. Splice variant c.4692+1G>C is a novel variant located at a highly conserved splice donor site. Prediction tools have indicated the variant to be damaging for splicing.

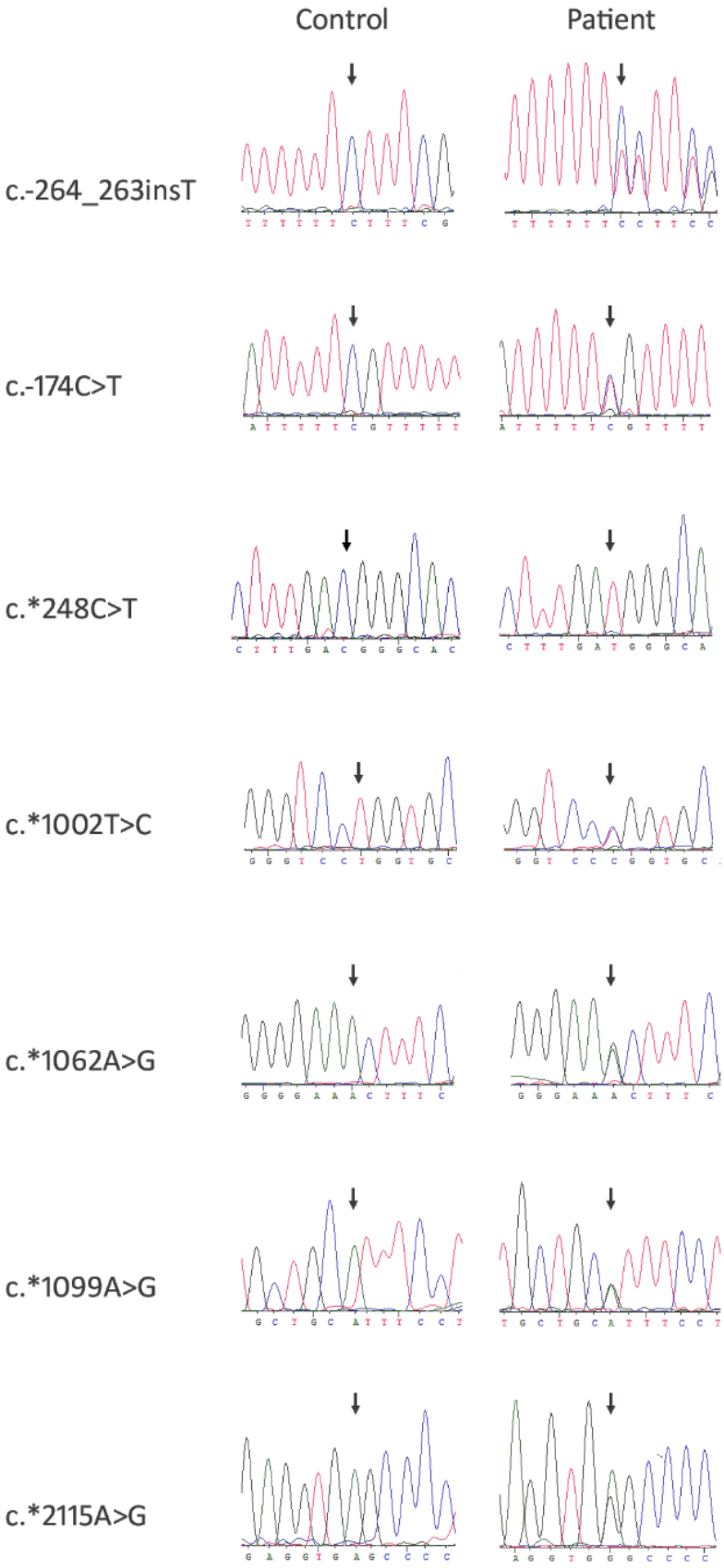


Figure 1. Electropherogram representation of 5'UTR and 3'UTR *CHD2* variants in the patient and control samples

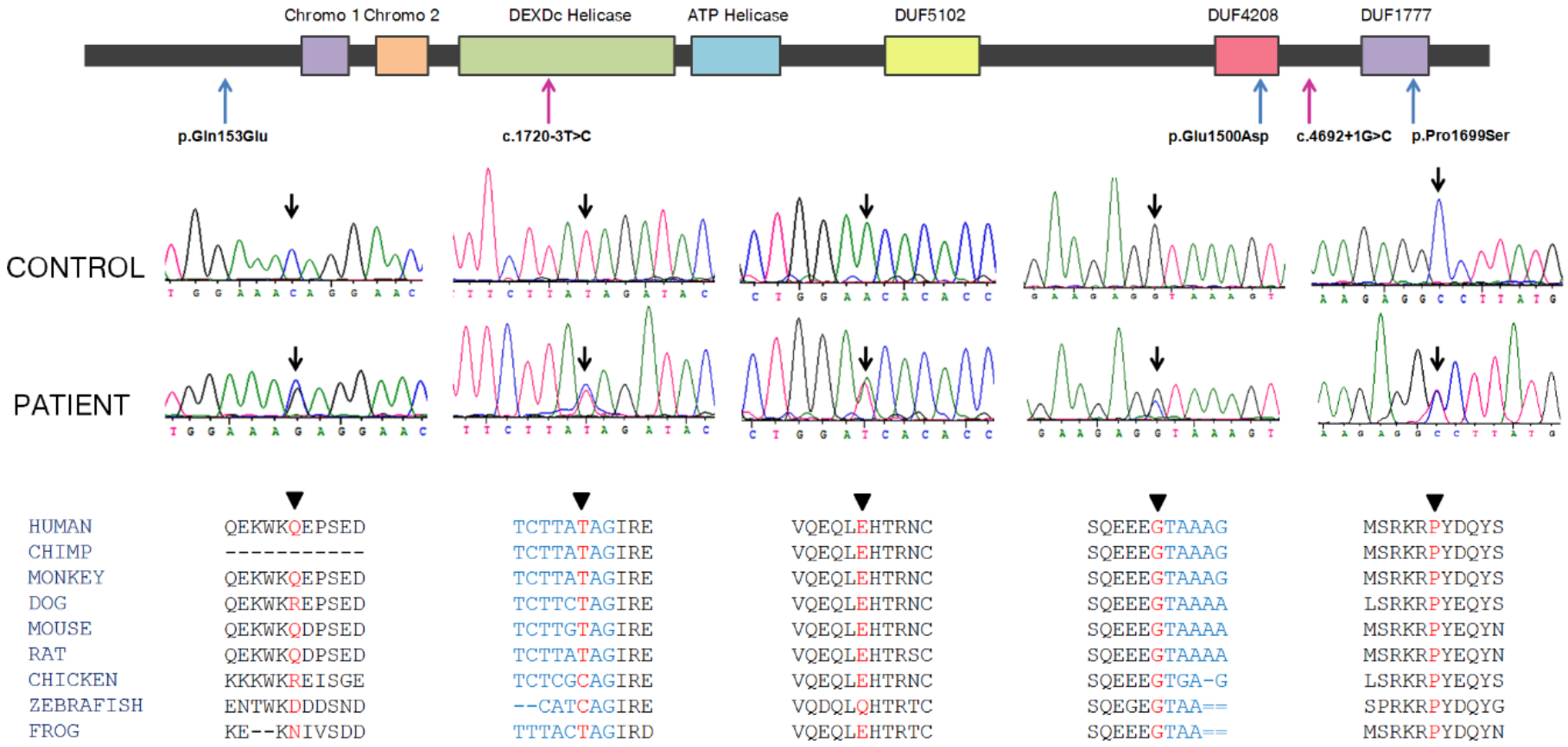


Figure 2. Schematic representation of missense and splice site variants along CHD2 protein and the electropherogram representation of the variants in patient and control samples. Conservation of amino acids or nucleotides across 9 species ranging from amphibian to mammals.

Table 1. *CHD2* rare variants in JME patients

Genomic position	Nucleotide change	Amino acid change	Exon/ Intron	Consequence	SNP rsiD	Conservation	Inhouse control Allele frequency	Minor Allele Frequency	GA100K / GA100K India / Indigene MAF	Mutationtaster
92900632	c.-264_263insT	na	Exon 1	5'UTR	rs1389168993	na	0/502	0.00006493 (GnomAD)	- / - / -	disease causing
92900722	c.-174C>T	na	Exon 1	5'UTR	rs540465851	5/10	1/502	0.00039936 (1000g)	0.006612 / 0.0192 / 0.0039	disease causing
92937531	c.457C>G	p.Gln153Glu	Exon 6	Missense	rs755510106	9/10	0/502	0.00002514 (GnomAD)	0.000575 / 0.0017 / -	polymorphism
92955420	c.1720-3T>C	na	Intron 14	splice region	rs759231634	8/10	0/502	0.00004119 (ExAc)	- / - / 0.0010	disease causing
93009231	c.4500A>T	p.Glu1500Asp	Exon 35	Missense	rs776113114	10/10	0/502	0.00007/9 (ExAC)	- / - / 0.0005	disease causing
93012445	c.4692+1G>C	na	Intron 36	splice donor	-	10/10	0/598	-	- / - / -	disease causing
93020200	c.5095C>T	p.Pro1699Ser	Exon 38	Missense	rs777481677	10/10	0/502	0.00003296 (ExAc)	- / - / -	disease causing
93024953	c.*248C>T	na	Exon 39	3'UTR	rs1051478439	2/10	0/502	0.00003185 (TOPMED)	- / - / 0.0005	disease causing
93025707	c.*1002T>C	na	Exon 39	3'UTR	rs1449923910	3/10	0/502	0.00003230 (GnomAD)	0.000287 / - / -	polymorphism
93025767	c.*1062A>G	na	Exon 39	3'UTR	rs1440077664	3/10	0/502	0.000007964 (TOPMED)	- / - / -	polymorphism
93025804	c.*1099A>G	na	Exon 39	3'UTR	-	5/10	0/502	-	- / - / -	polymorphism
93026820	c.*2115A>G	na	Exon 39	3'UTR	-	4/10	0/502	-	- / - / -	polymorphism

Conservation organisms: Human, Chimp, Monkey, Mouse, Rat, Dog, Cow, Chicken, Zebrafish, Frog
Databases checked for MAF: GnomAD, ExAc, TOPMED, 1000g, EVS, TWINSUK, ALSPAC, Estonian. Highest MAF for the variant across databases is listed.

Table 2. Bioinformatic analysis of missense and splice site variants identified in *CHD2*

Variation	SIFT[a]	PolyPhen2[b]	PANTHER[c] (preservation time)	FATHMM[d]	Mutation AssessorFIS[e]	Consurf[f]	PhD-SNP[g]	Provean[h]
p.Gln153Glu	Tolerated (1.00)	Benign (0.000)	probably damaging 97	Damaging (-2.44)	Neutral (0)	Variable	Neutral	Neutral (0.389)
p.Glu1500Asp	Tolerated (0.092)	Possibly damaging (0.580)	probably damaging 455	Damaging (-2.62)	Low (1.265)	Variable	Neutral	Neutral (-1.310)
p.Pro1699Ser	Tolerated (0.182)	Benign (0.004)	probably damaging 456	Damaging (-2.53)	Low (1.415)	Variable	Neutral	Neutral(-2.104)

Variation	CADD[i]	Mutationtaster[j]	UMD predictor[k]	SNPs&GO[l]	AlignGVGD[m]	MutPred2[n]	Envision[o]	SNAP2[p]
p.Gln153Glu	18.38	polymorphism	Polymorphism (30)	Neutral	less likely	0.111	1.00616	Neutral (-82)
p.Glu1500Asp	22.9	disease causing	probably pathogenic (72)	Neutral	less likely	0.348	0.905738	Neutral (-57)
p.Pro1699Ser	22.9	disease causing	Pathogenic (90)	Neutral	less likely	0.27	0.94965	Neutral (-40)

	MaxEntSCAN[q]	Mutationtaster[j]	HumanSplicingFinder[r]
c.1720-3T>C	-8.28	disease causing	no impact on splicing
c.4692+1G>C	-1.51	disease causing	probably affecting splicing

[a] Sorting Intolerant From Tolerant [SIFT]: scores range from 0-1. Variant scores ranging from 0.0-0.05 are considered deleterious. Scores between 0.05-1.0 are considered tolerated.

[b] Polymorphism Phenotyping v2 [PolyPhen-2]: Scores range from 0-1 as being benign to damaging.

[c] PANTHER-PSEP: predicting disease-causing genetic variants using position-specific evolutionary preservation

[d] Functional Analysis through Hidden Markov Models (FATHMM).

[e] Functional impact score (FIS): Neutral impact (FIS ≤ 0.8), Low impact (0.8 < FIS ≤ 1.9), Medium impact (1.9 < FIS ≤ 3.5), High impact (FIS > 3.5).

[f] Consurf Server; conservation scores for the amino acids based on the phylogenetic relationship from 41 homologous sequences.

[g] Predictor of human Deleterious Single Nucleotide Polymorphisms (PhD-SNP)

[h] Protein Variation Effect Analyzer [PROVEAN]: Variants with a score equal to or below -2.5 are considered ‘deleterious‘. Variants with a score above -2.5 are considered ‘neutral’.

[i] Combined Annotation Dependent Depletion (CADD) PHRED-like (-10*log10(rank/total)) scaled C-score ranking a variant relative to all possible substitutions of the human genome (8.6x10^9).

[j] Mutation Taster: Probability of prediction; values close to 1 indicates a high security of the prediction.

[k] UMD-Predictor: Combines biochemical properties, impact on splicing signals, localization in protein domains, variation frequency in the global population, and conservation. Score range 0-100, <50-polymorphism,50-64-probablr polymorphism, 65-74-probably pathogenic , >74 pathogenic.

[l] SNPs&GO: Predicting disease associated variations using GO terms

[m] AlignGVGD: Multi species alignment-based prediction

[n] MutPred2: ensemble of 30 feed forward neural network trained on subset of pathogenic and unlabelled variants.

[o] Envision: Large scale experimental mutagenesis datasets used to interpret variants molecular effect. Score range 0-1 from most damaging to wildtype like respectively.

[p] SNAP2: predicts impact based on amino acid substitution. Score ranges from -100 neutral to +100 strong effect prediction

[q] MaxEntScan: Maximum entropy model used on short sequences, greater the difference between wildtype and mutant’s score, higher the probability for the variant to have an effect.

[r] HumanSplicingFinder:

4.4.5 Minigene assay of the splice variants identified.

Heterologous splicing assays were conducted in cultured HEK293 cells to test effect on pre-mRNA splicing by the mutant and wildtype *CHD2* minigene constructs of variants c.1720-3T>C and c.4692+1G>C. RT-PCR of RNA from untransfected cells and empty vector minigenes expressing cells resulted in a single band of 260 bases. Sequencing confirmed the correct splicing of the artificial exons present in empty vector pSPL3. While the single PCR band in untransfected lane was unexpected, sequencing analysis revealed the primers to have amplified a delta globin exon, which shares sequence similarity to the artificial exons present in the pSPL3 vector. The delta globin exon was not amplified in cells overexpressing minigene constructs.

RT-PCR of wildtype RNA resulted in a single PCR product of 360 bases for exon 15 and 36, and Sanger sequencing confirmed the correct splicing of the transcript. RT-PCR of the mutant RNA for exon 15 did not differ from its wildtype counterpart in both amplicon size and expression, and sequencing analysis confirmed generation of correctly spliced CHD2 exon 15 (Figure 3). In contrast, RT-PCR of the mutant RNA for exon 36 resulted in a faint product of the expected size, and the expression differences between wildtype and variant spliced products was confirmed by qPCR. Sequencing analysis of the mutant amplicon revealed an insertion of first 5 nucleotides from intron 36 post exon 36 (Figure 4). No wildtype exon 36 encoding transcript was generated. This aberrant transcript would lead to insertion of nine novel amino acids followed by a premature termination codon that would result in a protein short of 254 amino acids at the C- terminal end thereby lacking the DUF1777 domain. The RT-PCR and qPCR analysis also reveals very poor expression of this minigene indicating that it undergoes nonsense-mediated mRNA decay. Cycloheximide treatment of cells expressing the minigene constructs did not drastically affect the expression levels of the transcripts generated by minigenes.

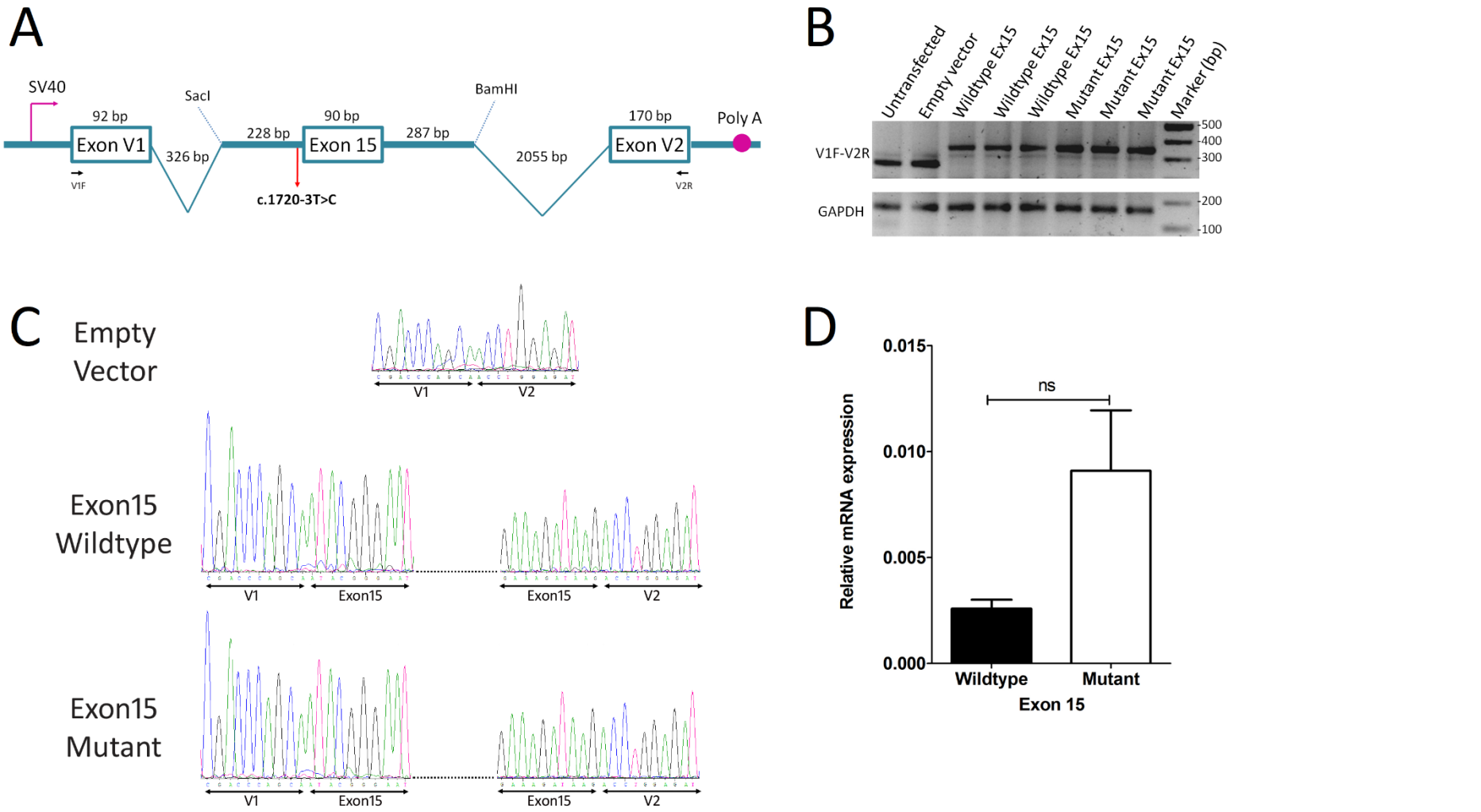


Figure 3. Minigene splicing assay for c.1720-3T>C. (A) Schematic representation of exon 15 construct in pSPL3 vector. (B) RT-PCR products separated on 1.5% agarose gel. (C). Sequence of the RT-PCR products (D). qPCR quantification of transcripts, unpaired t test

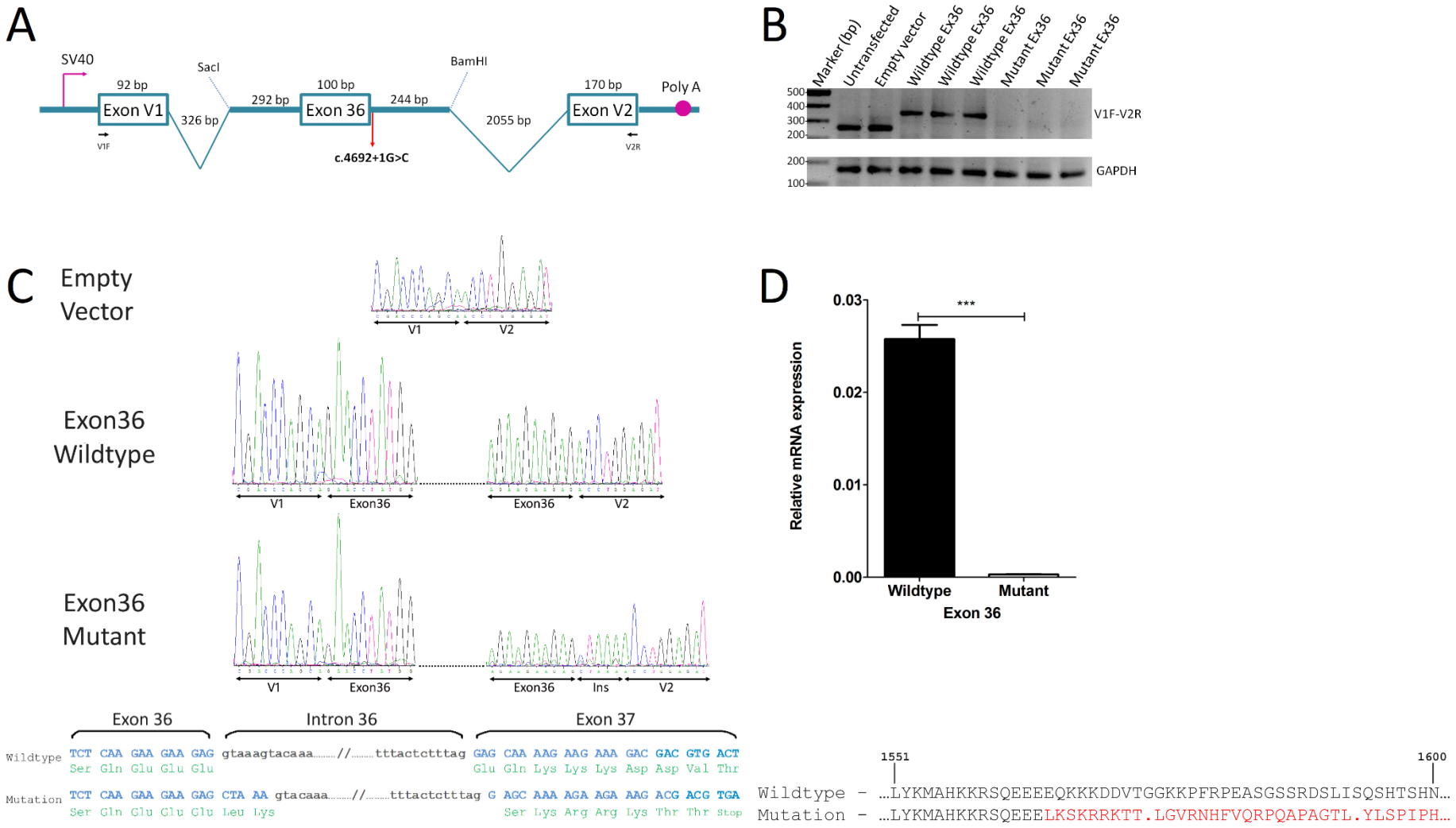


Figure 4. Minigene splicing assay for c.4692+1G>C. (A) Schematic representation of exon 36 construct in pSPL3 vector. (B). RT-PCR products separated on 1.5% agarose gel. (C). Sequence of the RT-PCR products D. qPCR quantification of transcripts, unpaired t test , $p<0.0001$.

4.5 Discussion

CHD2 was first reported to be an epilepsy-causing gene in studies where *de novo* microdeletions ranging from 511Kb to 5Mb at 15q26.1 in patients with seizures and developmental delay were identified (Veredice et al. 2009, Dhamija et al. 2011, Capelli et al. 2012, Courage et al. 2014). Rare *de novo CHD2* variants were also identified in patients diagnosed to have myoclonic-astatic epilepsy (Carvill et al. 2013, Thomas et al. 2015, Trivisano et al. 2015, Routier et al. 2019, Chen et al. 2020). Lennox Gastaut syndromic patients were found to carry frameshift and splice site *CHD2* mutations (Epi4k Consortium 2013, Carvill et al. 2013, Lund et al. 2014, Thomas et al. 2015, Chen et al. 2020). Dravet syndrome patients known to carry mutations in *SCN1A*, are also found to be associated with *CHD2* mutations (Suls et al. 2013, Thomas et al. 2015). Jeavons and West syndrome patients also carried pathogenic *CHD2* mutation (Chénier et al. 2014, Chen et al. 2020). A few cases with non-specific epileptic encephalopathy who also have intellectual disability or autism spectrum disorders harbour *CHD2* mutations (Rauch et al. 2012, Carvill et al. 2013, Hamdan et al. 2014, O'Roak et al. 2014, Caputo et al. 2018). Monozygotic twins with late-onset absence seizures and other neurodevelopmental disabilities had a frameshift and nonsense variation in *CHD2* (Pinto et al. 2016, Wang et al. 2017). A frameshift mutation has been identified in a patient with childhood onset schizophrenia along with febrile and generalised tonic-clonic seizures (Poisson et al. 2020). A missense variation was also identified in a patient with West syndrome (Chen et al. 2020). Genetic generalized epilepsy and eyelid myoclonia with absence seizure who exhibited photosensitivity were found to have *CHD2* pathogenic mutation. (Galizia et al. 2015). These studies indicate the role of *CHD2* in a rather broad spectrum of neurological and neuropsychiatric phenotypes.

Zebrafish larvae, where partial loss of *chd2* was introduced using morpholinos, exhibited abnormal movements that include twitching, trembling, poly-spike discharges, and photosensitivity, which was enhanced at 4dpf when compared to 7dpf larvae, recapitulating the phenotype seen in human *CHD2* mutation carrying patients. Other developmental abnormalities such as development delay, abnormal body curvature, microcephaly, excessive body pigmentation, pericardial edema, missing swim bladder, and stunted growth observed in zebrafish were not observed in human patients with *CHD2* mutations (Suls et al. 2013, Galizia et al. 2015). Mice models with homozygous deletion of the gene exhibit embryonic lethality, while heterozygous mice have growth retardation and kyphosis that are also seen in the zebrafish model. Other features observed were reduced body fat, cardiomyopathy,

glomerulopathy, enlarged spleens, but no apparent neurological irregularities (Kulkarni et al. 2008). Mice containing mutation leading to loss of DNA- binding domains exhibited compromised viability, growth retardation and multi-organ defects especially renal abnormalities (Marfella et al. 2006, Marfella et al. 2008).

In vitro and *in vivo* studies have found that *Chd2* knockdown in mice leads to decreased proliferation of radial glial cells and promotes generation of intermediate progenitors that leads to premature differentiation of neurons by regulating expression of REST (repressor element 1-silencing transcription factor) whose targets include genes involved in neurite growth such as *Celsr3*, *Mash1*, and *Ngn2* (Shen et al. 2014). *Chd2* haploinsufficiency results in abnormal excitatory and inhibitory synaptic transmission in the hippocampus, elevated cortical synchrony, and deficits in long term spatial and recognition memory due to alteration in genes involved in embryonic neurogenesis and synaptic transmission (Kim et al. 2018). Electrophysiological defects were observed in *Chd2* deficient interneurons but not in cortical excitatory neurons (Meganathan et al. 2017). These studies have suggested that *CHD2* plays an important role in cortical development and neuronal differentiation.

Considering these data, we were interested in examining *CHD2* in JME/IGE patients. Targeted sequencing was carried out on 189 patient DNA samples. We identified 12 rare *CHD2* variants of which three were novel among patients of south Indian ancestry. These variants were also rare or absent in our ethnically matched in-house control individuals. None of the variants were present in more than one patient. Two patients with *CHD2* rare variants had photoparoxysmal response. This is different from a published finding where most patients with *CHD2* mutations exhibited clinical photosensitivity (Galizia et al. 2015). No clinical characteristic was found to be common to all patients with *CHD2* rare variants. *CHD2* RVIS score is -1.75 with a percentile of 2.37%, pLI score of 0.07 indicates that it is highly intolerant to introduction of variations. Hence, the identification of variants in our cohort indicates these variants to be pathogenic.

The three missense variants identified have not been reported in any other *CHD2* -related neurodevelopmental disorder study. Prediction tools have indicated their effect to range between benign to damaging. A literature study identified only one JME patient with a missense *CHD2* variant located in the DNA- binding domain (Galizia et al. 2015). Although a recent report identified a nonsense variant in *CHD2* in a JME patient, he also had distinct developmental disabilities and drug resistance, which are not typical features of JME (Singh et

al. 2020). The relatively subtle *CHD2* variants identified in JME patients in this study probably reflect the milder phenotype in JME contrary to other brain disorders associated with the gene. Characterization of pathogenic human *CHD2* missense mutations affecting neuronal aspects using *in vitro* functional assays have not been reported. However, missense somatic variants p.His620Leu, p.Phe1146Leu, and p.His1270Phe associated with chronic lymphocytic leukemia were found to disturb the homogenous nuclear distribution observed for the wildtype protein, on overexpression in COS7 cells (Rodríguez et al. 2015). Functional analysis of missense variants identified in our study could probably show differences in nuclear localization and provide an indication for the mechanism by which the variants could lead to epilepsy. Splice-region variants were subjected to *in vitro* minigene assay in cultured HEK293 cells to study the effect of the variants on the splicing of pre- mRNA. Intron 14 variant exhibited no difference in both transcript sequence and quantity when compared to the wildtype. In our study, the intron 36 variant c.4692+1G>C was found to disrupt the coding frame by introducing premature stop codon. This occurs due to insertion of 5 bases after exon 36 by recognizing a new acceptor site in intron 36 altering the reading frame, while also exhibiting poor transcript quantities when compared to wildtype and was not accompanied by wildtype spliced mRNA. At the protein level, the variant would probably lead to loss of the DUF1777 domain. The function of the C- terminal region of the protein is unknown; however, identification of several missense, frameshift and truncating mutations in the region among epileptic patients indicates its importance for the structure or function of CHD2 protein (Lamar et al. 2018). The low subRVIS percentile score (8.74) of the C terminal region indicates its intolerance to variations. Functional analysis of a splice variant identified in fever-sensitive epileptic encephalopathy patients in zebrafish using morpholinos has identified abnormally spliced mRNA. The larvae with abnormal splice products also exhibited morphological and behavioural defects, including epileptiform discharges, as observed by field potential recordings (Suls et al. 2013). The intron 14 variant identified in this study did not exhibit any effect on pre-mRNA splicing in HEK293- examining the same in a different cell line or in stably expressing cells may capture the defect, if any. In summary, our study has identified rare heterozygous *CHD2* variants in JME patients, expanding the spectrum of phenotype of *CHD2*-related neurodevelopmental disorders. Although *CHD2* is not the major contributor to JME, a detailed functional analysis of these variants found would throw more light on the mechanism by which the gene may contribute to the disorder.

Appendix I*: SOX30* and male infertility

Introduction

Sox30 was first identified as a Wilms' tumor suppressor (Wt1) -associated protein in adult mice testis in a yeast two-hybrid screen (Osaki et al. 1999). *WT1* is essential for male gonadogenesis to maintain sertoli cell functions and regulate early germ cell differentiation (Gao et al. 2006, Wang et al. 2013, Zheng et al. 2013, Chen et al. 2017). Several variations in *WT1* have been identified that contribute to human male infertility (Wang et al. 2013, Seabra et al. 2015, Xu et al. 2017). Sox30 is highly expressed in foetal and adult testicular tissues and was therefore proposed to be involved in male gonad development. In mice testis, Sox30 expression is absent in somatic cells and type A spermatogonia, but it is expressed in Sertoli cells, Leydig cells and germ cells particularly in pachytene, diplotene and metaphase spermatocytes with highest expression seen in round spermatids and early elongating spermatids (Roumaud et al. 2018, Zhang et al. 2018). In common carp (*Cyprinus carpio),* Sox30 is expressed abundantly in female and male gonads especially in spermatocytes and spermatid/sperm of carp testis, and lower expression is seen in brain, liver, muscle and kidney (Anitha and Senthilkumaran 2020). Studies on *Sox30* null mice indicated them to be phenotypically normal but the male *Sox30* null mice were infertile. Histological analysis showed mature spermatozoa to be absent in epididymis and testes of adult males. Germ cell differentiation was found to be interrupted at the post-meiotic, step 3 round spermatid stage of spermiogenesis in *Sox30* KO mice with formation of multinucleated giant cells, abnormal acrosome and axoneme development, and absence of mature spermatids (Feng et al. 2017, Bai et al. 2018). Sox30 regulates expression of post-meiotic genes such as *Fhls, Tnp1, Tnp2, Prm1, Prm2, Prm3, Odf1, Catsper1, Smcp, Hils1, Ccdc54* and *Tsks* that are required for round spermatid maturation (Bai et al. 2018, Zhang et al. 2018). Human homologues of several *Sox30* testicular gene targets such as *TNP2, PRM1* and *PRM2* have mutations identified in them among sterile male patients (Zorrilla and Yatsenko 2013). Master transcriptional regulator of spermiogenic gene expression *Cremτ* and *Rfx2* are not regulated by Sox30 and vice versa (Feng et al. 2017). Sox30's expression is regulated by retinoic acid in mice spermatogonial stem cells (Wang et al. 2016). Mybl1 and dmrt1 expression is regulated by Sox30 in mice and fish, respectively, which are genes involved in male gonadal development (Zhang et al. 2018, Tang et al. 2019). In common carp, Sox30 regulates expression of steroidogenesis genes and testicular development transcription factors while Sox30's expression in turn is influenced by gonadotropins (Anitha and

Senthilkumaran 2020). *In vitro* studies indicate SOX30 to regulate *NR5A1* promoter (Sakai et al. 2008), a gene implicated in patients, with spermatogenic failure, of European ancestry (Röpke et al. 2013, Ferlin et al. 2015) and African ancestry (Bashamboo et al. 2010) but not in Indian infertile men (Sudhakar et al. 2018). Given these findings, epigenetic and genetic *SOX30* defects could underlie nonobstructive azoospermia (NOA) which is the most common type of male infertility in humans. *SOX30* promoter and CpG islands were found to be hypermethylated in testicular tissues of NOA patients wherein SOX30 expression was markedly reduced. Sox30 null mice are known to mimic the testicular size reduction and abnormal testicular pathobiology of NOA patients. However, no *SOX30* pathogenic variants have been found among the NOA patients, to date (Han et al. 2020). *SOX30* mutations have also not been identified among Sertoli cell-only syndrome (SCOS) with azoospermia in Japanese men studies, so far (Miyamoto et al. 2020).

Results

To investigate the role of *SOX30* in the genetic aetiology of male infertility, we examined genomic DNA of 494 male infertile patients of Indian ancestry with majority of them with a clinical diagnosis of nonobstructive azoospermia. Primer-sets generating eleven amplicons with an average size of 350bp were used to cover the five exonic regions of *SOX30* and the amplified products were Sanger-sequenced (Table S1.1). In addition to our control cohort of 496 healthy fertile individuals, allele frequencies were obtained from Exome Aggregation Consortium (ExAc), Exome Variant Server (EVS), dbSNP, 1000 Genomes, Trans-Omics for Precision Medicine (TOPMED) and Genome Aggregation Database (GnomAD). Population-specific allele frequencies of variants were obtained from GenomeAsia 100K (GA100K), The Indian Exome Reference Database (Index-DB) and IndiGenomes (IndiGen) databases.

Table S1.1. Primer sequences to amplify exons of *SOX30*

Primer name	Sequence (5' – 3')
SOX30-EX-1aF	CCTTGTGACGCAAGACTTCA
SOX30-EX-1aR	GCTCGGGTCTGGCTCTCT
SOX30-EX-1bF	CAAGGCTCTTAACCGAAAGG
SOX30-EX-1bR	GGGCCTGAACTGCAACAG
SOX30-EX-1bnF	GGTCGTTGTGATTGGGTGAG
SOX30-EX-1bnR	AACACCTGCTCTGGCTTCAC
SOX30-EX-1cF	GTGAAGCCAGAGCAGGTGTT
SOX30-EX-1cR	AGGAGTCTCTCGGGTTCCTC
SOX30-EX-1dF	CGGAGGAGGTCATGAGAGAC
SOX30-EX-1dR	CAGACTCTGCCCTGAAAACA
SOX30-EX-2F	AGACGGAGGTGCACTTTACC
SOX30-EX-2R	AAGGAAACTCATGCCCAGTG

Primer name	Sequence (5' – 3')
SOX30-EX-3F	TAATATCCCGGAGCTGGAAA
SOX30-EX-3R	TGTTCTCACTTTCCCTCTTCTTTT
SOX30-EX-4aF	GATTGATGAAGGCCTCTGCT
SOX30-EX-4aR	AAGTTGCAAATCTGGCATGG
SOX30-EX-4bF	TGCATTCTGAAGCCACTCAC
SOX30-EX-4bR	ACCTGGCCCTCTCCATCTAT
SOX30-EX-5aF	CAGTGGGAAAAGAGCAGAGG
SOX30-EX-5aR	GGTGACATTGACTTGCTGGA
SOX30-EX-5bF	ATAGCCACAGTGGGGAAGAA
SOX30-EX-5bR	CCCTTTCCCCTCACATACAA
SOX30-EX-5cF	GGCTGAAATTTGCATCAACA
SOX30-EX-5cR	CATGTTTAAGTCTCGTGTGAGC

Our study identified 15 rare heterozygous variants (minor allele frequency, MAF<0.5%), of which six were novel variants (Table S1.2). Of these, six were in 5'UTR, two in 3'UTR, one in intron 2, four were missense variants, and two, synonymous variants. These variants were unreported in the index-db database. Variants c.-318G>T and p.Glu744= were observed in more than one infertile patient and absent in in-house controls while variant c.-57G>A was identified in 4 patients and in 3 controls. Among the remaining variants, except for p.Leu315=, rest were absent in the in-house controls.

Table S1.2. Heterozygous *SOX30* rare variants in patients with male infertility

Genomic location GRCh38.p12	mRNA change	Protein change	RSID	*Global MAF	GA100K		IndiGen	In-house control	Patient sample no.s	CADD Score v1.6
					All	India				
g.157652426G>A	c.-348C>T	NA	Novel	-	-	-	0.0005	0/992	1	5.33
g.157652396C>A	c.-318G>T	NA	rs369079492	UA	0.00173	0.0042	-	0/992	3	3.91
g.157652375_ 157652376delCT	c.-294_-293del	NA	Novel	-	-	-	-	0/992	1	15.12
g.157652286T>C	c.-208A>G	NA	rs1399073665	C=0.00003/1 (GnomAD)	-	-	-	0/992	1	8.47
g.157652135C>T	c.-57G>A	NA	rs369841660	T=0.001997/10 (1000G)	-	-	0.0049	3/992	4	15.37
g.157652134T>G	c.-56A>C	NA	rs866162797	G=0.000128/4 (GnomAD)	-	-	0.0010	0/992	1	9.276
g.157651712G>T	c.367C>A	p.Pro123Thr	rs182220520	T=0.001198/6 (1000G)	0.00058	0.0008	0.0010	0/992	1	19.78
g.157651678T>C	c.401A>G	p.His134Arg	Novel	-	-	-	-	0/992	1	16.04
g.157651169G>A	c.910C>T	p.Pro304Ser	Novel	-	-	-	-	0/992	1	25.3
g.157651134G>C	c.945C>G	p.Leu315=	rs755219644	C=0.000027/3 (ExAC)	-	-	-	1/992	1	8.40
g.157646834T>C	c.1208-18A>G	NA	Novel		0.00029	0	-	0/992	1	2.597
g.157638627C>T	c.1483G>A	p.Val495Met	rs138471751	T=0.00231/290 (TOPMED)	0.00029	0	-	0/992	1	23.2
g.157626370C>T	c.2232G>A	p.Glu744=	rs552463131	T=0.000149/18 (ExAC)	-	-	0.0005	0/992	2	20.7
g.157626105T>C	c.*235A>G	NA	Novel	-	-	-	-	0/992	1	8.32
g.157625842A>G	c.*498T>C	NA	rs146515023	G=0.001776 (TOPMED)	-	-	0.0010	0/992	1	14.33

*Highest MAF among databases represented

All the nonsynonymous variants had CADD score greater than 15 predicting these to be the top 3.33% of the most deleterious substitutions possible in the human genome and were not located in the HMG domain (Figure S1.1). Variants c.-57G>A and p.Pro123Thr have also been identified in the JME patient set, in one and two cases, respectively; and functional analysis using reporter luciferase assays have indicated altered transcriptional activities when compared to the wildtype.

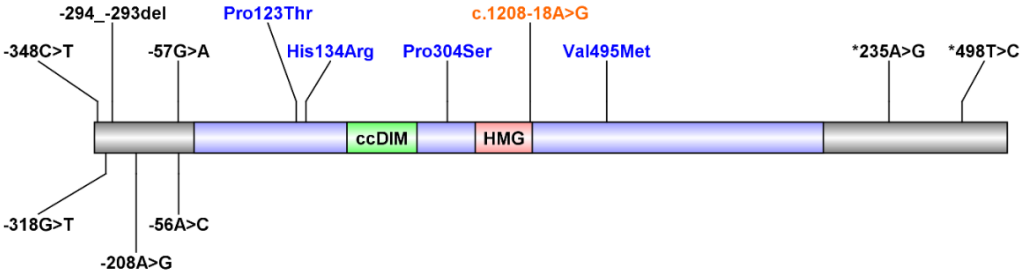


Figure S1.1. Schematic representation of *SOX30* mRNA with locations of the rare variants.

No homozygous and truncating rare variants were identified among these infertile patients. Variants did not cluster within any specific regions of *SOX30.* The electropherogram of the rare variants are represented in Figure 2. The 5'UTR rare variants are located in CpG islands, proximal enhancer region and in binding sites for transcription factors EGR1 and CEBPA (Table S1.3). Missense variants p.Pro123Thr and p.His134Arg are also located in CpG island region proximal enhancer region of *SOX30* and in binding sites for transcription factor CTCF since these variants were located in exon 1 (Table S1.4).

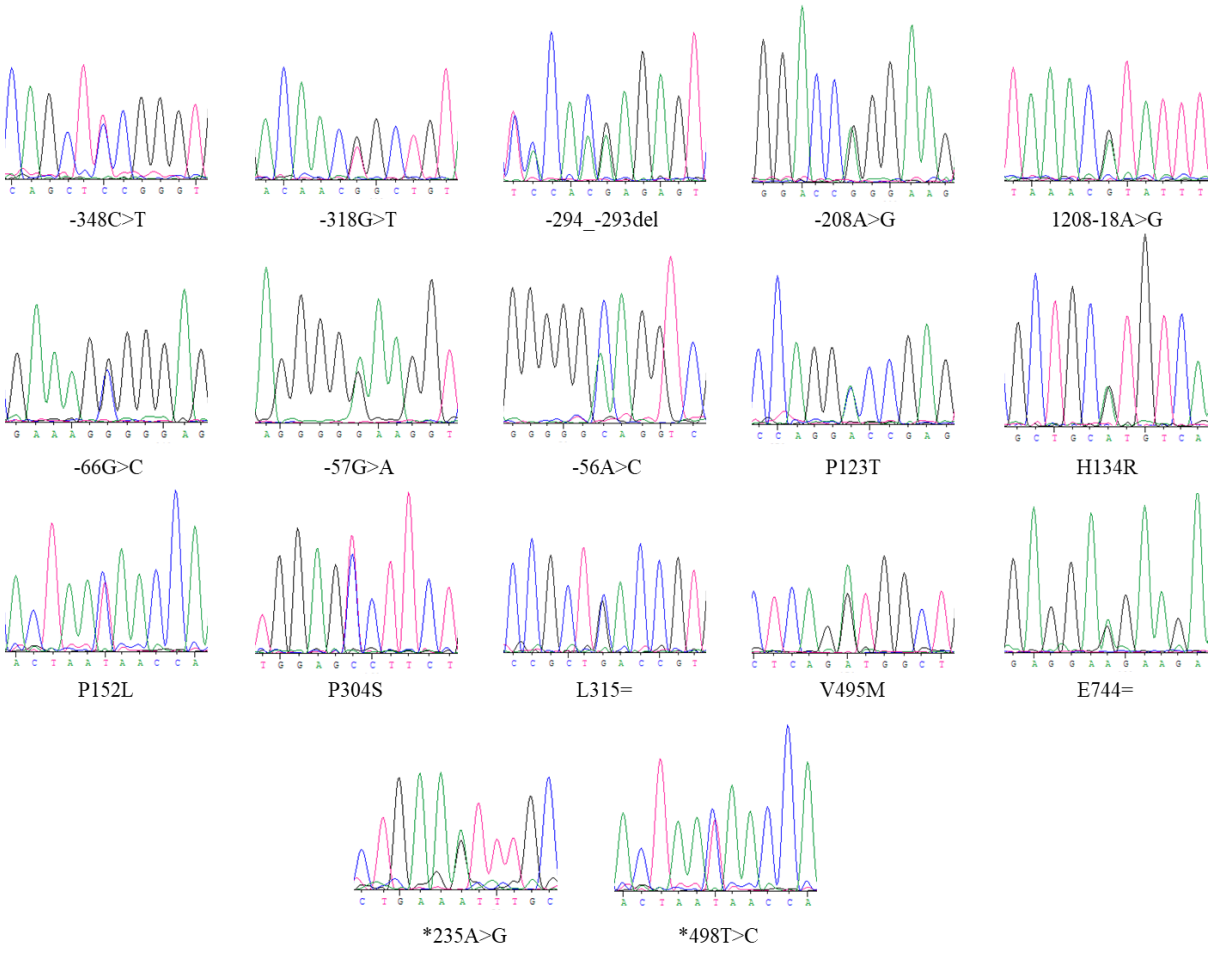


Figure S1.2. Electropherogram of heterozygous rare variants identified in the infertile patients examined.

Table S1.3. Conservation, sequence features and *in silico* predictions of the UTR and intronic variants.

mRNA change	Consequence	Conservation	*Sequence Features	Prediction
c.-348C>T	5' UTR	C - 7/10	Prom, EGR1, CEBPA	MT – polymorphism
c.-318G>T	5' UTR	G -7/10	Prom, EGR1, CEBPA	MT – disease causing
c.-294_-293del	5' UTR	AG -7/10	Prom, EGR1, CEBPA	MT- disease causing
c.-208A>G	5' UTR	A-2, G-5 /10	Prom, EGR1,	MT- polymorphism
c.-57G>A	5' UTR	G – 5/10	Prom, EGR1,	MT- disease causing
c.-56A>C	5' UTR	A – 5/10	Prom, EGR1,	MT- disease causing
c.1208-18A>G	Splice site	A – 3/10	Predicted Branch site region	HSF- Potentially altering splice site, cryptic donor insertion. MT – polymorphism, splicing affected
c.*235A>G	3' UTR	A – 7/10		MT- disease causing
c.*498T>C	3' UTR	T – 3/10	#Predicted miRNA binding site for hsa-miR-487a/b-5p,	MT- disease causing

The reference allele of -208A>G is poorly conserved hence, this is likely to be a polymorphism. Variants p.Pro304Ser and p.Val495Met, located at highly conserved residues are predicted to be pathogenic by at least 4 of the 5 bioinformatics tools. The intronic variant c.1208-18A>G was predicted to introduce a cryptic donor splice site while the synonymous variant p.Glu744= was predicted to introduce a cryptic acceptor site which could affect the isoform sequence. The 3'UTR rare variants are predicted to be damaging by MutationTaster and the variant c.*498T>C although poorly conserved was located at the predicted binding site of hsa-miR-487a/b-5p (Table S1.3, S1.4) that could dysregulate SOX30.

Table S1.4. Conservation, sequence features and *in silico* prediction of amino acid coding variants.

Protein change	Conservation	*Sequence Features	SIFT	Polyphen2	Mutation Taster	UMD Predictor	Human Splicing Finder
p.Pro123Thr	7/10	CpG, EnhP, CTCF BS	Damaging	Benign	Polymorphism	Polymorphism	No Impact
p.His134Arg	4/10	CpG, EnhP, CTCF BS	Damaging	Benign	Polymorphism	Polymorphism	No Impact
p.Pro304Ser	8/10	-	Damaging	Probably damaging	Disease causing	Pathogenic	Significant alteration of ESE / ESS motifs ratio (-6)
p.Leu315=	9/10	-	Tolerated	-	Disease causing	Polymorphism	No Impact
p.Val495Met	8/10	-	Damaging	Probably damaging	Disease causing	Polymorphism	Significant alteration of ESE / ESS motifs ratio (10)
p.Glu744=	9/10	-	Tolerated	-	Disease causing	Polymorphism	Activation of a cryptic Acceptor site

Conservation organisms: Human, Chimp, Monkey, Mouse, Rat, Dog, Cow, Chicken, Zebrafish, Frog
*Sequence features from UCSC genome browser, MT – MutationTaster, HSF – Human splicing finder, # miRDB, Brown font -Intronic/ Splicing variant, Green font – Synonymous variants, ESE – Exon splicing enhancer, ESS – Exon splicing silencer.

Nine common variants (MAF>0.5%) were also found among the infertile patients. Heterozygous variants c.-354G>A and p.Glu229Lys were absent among in-house controls and were rare in Indian population databases GA100k, IndiGEN and INDEX-db (Table S1.5). Multiple *SOX30* rare variants were not observed in the same patient sample while some rare variant carrying samples had common variants additionally present.

Table S1.5. Common variants identified in the infertile cases.

Genomic location GRCh38.p12	mRNA change	Protein change	RSID	Global MAF	Sample nos/ Zygosity	Inhouse control
g.157652432C>T	c.-354G>A	NA	rs143580325	T=0.008187/41 (1000G)	1 – Het	0/992
g.157652144C>G	c.-66G>C	NA	rs188639331	G=0.003594/18 (1000G)	3 – Het	5/992
g.157651916A>G	c.163T>C	p.Cys55Arg	rs184421438	G=0.011472/360 (GnomAD)	22 – Het, 2 – Hom	15/992
g.157651624G>A	c.455C>T	p.Pro152Leu	rs13181859	T=0.00404/994 (GnomAD)	6 – Het	3/992
g.157651467C>G	c.612G>C	p.Pro204=	rs3749797	G=0.174312/21888 (TOPMED)	118 – Het, 7– Hom	47/992
g.157651394C>T	c.685G>A	p.Glu229Lys	rs41275269	T=0.012869/3220 (GnomAD)	12 – Het	2/992
g.157646739G>T	c.1285C>A	p.Gln429Lys	rs12188040	T=0.05651/283 (1000G)	54 – Het, 2 – Hom	29/992
g.157638298G>A	c.1812C>T	p.Phe604=	rs35793864	A=0.126597/14587 (ExAC)	101 – Het, 5 - Hom	77/992
g.157625718G>A	c.*622C>T	NA	rs17054773	A=0.095185/2981 (GnomAD)	27 - Het	29/992

Discussion

Nonobstructive azoospermia (NOA) is a severe form of male infertility where there is complete absence of mature spermatozoa in the semen mainly due to impaired spermatogenesis. It contributes up to 70% of all azoospermic cases , while azoospermia affects 10-20% of infertile men , and 1% of all men (Kasak and Laan 2020). Majority of NOA patients exhibit primary testicular failure due to intrinsic factors affecting sperm production while a smaller fraction exhibits secondary testicular failure that occurs due to endocrinal and developmental defects. Primary testicular failure can be due to both acquired factors such as radiation, chemotherapy, varicocele, orchitis, trauma etc. and congenital factors that are mainly genetic in nature. Genetic abnormalities such as abnormal karyotype, Y chromosome deletions, and Klinefelter's syndrome account for 17%, 2-10% and 15% of NOA, respectively, and are major contributors to NOA. Syndromic NOA have other clinical manifestations in addition to azoospermia. Although several genes have been identified to cause NOA, their frequency in the infertile population is limited and varies with populations studied. Very few single gene mutations causing NOA exclusively have been identified (Table S1.5; Kasak and Laan 2020). In most

patients NOA remain genetically idiopathic. (Kasak and Laan 2020, Peña, Kohn and Herati 2020).

Table S1.5. Non-syndromic monogenic exclusive NOA genes.

Gene	Function	NOA phenotype	Inheritance mode	Expression	Variation	Mouse model
**TEX11*	Chromosome synapsis and formation of crossovers	MA, testis atrophy	XLR	Pancreas and testis enriched	LoF, missense	Yes
**TEX14*	Formation of meiotic intercellularbridges	MA, SCOS	AR	Testis enriched	LoF, missense	Yes
**TEX15*	Chromosome, synapsis, DNA DSB repair	MA	AR	Endometrium, smooth muscle, testis	LoF	Yes
CCDC155	Homologue pairing in meioticprophase	MA	AR	Testis enriched	Missense	Yes
NANOS2	Spermatogonial stem cell maintenance	SCOS	AD	Testis enriched	Missense	Yes
SPINK2	Inhibitor of acrosin	Post meiotic block	AR	Epididymis enriched	LoF	Yes
SPO11	Initiation of DSBs	MA	AR	Testis enriched	Missense	Yes
TAF4B	Transcriptional coactivator	SF	AR	Mixed	LoF	Yes
TDRD9	Repression oftransposable elements duringmeiosis	MA	AR	Testis, parathyroid enriched	LoF	Yes
WNK3	Regulation of electrolyte homeostasis, cell signalling,survival and proliferation	SCOS	XLR	Epididymis, testis enriched	Missense	No
ZMYND15	Transcriptional repressor	MA	AR	Testis, parathyroid enriched	LoF	Yes

AD- autosomal dominant, AR- autosomal recessive, XLR- X-linked recessive, DSB-double-stranded break,LoF- loss-of-function, MA- maturation arrest, SCOS- Sertoli cell-only syndrome, SF- Spermatogenic failure, * Established NOA genes.

The spermatogenic maturation arrest phenotype leading to male infertility in Sox30-deficient mice makes SOX30 an excellent candidate gene for nonobstructive human azoospermia. In this study, 15 rare (minor allele frequency ≤0.05%) heterozygous variants spanning the entire gene were identified in male infertile patients of Indian ancestry. Most of these variants were not identified in ethnically matched control individuals. Two missense variants, p.Pro304Ser and p.Val495Met were in highly conserved residues flanking the HMG domain, were absent in in-house controls as well as in Indian population databases and were predicted to be pathogenic by in silico prediction tools (Table S1.4). An intronic variant located 18bp upstream of Exon 3 in the predicted branch point sequence is expected to introduce a cryptic donor site that could alter open reading frame of SOX30. Several 5'UTR variants are located in SOX30's promoter region that could affect the binding of factors regulating SOX30's expression. A synonymous variant in exon 5 (p.Glu744)= is predicted to introduce a cryptic acceptor splice site. No homozygous rare SOX30 variants were known, although recently ClinVar database has reported a homozygous SOX30 variant p.Ser517Cys in an infertile

patient of Indian origin. It is interpreted to be likely pathogenic. No rare variants were identified in SOX30 among NOA and SCOS patients of Chinese and Japanese descent, respectively (Han et al. 2020, Miyamoto et al. 2020). The heterozygous variants identified in this study may be dominant or dominant negative in their effect because haploinsufficiency does not lead to infertility in mice. SOX30 is a mutation tolerant gene {RVIS - 1.29 (93.85%), missense Z score of 0.776 in the GnomAD database} wherein variations are frequently observed in the population, while it is intolerant to loss- of- function variants (pLI – 0.998). Hence, identification of several rare variants and absence of truncating variants in not an unexpected finding. Variants c.-57G>A and p.Pro123Thr were also identified in patients with JME. The p.Pro123Thr was identified in two female JME patients while -57G>A was identified to segregate with JME in the family NIH34 amongst which it was present in 3 males, one of whose fertility was confirmed. These variants exhibited altered in vitro transcriptional activity when compared to wildtype protein (Chapter 2, pg: 46-47, and chapter 3pg: 68-69) probably suggesting effect of a modifier gene or other genetic and environmental interactions. Further genetic studies are needed to clarify the underlying mechanisms of phenotypic manifestations of SOX30-associated human disorders.

Other SOX molecules involved in spermatogenesis and male fertility are group D (Sox5 and Sox6), group E (Sox8 and Sox9) and group F (Sox17) genes. Sox5 and Sox6 are highly expressed during spermatogenesis, but their functions are unknown (Jiang et al. 2013). Mutations in SOX5 and SOX6 are reported in patients with neurodevelopmental disorders (Angelozzi and Lefebvre 2019, Tolchin et al. 2020, Zawerton et al. 2020). Sox8 and Sox9 are involved in maintenance of postnatal fertility and Sertoli cell functions while Sox17 is involved in transactivation of genes in premeiotic germ cells (Jiang et al. 2013). Knockout mice models of Sox8 and Sox9 exhibit delayed postnatal male infertility while heterozygous missense mutation in these genes have been identified in azoospermic and 46XY disorders of sex development patients (Angelozzi and Lefebvre 2019, Portnoi et al. 2020, Zhang et al. 2020). Both SOX8 and SOX9 are not only associated with male infertility, but variants also lead to other disorders along the sex development spectrum such as sex reversal, primary ovarian insufficiency etc. (Angelozzi and Lefebvre 2019). SOX30 resembles SOX group D molecules in its association with both neurological (juvenile myoclonic epilepsy) and reproductive (infertility) disorders, although they are phylogenetically distant within the protein family.

Our findings suggest that SOX30 could be an important risk factor underlying male infertility. Although certain missense variants p.Pro304Ser and p.Val495Met have been predicted to be deleterious by computational methods, the interpretation of the actual functional significance of the variants is challenging. Further functional analysis in cellular and animal models shall shed light on the mechanism/s behind these contribution of these variants to the phenotype.

Appendix II for Chapter 2

Table A2.1. Primer sequences used to sequence the genes or gene regions not covered in the candidate gene study of the 5q34 locus.

Amplicon	Forward primer 5' -3'	Reverse primer 5' -3'
KIF4B-Ex1	CACTGTTGGTATATAGGAAACC	GACTTCTTCCTGCTCAGTAC
KIF4B-Ex2	ACTCAGGTGGTGGTTGGTAC	AGGCCACAGTCCTAGAGTTG
KIF4B-Ex3	AGATTGTGGGACTCACTGAG	GTCAGCATAGCGAAGGGTAC
KIF4B-Ex4	CTGCTGCAAGATTCTCTAGG	CTGGGTAATCAGTTGCTGC
KIF4B-Ex5	GGATCTTCAAAAGCTAGTGG	GATCAGCTATTTGACCCTCC
KIF4B-Ex6	ACCAAGCCAAGCTGAGTGAG	GGCTCAGGATGTGGTTCAAC
KIF4B-Ex7	AGGAAGCCAAACGCCATCTG	CTCTGCCATTTGGCTTTCC
KIF4B-Ex8	TAGAGACAGAGTTACAAGCTG	CTGCACCCACACTGCTTG
KIF4B-Ex9	CAAAATTAGTCAAGGTGTCCAGG	TGGAAAGCATCTCCCAAGCC
KIF4B-Ex10	CAACACCAGCTTCTTCTCTGGC	GAAACACTTACAATTCAGAGGGGGTC
KIF4B-Ex11	CTCTGCTTAATGGAGGAGCC	CTTTGTTAATGGCCAGGTGC
CCNJL-5UTR	CAGCGAAGGTGAGAGAGAAGC	CGATATGGAGCTGTGCAACAAC
CCNJL-Ex1	GGGCGAGAGAGGATAGGGAC	TCAGACACTCCGGGAGAAGG
CCNJL-Ex2	GGCCCTGAGGTCCCATTTCC	GCTAACATCCTTATCACTTGGTC
CCNJL-Ex3	TGTGACCATTTGTGTCTAGC	GAAATGAAGGCTGGGCACTG
CCNJL-Ex4	CTGGCCTGTCTGCTTTCTGTC	CCCTTTCTTTGCTTACTGCCTGC
CCNJL-Ex5	ATTTGTCCATCCATTCTCAGGG	TGAGACCCTGTAGCTGGAAG
CCNJL-Ex6	CCCACTTAGCAAGACAGGTAG	GCTCCCACTGAAGTAGCTGC
CCNJL-UTR3a	GTCCCTGCATCCCTTAGCATGC	CACTCTCCTGGCCTCTTATC
CCNJL-UTR3b	TGAGAAACGTCTGCCACGTG	CATATGCCTTCAAAAAGGAAGC
CCNJL-UTR3c	AATTGCCACACTGCCTGTCCC	CTAAACCAAAGAACATGCTGAGC
CCNJL-UTR3d	GCCAGTATTTAGGGATCTGC	CTTGAATCCAAACTGCTTGCAC
CCNJL-UTR3e	GATAAGACTCTGCCACCGTG	CTCCCGAGGTCTTGCTGC
C5orf54-Ex-1	ATCCCACAGCCCTTTGCTAC	TGGGATCGTGGCTTTGTCAC
C5orf54-Ex-2	CAACCTTGTAACAGCTCAGC	CCTTCAAACCAATCCTTAGC
C5orf54-Ex-3a	GCTGCAGGCTGTATCTTGAC	CTGATAAGGGTACCTGCTG
C5orf54-Ex-3b	GAATCCTCATACAGTAGCTG	ATTTATTACCCTCACCACAG
C5orf54-Ex-3c	CCGAGTTTGTTGCCTACGTG	CAGCTGCAATGAATATGCCTTC
C5orf54-Ex-3d	GTATCAGCTAGAGAGAAGTTATC	GACTTCTGCTGTAGCTGTAGC
C5orf54-Ex-3e	TGAGTGATGATATCCGTGTGGC	CAAAGACGGGCGAACTCAATC
C5orf54-Ex-3f	CCTAAGCACTAGAATATCCAGC	TAGGAAAAATTGAACCCAGAGG
C5orf52-5UTRa	GGTGTTGGAGATACTGAAATG	ACTGGGTCTCACTTCCCTC
C5orf52-5UTRb	ACTTGTTCAGAGTCATAAAGC	TGAGCTGAGCAGACCTTTTC

Amplicon	Forward primer 5' -3'	Reverse primer 5' -3'
C5orf52-Ex1	CCACGTCGAAGCGGAAAAGG	CAAAGAGTTAGGGTTCCCTC
C5orf52-Ex2	CAGAGCTGCACTTGTCCCAAG	AATGGGGCCTGCCTTAGAGC
C5orf52-Ex3	AGTAGCACAGATGTCTGCCC	GCCTTAAGCATGGAAGCTAGC
RNF145-1	CAAGACAAACACACAAACACC	AGCCTAGCCAAGCCCATAC
RNF145-2	CTGTATGGGCTTGGCTAGGC	AGAAGAGGAAGGAGCCTGCG
RNF145-3	TGCTGCTACTGCTGCTGCGC	CACTCCCACTCCTTGCTAGC
RNF145-4	CCTTAAGTTGAAAGATGGCC	AGAATGACACGTGACCCAAC
RNF145-5	TTCGTGAAAGGGTATCTAAAAGC	TCATGTGCAATGATAAGACAAAAC
RNF145-6	GAAAGTACTGAAACCTATTTGG	GCTTTACTTAACAGTGTTCATG
RNF145-7	ATGGAAAGGGGTATCATTAAC	TATGTGGGAAAAGTCCCCATG
RNF145-8	TTTTGTGGATGCCAAATACTC	GAACAGCAGCATCAGGGTTC
RNF145-9	GAATTTAGCTGAAACAGTCTTGC	GCCAAACCTAGGTATCCAGTG
RNF145-10	GAGTAATTGGGTGATTTCTCC	CCCAAATGTTGCAACCATTC
RNF145-11	ATGTAACAAGAGTCTAAGCC	AAGTCAGACATCTCATATAGG
RNF145-12	CCAGAGATGTTAGCTGCTG	GCCATCTAACTCCTGTGCTC
RNF145-13	GCACCCTAAAAACATCAGGAATG	CTGAGGTCCATTTTTTTTAATCC
RNF145-14	GAAAGAGATGTTAGAACCAGAAC	GCAGTTTCTGAAGTTGCTGAGG
RNF145-15	GGATCTAATCCAGGAAGTAC	CTGAAGACTCCGTCATAAAGTTC
RNF145-16	CAAAGATTGAAGCACGCTCATAC	CTTTTTGAGCCAAGCCTATTG
RNF145-17	CATTCATCATGGCTACAGGTAG	GCACAGGTAAGTTTGCTAACAC
ATP10B-1	GACTCGGCCTCACATTCACAG	GTCATATCCTGGGCCTAATTC
ATP10B-2	CTTTTCTGTCCTGTGTCGTGC	CTGCCCAAGAAGTAAAGATAG
ATP10B-3	ATTCCTCTCCTTCTCCAAGCA	TCCATCATCCTTAGCTCTGG
ATP10B-4	ATTTCTTTGCCACCTGGGCAG	TTCAACTGATCTCCAGCTTCC
ATP10B-5	TGAACAGGCCAACATGCTGC	GAGATTGCACGACTATACTCC
ATP10B-6	CAAGGCCAAACAGTAGGGAAG	CGAAGAAAAGAGACTCATCCTG
ATP10B-7	GTTCTGAGGAAACTTAAAGCTC	GGAAAAGGAGGCATCTACTAC
ATP10B-8	GTGACAAAGCATGGATTACC	CCCTCTTTATCACACACACAC
ATP10B-9	CCACTTCTAGTTCTGTGACAG	GTTGAGAAATGGGAAGTTACATG
ATP10B-10	TTATCTTAAGCTCCTGTGAGAAG	GTCTTGTTTTCATTCCATGCTAC
ATP10B-11	GACCAACTTTAGGACTTAGTAG	AAAGCGATGACCATCTCCTC
ATP10B-12	CTATGAAATGCCTTTGCTACC	GAAGCAACTTGTGGAAATGCC
ATP10B-13	TTACCACCCCCCACACAAAG	GAAAGCTCTTCAGGGTTTGTAC

Amplicon	Forward primer 5' -3'	Reverse primer 5' -3'
ATP10B-14	CTTTCCCCAGGCATGTGAAG	CTGTTGCTAGGACCCAAATTTC
ATP10B-15	GAGAGTAATATTCCACTTTCCTTC	CTATAGAACTCTCTGTGCCTGG
ATP10B-16	GAGCATTTGTATGTGGCCTCATG	GTGGCAAGTCAGAACAGCTAATC
ATP10B-17	GTCCACCCAAGATTGTTAAGG	AGGAAGCTGTGTGTGGTGTC
ATP10B-18	TAAACATTGCCTGGCTTCAC	ATATTGGGCTGGATGACTTCC
ATP10B-19	GCATTTCCAAGGTTGCTTAC	ACCTCTGGTCCAGCCTTTC
ATP10B-20	GTATGTAACTGGGGACACTTTAG	TGGATGTGGGTGGAAGTTCTC
ATP10B-21	GGCAGCCATGGATGGAAGC	GAGGTGCTTTGCCCACTGC
ATP10B-22	AAATGAGGCAACTGCCACTCC	ACAGTGGTCAGAATGCAGGC

Amplicon	Forward primer 5' -3'	Reverse primer 5' -3'
ATP10B-23	TCTGGTTGTCTGTCGCAGCC	GAAGGACACTACTGGTCTGG
ATP10B-24	CCTCCTTCTCCATTGACAAC	TAGGACATAGGGTAGGAGC
ATP10B-25	TGATCACCCTGCCTCCTAATC	TGCCAGCAGAAGTCAAACTCC
ATP10B-26	GGAGTTGGATTCTAAAGACTATC	CTGGCACATTGTTTCCTCTATG
ATP10B-27	AGGATAATGGCCTACTCAAG	GACACTGAGTTCAAATCCAC
ATP10B-28	TGAATCGGGGTTAACATCCC	CTATATAGGTACTAGCTTAAGG
ATP10B-29	TCCAAAGAAGCCAGGTCTGC	GTGTTACAGCTTTCAGAGGC
ATP10B-30	GTTTGAATCCTTGCTCTCTAGG	GCATCTAGCAGGGGCTTAAC

Table A2.2. Primer sequences for *GABRA1*, *GABRA6*, *GABRG2* examined in NIH34-C1 and *GABRB2*,in NIH34-D1.

GABRG2-seq-Int1a	CGTTTTGTTCTGAAGAGG	TGCTGAACCATGTTTGC
GABRG2-seq-Int1b	GTTCCTGTGTACTAAGGC	CGAATAGTCAGCTAGGC
GABRG2-seq-Int1c	CCTGAGAATGTGTATGC	GAAACTCCAAAGTGAGG
GABRG2-seq-Int1d	GCTGAAAAGGATGGAGG	GGTTACTGGAGAGAAAATAC
GABRG2-seq-Int1e	CAAGACACAAGGGAATG	GCTACTACTGTCTACCAC
GABRG2-seq-Int1	GTTGATCCTGAGGAAGG	CAGAAAAGTATAGACCAC
GABRG2-seq-Int1g	CATTCACATAGCTTGTGC	CCTTTAGGGTATTTTTACTCTG
GABRG2-seq-Int1h	AAGAAGATATTGGCTCC	GCCAGTTAAAGAAACATG
GABRG2-seq-Int1i	GATGTTTCTTATGAGCTGC	GAACAGAGACCTCTATC
GABRG2-seq-Int1j	GGCAGAAGTTGAAAATTGGC	GAAGAGGTCACATAGGTGTC
GABRG2-seq-Int1k	GTCGGCACAATTCAAATGC	CTACCTGTTACACTTCTAGG
GABRG2-seq-Int1l	CTTGGGTTTTTAGATCTGTC	TGGAGGACAACATTGAGC
GABRG2-seq-Int1m	GCATAGTTTAATCTGCAC	GAAGTCCCTTAATCTCTC
GABRG2-seq-Int1n	GAAACGTTTGCATTCACAG	GGTGGAGAGTGTTGTATC
GABRG2-seq-Int1o	GGGAGGGTTATTTCCC	GAACACGTGAAAGCATC
GABRG2-seq-Int1p	ATGGCAAAAACAAGCTG	AAACAACCCTCAGTGCC
GABRG2-seq-Int1q	CAAATAACCCTCTTCCAC	CTAGGCATGCTATGAAAGG
GABRG2-seq-Int1r	GAAGCCAGGACCCACAAC	CCTGGGCTGAAGTGATCC
GABRG2-seq-Int1s	TGGGCAACATGGTGTGACC	GGGACCTCTACCCAATTTGC
GABRG2-seq-Int1t	GTGGGAGTGTAGAAGAGG	GTGAGTGAATTGGAAGGTAGC
GABRG2-seq-Int1u	GCTGGTATTTTGCTATTCAGG	GGGGATACTGGGTAAATGC
GABRG2-seq-Int1v	CCCAGCTTCTTTTGCTC	GCCACATTGGAGTGAGC
GABRG2-seq-Int1w	GTAGCTTGGTAATTGCAG	CAATGAGAAGAGCAATTTGC
GABRG2-seq-Int1x	GGCTGGAAGTCTGAGATC	CCAGCATTTTGGGAGGC
GABRG2-seq-Int1y	GCCCGGATAATTTTTTAC	CATGGAAACTGTAAATGG
GABRG2-seq-Int1z	CCTTTCAGTCACATTGC	GCAACATAGTGGGTAGG

GABRG2-seq-Int1aa	CTCTGATTCAGGTGACC	GTAATCTCAGCTACTCAGG
GABRG2-seq-Int1ab	CAAGGTAATTGTGTACC	CACACCACACTTGTAAG
GABRG2-seq-Int1ac	GATGTCTGGGGGAAAGC	CATTCCCCAGATCATATAGG
GABRG2-seq-Int1ad	GGATAAGGGGTAAGACTTC	GCCTGGATTTATACACAG
GABRG2-seq-Int1ae	AAGGGATTCTCATAGAGACC	GCTAAATGAACTGTATGTGTCTC
GABRG2-seq-Int1a	CAACTACAAAATCCTTTACTCC	CTAGAAAACTGAATGTGATC
GABRG2-seq-Int1ag	TACACAGTAATTCCATGCC	GAAGAGGGCTGTTTGTG
GABRG2-seq-Int1ah	CTTTTAGCCTTATTGTGTG	CAAGGTTTTATTGTTCGC
GABRG2-seq-Int1ai	TATAACGAGAGTAGGAGG	GTATGAAAATGAAGGCTGG
GABRG2-seq-Int1aj	CCTGTCACCATAGAGC	GAAAACAGTGCAGATGG
GABRG2-seq-Int1ak	TTTTCCTAGCTGTTCAGG	AAGTGTCACATAGGTACAG
GABRG2-seq-Int1al	AAGCTTGCTCTTGTTCTC	TCACTCTTATCCCTGGC
GABRG2-seq-Int1am	GCTGAATACCAAAGTCTTGG	CATAACCTGCATGACCTTGG
GABRG2-seq-Int1an	CGTAATTATGAACCCACTTC	TAAAGTCATGTCAACTGTCC
GABRG2-seq-Int1ao	CTATCCTACCCAGCATC	CACGAGTATGTTTATTGTAGC
GABRG2-seq-Int1ap	GTTTGATAAGGATAGTGGTC	GGAGGTTAAAAGCAGAGC
GABRG2-seq-Int1aq	GGGTGTCTTGCCAAATGC	GAAATATTGCTACATGCAGC
GABRG2-seq-Int1ar	GCTGTCCTAACTTTGAGC	GCAGTGACGATGCTCATTG
GABRG2-seq-Int1as	ATGATAGGGAGGTGTAATCC	GACTCAGTTCCTTAACTACG
GABRG2-seq-Int1at	GTGGATACATACAAAGAAGC	CTCTATGCATGAGTGTGC
GABRG2-seq-Int1au	TTCCTGCTCTAATTCTACC	GATGCCTTTAGCACAAGG
GABRG2-seq-Int1av	CTTCAGTGTGGGGAGATC	CATTTGTTGCCCACTCACC
GABRG2-seq-Int1aw	GAAGACAGACAATAGAGAAG	CTATCAAGACATACATGCG
GABRG2-seq-Int1ax	TAGAGGTGACAATGGAGC	CTGTTCAATGAGAGGTTGC
GABRG2-seq-Int1ay	ATTCCAGAAACTGCATGAC	CCCGGAATCAGAATTGC
GABRG2-seq-Int1az	TGCCACTGGTCTAACCAC	CACAGATAGGCTATCTAGG

GABRG2-seq-Int1ba	CAGACATTCTTATGGAGTC	GCCTAACACGCAGTACTG
GABRG2-seq-Int1bb	CATGTTTGAAGCTGGTG	GAGTCAAGTATAATGGTGC
GABRG2-seq-Int1bc	TTTCCCTCCTGGTCTGC	GCATTAATTTCACAGAGTGCC
GABRG2-seq-Int1bd	TTTCTGGCAGCACACTGGC	ATAACATTTCCCACCCCTG
GABRG2-seq-Int1be	GTGGCAATTATGTCATAGC	CTCTTTGCTTTTGCTTCCC
GABRG2-seq-Int1b	TACCAGGTTCTTAACAACC	CAAGTTGACTTTTATCCACAG
GABRG2-seq-Int2a	AGTGTAGTTCAAGAGCAAGG	AAAGCGAGTGTGCCAGTG
GABRG2-seq-Int2b	CAGGTAGCTAGGACAACC	GGTTTTGGATCACAGTTCC
GABRG2-seq-Int2c	GTGTGAGGATAATGTGGTC	CACATAGAGAGGTGCAAG
GABRG2-seq-Int3a	CGCTATCAATATGGTGAG	CTCACTTAGCAGAGCATC
GABRG2-seq-Int3b	GTAGAGCAGAATGGAGTAC	CAACCAATTAGTGATGTGC
GABRG2-seq-Int3c	CCAGAACTGTTCTAGG	CCTGGCAGAGAACAAAC
GABRG2-seq-Int3d	TTACTGAAGTCAACCGCTC	CTAGTTTCTTGCCATGAGG
GABRG2-seq-Int3e	TTTGACTGTGTCTGGAGGC	CTGTCATACCACGTTTGCG
GABRG2-seq-Int4a	TGGAATGATGGTCGAGTG	AGAGGCTGAACCATAGC
GABRG2-seq-Int4b	TTATCTTTCATCTCCTGGC	CTCAGTTGTGATGTCAAGG
GABRG2-seq-Int4c	GAAAGAAAGCACAACCATGC	ACTTGAACCTCTGGGAGG
GABRG2-seq-Int4d	GTCTCACTGTGTCACCC	CACAAACACACTCATAGCTC
GABRG2-seq-Int4e	GTGTCACTAATGGAGATTC	CGGAGCAATTACTGATG
GABRG2-seq-Int4	GGAAGCAATTCACTCTACC	GTTGCACAATGGTGTGGC
GABRG2-seq-Int4g	ATATGTCTGTGTATGTGC	CTGATTGAAAACACAGAG
GABRG2-seq-Int5a	CACCTACAGTCTTTCTG	TCATCATGGATCCTCTTG
GABRG2-seq-Int5b	GTGAACTACTACTTGACAC	TCTAAAGCACAGGGAAGC
GABRG2-seq-EX-new-6	CATCAATACCAGAGGAAACC	GATGGAGAGAGATATGATAG
GABRG2-seq-Int5c	TTGTAGAGATGGGGCGTCG	TTGAAGACATCACACTGGG
GABRG2-seq-Int5d	ACTTGTCCAGGGTAGAATGG	CTGAGAAAGATAGTTACAAGC
GABRG2-seq-Int6a	GGGAAAATTGGATTCAAATTGG	GACAGAGGGTTTTAAATCAAC
GABRG2-seq-Int6b	GATTATCAGATGTGCATG	ACTCAAGCATTACTATGACC
GABRG2-seq-Int6c	GAGCATTTGCAACTGTCTCC	CACTCTTGGGCTCAAAACAAC
GABRG2-seq-Int6d	CCTTAGAACTGAAGTCTTCC	GAGCCCGTTGAAATAGTAC
GABRG2-seq-Int6e	CAACTCCTTTGTGCATCC	CCAAGACACAAGTTTGCCTG
GABRG2-seq-Int6	GTGACTGTACTTCTGATTCC	GAGAATGTCTGGATTTGGTAC
GABRG2-seq-Int6g	CCCTAGACTTTGAATATGTTG	GTAGTTAGTGCTTAGTAAGTGC
GABRG2-seq-Int6h	CATCTGGAATGTGTCAGTAGC	GGTAGTGATGCATTAAGAAGTC
GABRG2-seq-Int6i	GTTTTCTACAATTTCCCAGC	TGTATATGTGCCCAGCCATC
GABRG2-seq-Int6j	AATGGGATGGCTGGGCAC	CTTTAAGTCAGCCACCAGG
GABRG2-seq-Int6k	CATCTCTAAGCCAGTCAC	ATTGTGATCTGCATGTCAC
GABRG2-seq-Int6l	TTCCTTTCCACCAGTGAC	TATTACGCTGCTTCAAGTAG
GABRG2-seq-Int6m	ACAGTCTCTTGTTACTCCTG	TCTCCCTGCTAGATAGTGTG
GABRG2-seq-Int6n	TCTGGAACATTTGTATGCTGTG	AGATTTGAATGGTCTTGTGC
GABRG2-seq-Int6o	GTGGTCATTACTACTCAC	TGAATGGCAGGTGAAAATGG
GABRG2-seq-Int6p	ACAGATTAGGGGCTGACTAG	GTAGTGGGCATGGAAGAAGA
GABRG2-seq-Int6q	TAACTACTCTTTTCAGTGCACC	CATGGGAATGAGGGCTTTC

GABRG2-seq-Int6r	GGTCTTTAGGATGGACAGAG	CATGAAAGTGGACAGTAGAG
GABRG2-seq-Int6s	CTCACCTTTTGACATTCTGC	GCTAACTTATCAACCATCTC
GABRG2-seq-Int6t	TTTACTTTGTCTGTCCCGTG	AAAGAAATAGGGCCAAGAG
GABRG2-seq-Int6u	AGACACTCATCTTTCCTGAC	GAAGACTCCATGAGATCTGG
GABRG2-seq-Int6v	AGGGAAGGCCTCTATAGG	CTGACAAGGAGATGATTCTGG
GABRG2-seq-Int6w	TTGTTGTACTGCTGGGACAG	AAGCGCGTAACACGGTGTC
GABRG2-seq-Int6x	GGTAATGATATGGTGATG	AATTTGCTTACTGGCACAG
GABRG2-seq-Int6y	CTAGAAGTAAAAAGATCTGGC	GGTTATCAGTTTTCAGGAGG
GABRG2-seq-Int6z	GTGCGTTCTCTTGTCTATATG	TGGAAGTAGGCAGCTTGC
GABRG2-seq-Int6aa	TAATTGTGACGGTTTCAGGG	GTAAACTGTGGCTAAGGTAG
GABRG2-seq-Int6ab	TAGACTTGGGAGGATTCTGG	GAGTCTGCCAACTGGTTATCC
GABRG2-seq-Int6ac	TGTCCACATACTTCTGCTGAG	ACTTTTGCCTTCAGTGACAC
GABRG2-seq-Int6ad	GCTAGCAGTCAAAGGGTC	CTGATGAAGCCATTTATCTAGAG
GABRG2-seq-Int6ae	CAGACCACCCATTATATTTC	GCCTAGAACATTATTAGAGTCC
GABRG2-seq-Int6a	TCTGTAAATCATCCTCCTG	AACCCTAGTTATTCCTCACC
GABRG2-seq-Int6ag	TCAAACGACTGGTCATCAGG	CTTCACTCCTCATGTTTCC
GABRG2-seq-Int6ah	ACAGAGAGAGGCTGGTAG	AAAGGATAGATGAGCAAGCAAC
GABRG2-seq-Int6ai	TCTGTTTGGTGCGAGTGTTC	CCTCTTCAATGGCATACTC
GABRG2-seq-Int6aj	TAAGAGATGTTGTTCCATAG	CTCGGAGAATTGAGTGTG
GABRG2-seq-Int6ak	ATCGTCCCCAGGAGTAATTC	GGGAAATGCCATTAGCCAC
GABRG2-seq-Int6al	ATCAAATCTCTGCACAGTGG	CTGTGCTCTCACTACGTCC
GABRG2-seq-Int6am	AGGAAACTGACGATCGAAGG	AAATGCTCATGTTTCCCTCC
GABRG2-seq-Int6an	AAGTGTGGTAAGACTGTTTG	TTTCCTATGCCAGTTTGGC
GABRG2-seq-Int6ao	AACACAGGAAGCCTTTTCG	GATCTCTGAAAACCCTATG
GABRG2-seq-Int6ap	CATAGGGTTTTCAGAGATC	AACGCCAAACCCAGAAACC
GABRG2-seq-Int6aq	AGAAAATGTCAGTGCCTGGG	GTACACATATTCAGCTGGAC
GABRG2-seq-Int6ar	AGACAAAGGACATTATGACTCC	TTGGAGTAGGTGTCTCACC
GABRG2-seq-Int6as	CCCTTCCAAACTTTGCAGC	GGTAAGTGCTAGATATGCTCC
GABRG2-seq-Int6at	CATCATTTCTATGTCAGAAGGG	CTTCTGTGACATTGGTAGGTG
GABRG2-seq-Int6au	TACTGGCTTTCATGGATGC	TAGGTCTTTGCTGATATCCC
GABRG2-seq-Int6av	AACTTGACTATGGACTCAG	TATGTCAAAGCAGAGTAATC
GABRG2-seq-Int6aw	AGAGGATAGAATTGGTTGG	TTTGCCGTTCTGATTTCATC
GABRG2-seq-Int6ax	GGTGGATATTCAGATGGAG	GCCTCAACTCTAATATCCC
GABRG2-seq-Int6ay	GAGTATAATCCAGTATTCACC	CCACATTTCAAGTGCGCG
GABRG2-seq-Int6az	CTCTAGAAAGAAAAGCTTGG	AATATCTCCTGCCAACCAG
GABRG2-seq-Int6ba	AGCAGTGACTTTCTTGGCC	TTTGAGTGTGGTGAGATATTTCC
GABRG2-seq-Int6bb	CTGATCCCTTCTGAAGTTGAG	CTGAAATTAGATCTGTACCTCAC
GABRG2-seq-Int6bc	GACAAAGCTTTTTCCACACTAAG	GTTACCAGTCATGAAAGCCT
GABRG2-seq-Int6bd	AATAAGAGTGACGAGAGAGG	GCCAGCAAGTCTACAGC
GABRG2-seq-Int6be	ACGTGGCCTTGGAGATATG	TCCCACAGCATATGCTTCTC
GABRG2-seq-Int6b	TAGGACAAAAGCAGGGACC	GTCATGATGCATAAGTTGC
GABRG2-seq-Int6bg	AACTGAGATGTGTTTCTC	TTTATCTGCATGAGTCTAC
GABRG2-seq-Int6bh	GACAAATGTTACTGAATGCC	CCATTCTTCTGCTCAGATC

GABRG2-seq-Int7a	CATCTTTAGGTGAGACACC	CGTGCAGGTTTGTTACATATG
GABRG2-seq-Int7b	GGAGGAAAAGCATTAGGAG	GAGCCGATTACTTCAAATG
GABRG2-seq-Int7c	GGCCTTCTTTTTTCTAGAG	CCAATAAGTTAGGTACAGG
GABRG2-seq-Int7d	TATGGGTTTGACATCACAC	CAATGTATTTGCTCCAGCC
GABRG2-seq-Int7e	CTTCTATGTCTGTCATCAG	CTGATGAGACATAGGTTTGC
GABRG2-seq-Int7	GGAAGTAGAGGTGCTGC	TGCTCAGAAAACAGCCAG
GABRG2-seq-Int7g	CCTTAAGCTTCTTGCCACG	GCCATGTAGAAGTAGACTC
GABRG2-seq-Int7h	GCCTCTCACTATTGATGTC	ATCTACAAACTGGGAGTGG
GABRG2-seq-Int7i	GTATTGGTACCCTATGCAC	CATGAAAGATCAAAGATGCTC
GABRG2-seq-Int7j	TGCTTGGCACAGTATCTGG	TCTCCCTAATCAGTTCAAGTC
GABRG2-seq-Int7k	ACCCAAGAGAACCACATGG	GCAGGAGAATCACTTGAAC
GABRG2-seq-Int7l	TGCAATGCCACGATTTCAGC	GACTAACTCTTTGAAGGTCC
GABRG2-seq-Int7m	AAAGATGCCTCCACTCAGG	GTGGGTTATGTGTAAGTGAC
GABRG2-seq-Int7n	GCAGAAGGAAGTTTTCAG	ACTCGTGAGACTACATTG
GABRG2-seq-Int8a	CCCTGTATGTATCATTTTCC	AGTTCGAGACCAGCCTG
GABRG2-seq-Int8b	GCAATTCTCCTGCCTCAG	AGTTTCCAGCCTGGATTGAC
GABRG2-seq-Int8c	GTGGCAATGAAGGGCAAAAC	CTTGGGTTTCTGCAGCTGC
GABRG2-seq-Int8d	TACCTATCCAACCCAGAGAC	GCCATTTCCATGACCCTG
GABRG2-seq-Int8e	CAAGTTCTGATGAGTCC	AACACAGAAGGAGAGTTC
GABRG2-seq-Int9a	CAGACTTCTAGAGTTGATTC	CATGTGAATGGAAATTGAG
GABRG2-seq-Int9b	CTGGTCTAGCAGGAAATTTG	CCTAGGCTGTTTTTCCAC
GABRA6-seq-Int2a	GCAACTTCTACTCAGAAAACG	GACAGCACCTGAAATTAGC
GABRA6-seq-Int3a	GGAGAGTGGAGTCCCAAAAC	CCTCAATGTAAAAGGTGGAGAGC
GABRA6-seq-Int3b	GGCAACATCATTATCAG	CTTGCCAACTTAAAAAAGC
GABRA6-seq-Int3c	CTTTAAAGAGGTTCACC	CGTATTATGGCAAGCAG
GABRA6-seq-Int3d	GCTTGTGTCTGTTTTTACAG	GATAAACACACTATAGCTGC
GABRA6-seq-Int3e	GCAGCTATAGTGTGTTTATC	GCTGGAAAGATAGTTTGC
GABRA6-seq-Int6a	TGTTTTGCAGTGATGAGTAGC	CCATCTTCCTTTGCAAGTGG
GABRA6-seq-Int7a	CCAGCAAGAACTGTTTTTGG	GTGGAACTAAATATGTAGGTTG
GABRA6-seq-Int7b	GGTTGTATACCCCTTTTATTAGG	GCAGTTTTCCTTCTTGAACTATGG
GABRA6-seq-Int7c	CCTAATAGTATAAATTCCCCAGG	TAGTTGTTTAGGGCTTGAGAG
GABRA6-seq-Int7d	TGTTAGCCAGGATAGAATGG	GTTAAAACAGTGGTGATCCC
GABRA6-seq-Int8a	GTAATGTGAGTTGAGCACAG	CCTCTTCTAACAGCTACTC
GABRA6-seq-Int8b	GTTAAATTTGCTGTGGC	GGATGCCTCTCATTTC
GABRA6-seq-Int8c	GGCACTTGTTGACTGAAAAG	GGAGTTCAGCGGTGTGATC
GABRA6-seq-Int8d	GCAGTTCTAGCTATTCAGG	GCTGAATCCTGATAAGCAC
GABRA6-seq-Int8e	GATTCTGTTACAAGGTATCCG	GCAAATGTCCTATGTAATCACC
GABRA6-seq-Int8	GCAACACATAGGGAAATTTTG	GCTTTAAGGTGCCTTTTTGG
GABRA6-seq-Int8g	CTTACAGTGAATGTGGG	GCTCAAAGGATTGCAG
GABRA6-seq-Int8h	GCATAGTTTGCTAGTTTCC	GCTTAGACATCACTCTGC
GABRA6-seq-Int8i	GTGAGAGAGATCAATGAGCC	CCTACAACCCTCGTCTAATC
GABRA6-seq-Int8j	TTTGAGTACTGTCATTCCTC	GAGAGGTAACCTATCACC
GABRA6-seq-Int8k	GGTTGATCCCTTCAGTAAGG	CAGTTTAGAAGGAACACGC

GABRA6-seq-Int8l	GGAAGGGACTGAAGCAAAC	GATAGATGAGCAAACGTAGGC
GABRA6-seq-Int8m	GCCCAAAACGCTAAAGC	GCCTTATGCATTTTTCTGC
GABRA6-seq-Int8o	GCCTTCTACCATTAAAGAG	CAAGGAGAGTTGGAAATCC
GABRA6-seq-Int8p	GCTGAGAGAGAATTAGAGC	GAGTAGAAAGAAGCCAGCC
GABRA6-seq-Int8q	GCTTTATTCCAAACAACTAGGC	GCTTGTAATCCCAGCTTTGG
GABRA6-seq-Int8r	GCAGTCTTGCTCTGTCCC	ACCTGCCGATGTGGTGAC
GABRA6-seq-Int8s	GTGAATTTAAGCTCTCTGC	CAGTGCCTGTTTCATTCC
GABRA6-seq-Int8t	CTAGGCCCAATACTTTAG	CTTTGCACACTCATAAGC
GABRA6-seq-Int8u	CTTTGTCACTGGCCTTTGGC	CAGATGATATTTGGAGTCAGG
GABRA6-seq-3'UTRa	GCTGACACTTCCAAAGG	CTTCACTTCTTCACACC
GABRA1-seq01	GACAAACAGTTGCCTCCAAAG	CCCAAATTTGCTTCCCAATA
GABRA1-seq02	CCTTAGGTAAGTGCGACTTTGG	AAGCTGCAGTCTTACCACTTTC
GABRA1-seq03	CCATTCTGAACACGAAAACTCA	ACAAAATGCTTTCTGCAGTCTC
GABRA1-seq04	AGCTATGACGAGGCTCAGAGTT	TCTGTTCCTGTGACTGCTTTGT
GABRA1-seq05	GGCATGTTTGTGTGCATTTG	ATAACCACCAGGGCTGACTG
GABRA1-seq06	CATTTATTTTATCGGGGGAAGA	GGGTCACAGGAATAGGCTACAC
GABRA1-seq07	ATGATCCAATCAAGCATGTGAA	AAAGAGACCACTGCCTTTTCC
GABRA1-seq08	CAGAACTGGTCTTGGGGAAA	TTAATTTCGTATCGGCAAAATG
GABRA1-seq-8aN	CGTAAAATAAAGATTTCCATTCCA	CCAGTATGTCTCATTGGCAGAA
GABRA1-seq-8bN	GGCAGTTGATTTTGTGACAATG	TTTGACAGCAGCAAGATATATGA
GABRA1-seq-9aN	TGGCATCAATGTTGAAATTAGG	GCATGACAACTCCAAAATCCT
GABRA1-seq-9bN	TGCTTTTGTACTCAAGGTCAGA	GGCACACCTATTTGTGAGCAA
GABRA1-seq-9	AATAATAGCCATTGCATTTTGC	TGTTTTGTTTAAGAAGGCACACC
GABRA1-seq-10	TTTCAGACTAGAATTTGAAGAAAGC	AAAACCTCTTTCCAAATAAATCCTTT
GABRA1-seq-10n	CATCTCCAACTAGCAACGTTTA	TTTTTCACAAGGTAAAAAGGAATG
GABRA1-seq-11	TCATTTCTTACAATTTTCCTTTCA	GAGATAGCTGCAGTCATTGTGC
GABRA1-seq-12	AGCACAGTACTTTCTGTGTAACGA	TGGCTGTCTTAATATTGCCATT
GABRA1-seq-13	TTCAAAGTCACACAGAAAACCAA	GCCTGGCTATTTCTTCTTTTCA
GABRA1-seq-14	TTTGCGATAAATTACAAAACTAGC	TTTTCTGGGATTAACTGGATGA
GABRA1-seq-15	CTGCCACACTGCTATTTTTG	GGGATCTACTGGGATGCATGT
GABRA1-seq-16	AAGTTGACATTATTGTGGCTTTTT	ATGGTTGCTCTTATCCCTGAAA
GABRA1-seq-17aN	TGATAGGAATCAGTTTGCAAATAC	ACACCTTGCCACCAAATAAAGT
GABRA1-seq-17bN	TGCAATGTCAATGAAGACACTTT	TAATACGGGCAAAGAGGAGAAA
GABRA1-seq-17	TGATAGGAATCAGTTTGCAAATAC	ACATCCAGGCTTCCTTCATTTA
GABRA1-seq-18	GAGCCTGGGGAGACTCAAG	GGGAGAGAGAGGATGAAAAGG
GABRA1-seq-19	TGGAAAACAAATTAGGGCAGA	AGATCTCTTGCAACAATTAGGC
GABRA1-seq-20	TGCCTCAGAGCAGACTTTCTC	CCACAGAGAAAACGACAGCA
GABRA1-seq-21	TGTTGTTGTTTTTGCTGTGTTT	GGATATCCATCACCTCAAGCA
GABRA1-seq-22	GGGGATGGTTAATGGGTACA	CTCAGGCACCCTATTTCACA
GABRA1-seq-23	TTCTCGTATAACAAAATGCTACCC	AGGCACCACATGGAGACTAAAT
GABRA1-seq-24	TGTTTATGTAGGAGCCCAGGAT	TGAGACTTTAAAGATGGGAAAAA
GABRA1-seq-25	TCAGATCCTTTGCCTCCTTC	CCCCTAACTCCCAGACAACA
GABRA1-seq-26	TTTGAATTATCGCTTCAAAACA	AATGCCACCATCCACATAAATC

GABRA1-seq-27	CCAGCTCCAAAATTGGGAAT	TGCGCACATGTACCCTAAAA
GABRA1-seq-28	CGCATTGTGCAGGTTAGTTACA	GGGTGAAAACGAAAAATATGAC
GABRA1-seq-29	TGCACTGCATTTTTCAGAGCTA	CTCACCACCACATACTGACACC
GABRA1-seq-2aN	TCCCCGGTCTTAAGAGATCC	CCAAAACAACATGCCACCAC
GABRA1-seq-2bN	TTTTCCCGCTGTCTTTCTTC	AAGCTGCAGTCTTACCACTTTC
GABRA1-seq-30	AGGGAGCAACCTGAAAAGTCT	GGGCTTTTACTCTGCCATTC
GABRA1-seq-31	TTGAAATGCCTCATGTTAGCAC	CATGAGTCCTCAGAGAGTTGTCA
GABRA1-seq-32	TTATGGTCAGTCTGTAGACCACTT	GGCCGAAAGCTAAAAGGTATTT
GABRA1-seq-33aN	AGACCCAAACGCTTTATGAAAT	CCATGCCTTCTACCCTATTGTC
GABRA1-seq-33	AAATCAAGACCCAAACGCTTTA	TCCTTTTCTCTTGTCTTTTCCA
GABRA1-seq-34	CCTTTGCTGGGGTGTACATTAT	GCAATTGAAACATTGTCTTGTTC
GABRA1-seq-35aN	GATACACCTTCAAACCTGTGATT	AGCTAAAAGGTTTGCACAAGGA
GABRA1-seq-35	AATAAGGAACAGCGTGTCCAAC	GTAGGTCCCTGGAGAATGTGAG
GABRA1-seq-36	CAGGCAGGAACACTTCAATACA	CCAAAGTGCTGGGATTATAGGA
GABRA1-seq-37	AATTACCTAAGGCCACTCAGCA	TTCATTTCCATTTTCTGCACTG
GABRA1-seq-38	TCTCAGAAATGGGAGACCAAAT	GCTTGGGCACTGTAAGAAAAAC
GABRA1-seq-39	TATCATCCACAGTCTCCCTTCC	TTCACAGCCAGGGAATCTATTT
GABRA1-seq-40	AGCTGGTAAGTCACATTGCTTG	TTGGAAAATGACGAAGAGTTCA
GABRA1-seq-41	GCTGTACACCATGAGGTAAGGA	TGCTGCTCTCAAGAATGAGTTC
GABRA1-seq-42	TGAAGGAAACATGAATCAAATAAA	TTTTAGCAAGCATATTTACCCATA
GABRA1-seq-43aN	TGCATAGAAGAAGGTAGGAAAAGC	TCACAAAGCCTGAGGACACA
GABRA1-seq-43	TTTTCCTTTGCTAATGTGTTTTT	GAAGGCATAATTGCAGGGTTA
GABRA1-seq-44	TTGCCCACTAAAATTTGGAAG	TCACTCCATGCATGTAAACCA
GABRA1-seq-45aN	GGGAGAAGGGAGTTATCACCA	CAGCATCAGTCCTGGAGAGAG
GABRA1-seq-45	ACTTCCTGGGTTTCAGCTGTT	AAAGCAGAACAATTGCCTGAG
GABRA1-seq-46	TCGGTTTATGCCTAATATACTGGT	GGGAAAGCTTTTATCTTTTTCG
GABRA1-seq-47	CAAGATGGCTGCTGTAACTCC	CATGGTGTTCTTTTCCGCTAA
GABRA1-seq-48	AAGACCTGTGGGGAATGAAAC	TGAGCCACATGAACAAAATGA
GABRA1-seq-49	AATGCATTGCAAACACACTGA	GGGTAGATTCACCTGATGCAA
GABRA1-seq-50	TGCTGATGTCTCTGGTCAAGT	GGCAGAGCACACTAGTTTGAGA
GABRA1-seq-51	TCTTCTTAGGTAAGATGCTTCCA	TGTAGCAACCTGGTCATGAAA
GABRA1-seq-52	GGGTCAAACAGAGGTGCATTA	GGTTAGAGGCAGTAGCTTTGG
GABRA1-seq-53	CATTTATCATGACCTTCGCATT	AGAGGGCAGGCTAGGAAAGAT
GABRA1-seq-53N	ACTCTTCCCCAAGATTTTCACA	TTCACTCTTCCAACAAATGGTTA
GABRA1-seq-54	TCAGTTATGGAGGAAGAAGAAGA	TTTAAAGATGTGGGTCCAGCTT
GABRA1-seq-55	TGATCACTCCATGTGGGAATTA	TCCAACTGAACAAATGACTTGC
GABRA1-seq-56	TCCAAGTGGCATAGTGGAAAG	TGTTTCCCAACATTCTCCAAC
GABRA1-seq-57	GGCAGACCAAATAATGCACAC	TGTGTGTTCTCTCTCACCATCA
GABRA1-seq-58	ATCTCTTTGTGGAGCCCTGAA	CTTCCCAGCTGTATGACCTTG
GABRA1-seq-59	TGGCACTGTTCTATTCTTCAGAT	GTTCAGGTGATCCTCCCATCTT
GABRA1-seq-59N	TCATGAGCTATTGTGAGATGGAA	TTGGAGTGCAGTGGTGTAGTCT
GABRA1-seq-60	AATGTGGAGAAACCCCATCTC	GGCTCTTTGGATGAATGATGA
GABRA1-seq-61aN	CCACTTGGTCATTCAACAAAA	TGACAGCTTCACAAGGAATCA

GABRA1-seq-61	ATGACCCAGAAGGGTCAAGA	CCTGAAGGAAGCTTGCAAAT
GABRA1-seq-62	TTGCTTTTCTTCCACTCTCCTT	GGACAAATGGAATGTTACCAGA
GABRA1-seq-63bN	AAGTGTTTAGTAAAGGGCAGGTG	ACCAGGTCCTCAGTGGAATAGA
GABRA1-seq-63	TTTGGATCACTTTCTTCTTTTCA	CAAGTACCAGGTCCTCAGTGG
GABRA1-seq-64	ACACGGCAAGAAAATAGCAGA	TTGATCTGGGAAACAGTGGTC
GABRA1-seq-65	TTTGGGTTGTGTGTCTACTGCT	TCAAGTGGAAATGAGTGGTCA
GABRA1-seq-66	TACGCAAGGAAGGGTATGGA	CAGAAGTGCACACATCAGGT
GABRA1-seq-67	GGGAGAGCCAATTAATGACC	CTTCGTCCTGGCTCCATCTA
GABRA1-seq-68	TGCTTTTATACTAGGCGCTATTC	CGGAGCAATGATTTGGTGTT
GABRA1-seq-69	TGCTACAAGTACCAGAAACGTG	AGGGTTCAAGCCATTCTCCT
GABRA1-seq-70	TGTAGTGGTGGGTGCCTGTA	GGCTTCTCACAGAGCTCAGAA
GABRA1-seq-71	TTGTACACATGTTTCTATGCTTTTT	GAACACAGATTTGTTGAACAGCTT
GABRA1-seq-72	TCCTCTTATGGAGCTTCAGTTG	ACTGCTGAATGACGTTGAGAAA
GABRA1-seq-73	GCTAAACACTTCATTGCATTACC	CAGGGTGGATCAGGCTACAT
GABRA1-seq-74	CAGTGGAGAGAATGAAAAACAA	GCAGTACCAACATCTGAACACC
GABRA1-seq-75	TCACTAGGGATGGATTTGCAG	CCCCAGTGTAGTCTGTGATTC
GABRA1-seq-76	TTTGGCCTGTGGTATTGGAT	CCATGAGGCATTTCTTGGTAA
GABRA1-seq-77	GGCAGGCACATTTTCTCTTT	TGCTGTGACCAGGAAATACAG
GABRA1-seq-78	TTTGTTGATGCCCACTTGAA	TTAGGGGTGTAGCTGGTTGC
GABRB2-seq-Int1a	TTCAGGAGGAAGGGAAGAG	GCAAGCTTTTCTTCACTACTC
GABRB2-seq-Int1b	CCTGCTATAAATAACCCGGAC	CCAAAACATCAAAGGGGAAGC
GABRB2-seq-Int3a	CAATGACCCTAGTAATATGTCG	TGCAGCCAGAGAGAGATTC
GABRB2-seq-Int3b	TCCCAGCCTCACAGAAATG	GACCAGAAAGAGTCAACTCTC
GABRB2-seq-Int4a	GAAATGAGCCAGCTATTGGG	TCATCGCATTGCATGTCTGC
GABRB2-seq-Int4b	TGTGTTCCCAACTAGGAAAC	CCCAAGAAACATTCTCTAACC
GABRB2-seq-Int4c	TATTGGATGTTTCAGAACC	TTCAACACTTGGTTCTAATC
GABRB2-seq-Int4d	TTAAGGCAAGGTGTCATCAC	GAGCCCAGTCTATCTGTG
GABRB2-seq-Int4e	ACAATGATAGAAAGACCCTG	GGACTACCACTTTCAATTTCC
GABRB2-seq-Int4	GATCAGTGAGGAAAGTGTG	TGAGCAGTAAGTCAGTAAG
GABRB2-seq-Int4g	AATGGAAAACCACAGCCCAG	AACCTCCTCTCATGAATCATG
GABRB2-seq-Int4h	GCAAAGTATTGAAAACCTTCTGG	TGAATATCCCAATAAAGCGAGTC
GABRB2-seq-Int4i	CTCCGTGAAGGCTCAGATG	AAGAGGAAGCTATTTCAGTCC
GABRB2-seq-Int4j	GTCGTGGTTAAATTGGCAGG	CAAATACCCAGACAAGTGAC
GABRB2-seq-Int4k	CAGGTAGGCGTTTACTATTC	TACAGGCTTTGGACAAATATTC
GABRB2-seq-Int4l	TGGTATTAGGGATTTGGAAGTC	TGTACTAACTCTCTGACTGC
GABRB2-seq-Int4m	AGGCAAAATGCAAAGACAGC	GAGAATATGTAAGTCTTCAGGC
GABRB2-seq-Int4n	ACAGAGTGAGAGTCATATATAG	GTTCTCATTCTATCAGTTCCAC
GABRB2-seq-Iot4o	CTTCCTGTGCAGAGACATC	ATCAACATCTTCAACTCCTC
GABRB2-seq-Int4p	TATTACATCTCTTGAGCAC	GAATATTTAATCCTCTCGGG
GABRB2-seq-Int4q	CAACAATCCTCTGACTTGAC	TCAGTGGTAGTAGTGGTATG
GABRB2-seq-Int4r	CATATCTTACACATTGCATTCAC	CCGATAGAATGTGAACAAAGAG
GABRB2-seq-Int4s	GGTACATTTAGCCCATTGGC	TGAACGCCATCTCGGACC
GABRB2-seq-Int4t	ACACCAGAATCCGAGTCAG	CAGGGTGAATATGAGATGAAAAC

GABRB2-seq-Int4u	GGTAAGCAGCGTTGTTGAG	AACCTAAAAATTGGCACCTGC
GABRB2-seq-Int4v	AATTGTCAGGTTTCTGTGAGG	TGCCAGGAAATCCTACTCTG
GABRB2-seq-Int4w	CTTAGACTGCAATTTTGAGGC	ACACATTCTTACATGGTAAGTC
GABRB2-seq-Int4x	GTGGTTACTATCCTTAAGG	CAAAATGACTTAGACTGTG
GABRB2-seq-Int4y	TAGTCAGCATGCCAGCTTCC	ACCTAAGGATTGTTTGCTTTCC
GABRB2-seq-Int4z	GATCGGGAATAATCTGTTGTTC	ACAAGGGAATTCTATGCAACTG
GABRB2-seq-Int4aa	GTTTTGTGAACATCTCTTG	AAAGTCTGAAAACCAGGATC
GABRB2-seq-Int4ab	GAGCCTATATCACATATTGCAG	ATTATGTAAGACGTATGGGCAC
GABRB2-seq-Int4ac	TGTTCTGTTATGTTCTTCGG	CAACACAGGAGTTACAATTC
GABRB2-seq-Int4ad	TTGATCCACTCAGCTAAGAATG	CACAATTGTCAAAAGACCAC
GABRB2-seq-Int4ae	CAGGCTTCTGATTTGATG	CTTACACTTTCCATAACATG
GABRB2-seq-Int4a	ACTTTACTCGAATGAGCTATGG	AGCACAGTCTAATGGGATCTG
GABRB2-seq-Int4ag	GAAGTTTTATGGCCAAATACC	TTCATTGCTGACTTTCCTGC
GABRB2-seq-Int4ah	GCAGGAGTACCAGATTGTAC	CACAGGTCATGAGTGTTTCG
GABRB2-seq-Int4ai	TTACACACACACACACATACAC	AGGCCCCTAGGTGTTATCAC
GABRB2-seq-Int4aj	GGTAATGTCAGTTCACAAGGC	GGGAAAATCAAATGATGTGGATC
GABRB2-seq-Int4ak	GTGATGATCCACTTCGTTTG	TCATCACAGCTTGCCATATTAG
GABRB2-seq-Int4al	TATCCCTGTTTTGTGAGTC	AGGAAGTATGTCTGCACC
GABRB2-seq-Int4am	ACACAAGTAATCTGACATCCAG	TGAACCCTCTAGGATGTAACAG
GABRB2-seq-Int4an	GATGAGATTTACAAGTAACAAGG	GAAAAAGACATATGACAAACTGC
GABRB2-seq-Int4ao	TAAGAAACCATCACTACTGC	TTAGCCAATCATTCTCCATTG
GABRB2-seq-Int4ap	TTGGAATTTTTCTAGCTCAGGG	CCCAGTGACATAGGTTGAAG
GABRB2-seq-Int4aq	CAAGTGTTTTTCCAGGCAAC	TAGTGTGGACTGCTATCAG
GABRB2-seq-Int4ar	GGGAACTAATAGACATGGGTAC	GCAAGTCTGCATATAGTAGAACC
GABRB2-seq-Int4as	GTTTCTTTCAGGTATGTACTACAC	CAACTGTAGCTATGAGAAACTC
GABRB2-seq-Int4at	GCATACCAGACCTTAAACATCC	TCTCCCAGATGTCTCAGCC
GABRB2-seq-Int4au	GTGACGAAAGTAGAGATTAG	GGTTACAAATGTGTTTGCTG
GABRB2-seq-Int4av	AAGAGGTGGGAGAAGAGTGTC	CAAAATCATGTCAGTTGCTGC
GABRB2-seq-Int4aw	ATAGTGAGCACAGTATGCAACAG	GGTCTAGTACTCAGAATCTAC
GABRB2-seq-Int4ax	GTGTCTGAACATATTTTTTGCCCAG	TCTACCCCTCGAATCTCCTG
GABRB2-seq-Int4ay	GCAGTATGGCCATTTTAAGTAG	CACACAAACTTGGGTAATCACTG
GABRB2-seq-Int4az	CAGTTTGAGTGATGAACAAGCC	AGGTTTTCACTATGGCCTTGAGG
GABRB2-seq-Int4ba	GGTGACAAGAAATGAGACTCC	TCCCTACCTCATGACCTGC
GABRB2-seq-Int4bb	CTGGGAAGGTGAAACTCC	AGACATAAAACACGGCATGG
GABRB2-seq-Int4bc	TGACATCCAGGGATCTACAC	TCCTCTCAAGTTGTTGCCAC
GABRB2-seq-Int4bd	ATGGAAACCACAGACTCAGTC	GACTAAAACAGACCCTTCAC
GABRB2-seq-Int4be	TTCTAGTGGGGGTGTAGTGG	CAGAACCTTCAGATAGTGACAG
GABRB2-seq-Int4b	CTTAGGTATTTTCAGAATGTCC	AGGTAAAACTCCCATTAGGC
GABRB2-seq-Int4bg	GTTCCTACACAGCCTCATG	CTTGAGACCAGCAAAAACCC
GABRB2-seq-Int4bh	TTCCTGTGAAAAATAGGTGCAC	CCTTGGGAAATCGCACAAC
GABRB2-seq-Int4bi	CAGTTTTTTGGCTGTGCTTCAAG	TTGCCCTTCATAAGAGAGGC
GABRB2-seq-Int4bj	GAGATGTTCCTTATGTAGAAGTC	GTGAAAGAAGATCAAGAGGCG
GABRB2-seq-Int4bk	CACATCCTTTTCCCTTGAATAC	CTGGTTCTGACTATGAGGAG

GABRB2-seq-Int4bl	TAAAGTAGGCAGTAGAGAGAG	GGTGTATTGTGTTTGTGTCTGC
GABRB2-seq-Int4bm	CAGACACAAACACAATACAC	CCAAACATAACAGGCAATGTC
GABRB2-seq-Int4bn	GATATGCTTAATGCTAGATGAG	TTCCCGAGTCCACTACAG
GABRB2-seq-Int4bo	GACACACAGTAAGTAGTAGTC	CTTGTCAGCTGTGAAAGTTC
GABRB2-seq-Int4bp	GGAGATAGTTACACTGCATAC	CTTTGGAACACAAGGCTCTG
GABRB2-seq-Int4bq	ATGCCTGGGCTTATATTGTGG	CAGCTGTCACTTGTCTATCTC
GABRB2-seq-Int4br	GGGTGAAGCTATGTTTACTC	ACAGTTCTCTTAACATCTGGC
GABRB2-seq-Int4bs	CCTGAGAAGTGCTAATGTCC	CTGTTAGTGGCTGTCATATATCC
GABRB2-seq-Int4bt	AACTGAAGTAACTACTCTGCAG	CTGAAGCAGATAAGCTGGC
GABRB2-seq-Int4bu	GTGGGTTATAGCAACATTCCTG	CTTTCCTTGAATAAGTCCTACC
GABRB2-seq-Int4bv	CAGGATAATTTGGGGCAAACG	TCAAACTTTAAATGCCACCTGG
GABRB2-seq-Int4bw	GTCTTGGAGGAAGCATTATGAC	CATCTTTCTCAAAACACACTGC
GABRB2-seq-Int4bx	GTACCTACAATGCTTGGCTGC	GTTTTGCCATGTAGCTACTAATC
GABRB2-seq-Int4by	AAGTAGCTCAGTGAAACCTAGC	CCAGTTTTACATCTACACCAC
GABRB2-seq-Int4bz	GAGCTAAACATGTGCAGGACG	ATGTGCTATGGAGGGTGAC
GABRB2-seq-Int4ca	CAGATAGAATTCACAAACCTAC	TATGATGCATGAGTTCCAGC
GABRB2-seq-Int4cb	CTCATAGGATCAAACTTTGG	AACCCTTATGAGGCTTAATTG
GABRB2-seq-Int4cc	GATTCTGGATGCTTATGAGC	GTTTAGCCTTAGGCTTATC
GABRB2-seq-Int4cd	CAAGTGGTGACTAAAATAAGCC	ATGAAACCGCTGCCATTCTTC
GABRB2-seq-Int4ce	AACTAGGCTGATCTTCTCTG	GGATATGCTATTTCTCTCACTG
GABRB2-seq-Int4c	GCCTAGAGGTTTCATAAGTATC	CCAATTTCCAGAGTAGTTGTG
GABRB2-seq-Int4cg	ACAGCAGTGGGAGAGTTCTC	AAGACTAGGGAGTTCCTAGGC
GABRB2-seq-Int4ch	TTCACAAAGGCAGGCAGACC	ACCCTGTACCAGATCACTTG
GABRB2-seq-Int4ci	TACAGTGCGCATTTTCCAGG	GCATCCATTTGTAGCCATAGC
GABRB2-seq-Int4cj	ATCTCTCATGAACACCTTAG	AGCATTTTCCACACTTTTGC
GABRB2-seq-Int4ck	GAACCCTATGCTGAAGGCTC	CTATACCCACTCTGATTGACAC
GABRB2-seq-Int4cl	ATACCTGTGGATGTCTCCAG	TAAGACTCCTGGCCTGACC
GABRB2-seq-Int4cm	ACCAGAATGCTGAATCCAGG	AATGCAGAAGTTTGACAAGGC
GABRB2-seq-Int4cn	ACAGTTGCTGCCTAGTTTTC	GAAACTTGTCATTTTGCCTG
GABRB2-seq-Int4co	TAGCTGAGAATGGGAAAGGAG	GTGAAGAAGCCAACTGGCAG
GABRB2-seq-Int4cp	GGCCTATTGAAGAGAAAGTAC	ACTTTTGTGTGGCTGCCAAC
GABRB2-seq-Int4cq	GTATCAGACACAGAAGTGC	ATAGCAGAGTAACCATGAAC
GABRB2-seq-Int4cr	GTTCATGGTTACTCTGCTAT	TTCCTTCAATTCTAAGCAGTG
GABRB2-seq-Int4cs	GCTTGCTGCTTCAGAGAGAATG	CTGGAGCTATTCTTCCCCTC
GABRB2-seq-Int4ct	GTAGCAACAGAGACTTTGTG	AACCCTTATACTGTTGGTAGG
GABRB2-seq-Int4cu	GGCTTGCTGGATTGTATAGC	AGAGCAAAGTTGGAAGTCTC
GABRB2-seq-Int4cv	CCATTGGCCATACATTTGTG	CTGAGCTCAATAGCACATGAG
GABRB2-seq-Int4cw	CCTTTTTAGTTTGTTGAGGG	CATACCCCAAAATCCTTTTCG
GABRB2-seq-Int4cx	AGAACAACCAGCTCTCAGAG	GATTGGGAAGAGACCAGCTC
GABRB2-seq-Int4cy	TTCCCCAAGCTCTGATAAGG	CCCATAGAGAAACTCAGCAG
GABRB2-seq-Int4cz	CATCTTGGATCATGGCCTAC	CTAGAACTATTTGGGACATAC
GABRB2-seq-Int4da	GTGTATACCTGCACAACTTCAG	TGCCTATCTCTGACCCTTATC
GABRB2-seq-Int4db	AGGTGAACTGAGTCCTCTGC	GTGATCCAGGCGTCTTATC

GABRB2-seq-Int4dc	CAGGAATAATGATGACTAAAGAGG	CTTTCAAACTTCCATGAGGATCC
GABRB2-seq-Int4de	TTGTGGAGATTATTCCCCTCATC	TAAGAGCACCAATTGTTCCACC
GABRB2-seq-Int4d	GAAGACTGAAGTTAGGTTGCC	CCCTATGCCTAAAACAGAGC
GABRB2-seq-Int4dg	CTTAAGGTCTATGTGGTGTC	TTTCCTGTTGGATACATGGG
GABRB2-seq-Int4dh	AGCTGTCAAAACTTGCTGCTG	CACTCTCTCTGCTATCAGGC
GABRB2-seq-Int4di	CAGAAGGTCAGATCCTCAGC	AGGAAACACCAACATCAGCC
GABRB2-seq-Int4dj	ATGCTGCACCACACATCCAG	ACCCAAGAACCACCTTACAC
GABRB2-seq-Int4dk	AAGACATGGAAACATTGAGG	CAAAAGATATCTGACATCTGG
GABRB2-seq-Int4dl	GAGAGCCTATAACCATTATGG	GGAAATACAAAACTGTCAGCC
GABRB2-seq-Int4dm	CCTGTATCGGCTGAAGAAC	AACCCATGCACCAATATGTTC
GABRB2-seq-Int4dn	AGTTCCATTTGATCCAAACC	TTTGGGAACTTGTCTCAGAG
GABRB2-seq-Int4do	TTCGTGGAAGAAATAACCTGC	AGGATCTCTAGAGGAGCATG
GABRB2-seq-Int4dp	TTTGTAACATAGGCACACAC	GCTTTTAATACATGCCTGATC
GABRB2-seq-Int4dq	GTACCCATAGGCTTCTAATGG	GACAGATTAAATGCCTGCTCTC
GABRB2-seq-Int4dr	CCATTCCAGTTAGATATTTCGC	TTGAGACAGGGTCTCACTCTG
GABRB2-seq-Ex5	TGTTTTCTTGGCCGGTTGTGG	AGCAGCTTTAGGATCAGCAGC
GABRB2-seq-Int5a	TTGAGAGCTTAAGGTAGCATGC	GGTTTATGTACCATCTCCAGAG
GABRB2-seq-Int5b	GTTATGGGTCAGACATTGAGC	CTGCTATTTACAGTGTAGGTGC
GABRB2-seq-Int5c	GCACTCTTCCTGTTTCTCC	CTACTATTTGCCCATTGAAGC
GABRB2-seq-Int5d	GATTGTTCTCAGCTGTATTG	CAGCAAAATTCCCTACTAAAC
GABRB2-seq-Int5e	AAGTACAGGCTTTATTCTGGAC	CAGTGGCATGAATTCGCTC
GABRB2-seq-Int5	CCTGTAGTGCCTGCTTCTC	AAGCCACCGACAAAAGTCC
GABRB2-seq-Int5g-a	CCT CAG TTG TAG CAT GCC TTC	TGC TTT CAC CTG GTG AGG AC
GABRB2-seq-Int5g	AAAGCACACGATTGCAGC	AGGAGTGATAAGACATGATG
GABRB2-seq-Int5h	AACTCTTAACTGAGACCCTC	TCCAATGTAATGGTAGTGGG
GABRB2-seq-Int5i	GAGGGCAGGTTTCCTGATTC	ATGCACTTATACCTCCTGAACC
GABRB2-seq-Int5j	GTATCATTTTTGTAGAGGTGAGG	TGTGGTAGGATTGGGGTAC
GABRB2-seq-Int5k	TTGTGAGCATAACAGTGC	TGTGAGCATTTAATCAGAAG
GABRB2-seq-Int5l	ATGTTCTCAGTTGTAGCTG	TTTCTATCATCAGTCAGGC
GABRB2-seq-Int5m	TTGTGCCTAACATCCTAGAG	GAGTAAAATGAAGACCTGGC
GABRB2-seq-Int5n	CTGATTTCATTAGGAGTGAACC	ATCAACACTTTCAGTACCTGC
GABRB2-seq-Int5o	TCAGGGCTCAGATTGAACTC	ACCAGGGATTCGAAGCTGC
GABRB2-seq-Int5p	CTCTGCATGACTTCTAAGGC	GCTTGCTTTATATGTGATCTCC
GABRB2-seq-Int5q	CAACCTATGGAAGTTACTGCC	GCCTGGCATGGTATTGATAC
GABRB2-seq-Int5r	TTAGCCTCACTGCTGGTACC	CTCCAAGTACAAGCTTATGG
GABRB2-seq-Int5s	CCATCCAGTCAAGTCCTGAG	CCTCTCAGCAAGGACATTCC
GABRB2-seq-Int5t	CCAGATAGCAGCTTTCAGAGAG	CCAATACAGTGAAATGCCACACG
GABRB2-seq-Int5u	GCAATTCATACAAAATAGGGTC	ATCAATGTATAGGTCCAGTC
GABRB2-seq-Int5v	TTCATTGGCATAGGGTATGTC	GGATATAGCCTCTTCCTTGTC
GABRB2-seq-Int5w	GATTCTGATTTGGGGAACCTC	CATTGCATTCCCTCATTCCC
GABRB2-seq-Int5x	GAGAGATTTGGGGACACTAG	GGGGACGGTTATTGCGTAC
GABRB2-seq-Int5y	TCACCATTGACTATAGTCACCC	CTCAACATCTCAGATCATCAG
GABRB2-seq-Int5z	CACATCGTCATCACCATTTG	CAACAAAAGACCCAGAATAGCC

GABRB2-seq-Int5aa	GTCTATGTGTCTGTTTCTATGCC	CTAGCACATGTACTCTGGAAC
GABRB2-seq-Int5ab	GTGAAGGTGATTTTCTCTGGATATG	CATCAATGCCAGTCCTGTCTTAC
GABRB2-seq-Int5ac	TAGCTTACACACCCCAGTTAC	ACACAGAGAAGAAGTTCAGAAC
GABRB2-seq-Int5ad	GGACCCTGATGCATTCTTTGG	GTACTGCAAGGCACTTCTGG
GABRB2-seq-Int5ae	AGGACAATGCTGGGTCTTGG	CACTAGTTCCCAGACAACATATG
GABRB2-seq-Int5a	AAGGTCCTCAGTGTAGTACC	AGTGAGCATTCGTAGTACCTG
GABRB2-seq-Int5ag	ACTACTCTAAGACAGCTAGGG	AGGGGCAATACAGAGTAACTG
GABRB2-seq-Int5ah	CACACGAAGAAAGTGAATCTCAG	GGTAGATTATGTAGCCTTCAAG
GABRB2-seq-Int5ai	CTCATCACTAACAGGTTAGGC	CTGCTCTTCTTCACACTCTAC
GABRB2-seq-Int5aj	GTCCCATCAGATGAGGAAGTG	GAACCCTGCAGGATGATTCTC
GABRB2-seq-Int5ak	TGAGACAGAGATTTAGCATG	TGTGTTAAGAAGCACTCCAG
GABRB2-seq-Int5al	CTTTCGTATGGCCAGAGGTC	GTTATCAGTGCAAAGGACAATC
GABRB2-seq-Int5am	TACATTGAGCCCTTTTCCAG	TCTTAATTCCGGATAACACGG
GABRB2-seq-Int5an	TGCACATGGCCTAGTTCTTC	GCTTGGAAATAAATCTGGACC
GABRB2-seq-Int5ao	TTGAATTATCACAGGCACTG	TAGCAGATGAACACAGGAC
GABRB2-seq-Int5ap	GTGTTTGTCCCCAAGTCTAAG	TTAAACCTGGTGAAGGATGG
GABRB2-seq-Int5aq	GTCTGAGTTTGTGGCTCCTTG	GGCATCTACCTCTGAACATGG
GABRB2-seq-Int5ar	TACTTACCGTGTGATGCC	TTACCAAATGTATGCCATCTC
GABRB2-seq-Int5as	ACTGTCTATATGTATCCTGTC	GATTACTTGGCGGATAGTG
GABRB2-seq-Int5at	GAACCTATTCTTTACCCATC	GGCATATTTAGTAAGAAGCTC
GABRB2-seq-Int5au	GTGTGTAATTTCCAACATGAG	GGATATCAAGGATCTCTCAC
GABRB2-seq-Int5av	TAGTGATGCAGATCCTTGTG	ATGGTTTGAAGGATTGACTCC
GABRB2-seq-Int5aw	ATGCTTTGCTGCTTGATAGC	CACTTACCATTCCCAAAATC
GABRB2-seq-Int5ax	AAAGGAACCTATGTTTTCTGAG	CAATGGCCTCTAGGTGTTC
GABRB2-seq-Int5ay	TTCATAACATGGCTTTGTGGC	GTGCTTTGGAACTCATTCTC
GABRB2-seq-Int5az	CACACACACAAAAACAACCTGC	CTTTCTGGAGGGTAATTAGAC
GABRB2-seq-Int5ba	GTTTAACATCCTGGCATACAG	GCAGGAGTGTTGCCATTAG
GABRB2-seq-Int5bb	TTTCCTCCATCTACTGATGC	GGTTGCCCTGTCTTGCTTC
GABRB2-seq-Int5bc	CAGAGATAGAATCCCCAGG	CCTACAAAGGAAAGCAACACA
GABRB2-seq-Int5bd	CCCAAATAGTGAACGCTGAAC	GGGGAAGCATGGCAAAATCC
GABRB2-seq-Int5be	GTCTCAACCAAGGCTCACTGC	CCTGGCGGTCTCATCTATAGG
GABRB2-seq-Int5b	TTTCAGGGATGCAGACTTCC	GCTTCTCTGGAGGATTGATC
GABRB2-seq-Int5bg	CAAGGTCACAGAACATGTAATG	TGCCTTAACCCTAACTATATTC
GABRB2-seq-Int5bh	AAGCATGCTGTGTATTACAAC	CGGCCCCAAGTTTGTTTTTG
GABRB2-seq-Int5bi	GTGAGCTGAGATCATTCCAC	GTGTCTTTGAAGTGAGGCTAG
GABRB2-seq-Int5bj	CTACAGAAGGCACATTTCCAC	CTGCTAACGTGTAGATACTG
GABRB2-seq-Int5bk	GGAAGCTATTTACATCCATC	CCTGTCTTGAGAAACTAGAG
GABRB2-seq-Int5bl	CTTCTATACTTGGCTCTCTC	ATTTCCTTCAAGCAGTCACC
GABRB2-seq-Int5bm	GGAGGATGTGAGAGGATTGC	TAACCCAGGTAGACTGAGCC
GABRB2-seq-Int5bn	GAGCCAAGAATGACTTACTTC	GACTGAGAATGTGGGCAAAG
GABRB2-seq-Ex6	TAGATAAATGGCTGATTGTGTG	TCCTTCCAGGCTGTTAGC
GABRB2-seq-Int6a	CACCCTCTTAGTCTCATCTC	GCCATATTTCTGCTCACAGC
GABRB2-seq-Int6b	GTAACTGGACTGTTGTGAGGG	GAGCCACTGTACCACTAAG

GABRB2-seq-Int6c	CACTTCAATTCACAATCACCTTG	ATAAAGGAGGCCAGGGATATTG
GABRB2-seq-Int6d	CAGGGAACATTAGGATTGAGC	GTCCCTTTCCCAACCTGTAC
GABRB2-seq-Int6e	TTGCCGTTACCTGTGTGGTC	TAAGGGAGTGACTTTGCCAG
GABRB2-seq-Int6	GTCCTACACACTGCCCTTTG	ATTCAGTAACACAACAGGCAC
GABRB2-seq-Int6g	AAGGTACTTGTGAGTGTGTG	TTCCTCAGGGTAAATGTTCC
GABRB2-seq-Int6h	GAGGGATATGTAGCAATCAG	TGACACCATGGAGCATTTGC
GABRB2-seq-Int6i	GTGTAAAGTGGAGGTGTCAC	AGCCAGCACACATCGTCAC
GABRB2-seq-Int6j	CATAGTCACTTGTTCTCATGG	AATCACGCTCCAGAACACTC
GABRB2-seq-Int6k	CAAATGATCAGGCCACACCAC	GTTACATGTGGCTATTGAGCAC
GABRB2-seq-Int6l	TAATTTCTGTAACCAGTTGCC	CATACAAACAGGAATGCGTAG
GABRB2-seq-Int6m	CGGTTGTAAGTTTCACTATTAC	GAGCCTTATCAAACTCTCAG
GABRB2-seq-Int6n	AGGAGCAGTTACAAGGATACTG	GTACACAGCACAAGATCTGGC
GABRB2-seq-Int6o	TCATGATAGCCTCCAGAATC	ATCATGCAAAGAGGAATGTG
GABRB2-seq-Int6p	CCAGTATCAAGCAGAGTGGC	TGCAGTCAGCAACAATGCCC
GABRB2-seq-Int6q	GTGGCACATGCTGATGGTATC	ACCAGTTCTGACAGTGCTCG
GABRB2-seq-Int6r	CAGTCCCAGACAGTCTATCC	TGTCCAAGGGGCTGCCATG
GABRB2-seq-Int6s	CAATGTAAGATGAGCAATAGC	GAAAAATAACAGCTCCATCAGG
GABRB2-seq-Int6t	CAGTCCCTCTAATCACAGTAC	CTCAGCACTTTGGGCATAATG
GABRB2-seq-Int6u	CACAGGAGTTTGCACAATTG	CGAAGAAACAGAAAGTCTGATC
GABRB2-seq-Int6v	GTGGGGACAGTGATCCATC	CCTTCAACTGCACAATACAC
GABRB2-seq-Int6w	CCAGTTAGTAGGCGTTTTCC	GGATTTGGCTTATGACTCATAG
GABRB2-seq-Int6x	CCTTTCTGACATTATCCATAC	CTTCATTTACAGAAACAGCTG
GABRB2-seq-Int6y	GCACTCTGTCACATAAATG	TGACAGATTTTGTGTGTGC
GABRB2-seq-Int6z	ATACTCTGCTGTCATGGTAC	TTGGCATTTGACAGGTTTGC
GABRB2-seq-Int6aa	TCGTTAAATTGACCCACAGTTC	CACCCTAGGTTGGATTCATG
GABRB2-seq-Int6ab	GGGGCAGTCATCAAAACTAGG	ACAGGATGGCTTTTGGAGTCAG
GABRB2-seq-Int6ac	AAGTTCCCAGTTGCACAGC	GTTCTAGCCCTGATTCTGTC
GABRB2-seq-Int6ad	GTTGCTATAAGAAAGTTCACTG	GCTGTCAATAGAAGAAGGC
GABRB2-seq-Int6ae	ACACTTGTTCCAGCAATATG	CTGAAATGGGTAATAAATCCC
GABRB2-seq-Int6a	TGAAAGATGGGGTCTCTCAG	GTCTCCAATCCCTGAGCTC
GABRB2-seq-Int6ag	GGTTTTGTGTTTCTCTTATGATG	TACTGTTCTCCCTTTATGTCC
GABRB2-seq-Int6ah	ACCTCAAATGGGTGAAAATG	GCATCTCTACATACTCTGC
GABRB2-seq-Int6ai	GCTTCTACATATTACCACAGATG	ATTCATTCATCCGACACCTGC
GABRB2-seq-Int6aj	TACTAACTGTCCCAAGAGCC	CTTGCTGAATGACGCTGAAC
GABRB2-seq-Int6ak	TCAACTACATGTTCAGCGTC	TCATCCTGCAAAGCGAGAC
GABRB2-seq-Int6al	CTCATTATGTTGTACAGCTTG	GAAGTGCGCTATCCTCTAC
GABRB2-seq-Int6am	ACTAGGAAACAGGGAGAAGC	AAGGGTGAAAAGAGAGTCTGC
GABRB2-seq-Int6an	ATAAAGCCCATCTCCTCTTC	CAAAGACAAGGGTGAGCAC
GABRB2-seq-Int6ao	CCCACATTCAGTAAAGCCC	AGTTTACCTCATTCAATTCCC
GABRB2-seq-Int6ap	CTGGGCCATATTTCAAGC	TGCTCAGTGAAAGACAGAAC
GABRB2-seq-Int6aq	GCATCTCATTGTAAGCTCCAC	CTAATGGTCACACATGTTTGGC
GABRB2-seq-Int6ar	ATCACCAAGGCCATTATCACC	TGTGTGGCATAGGATCCTAG
GABRB2-seq-Int6as	GATTCACGTACCAAATTACTGG	GTTTGAAAAGGTAGCTTGGG

GABRB2-seq-Int6at	CCCTAGGTTCCATTCCTAC	GGTTCTGCATTCATCACTG
GABRB2-seq-Int6au	GTCCATCAACCCATCACATG	GTGGAAGAATGGCTCATTG
GABRB2-seq-Int6av	GGAAGAATCACAAAGACAAACC	CCTACCTTATTGGATTGACC
GABRB2-seq-Int6aw	TACTGTAATAGAGGAAGCAG	TTTGTTGAGACTGAGTTTTGC
GABRB2-seq-Int6ax	CAGGACAATTGCTTTAACCC	TCCCTTCTCATCTCATACAC
GABRB2-seq-Int6ay	AATGCTTCTGCAAGTCACAG	TCTAAAATGTGGCTGTAAGTC
GABRB2-seq-Int6az	GGAAGACAGGGACATGAATGC	CCTGATACTAGAACCAAGGG
GABRB2-seq-Int6ba	CATATAGTTGGCTTTCCCTTGG	GAGCCATCCATGTGAACAAG
GABRB2-seq-Int6bb	CATTATTTCTTGCTCTCTGGG	CAGTATGATGAAACTTTCAGC
GABRB2-seq-Int6bc	GTCATATCCCAATGCCATGC	CTGCTGTGAGACCATTGAATTG
GABRB2-seq-Int6bd	GCATATTCTTATCTACAGAGTC	TGATGGAAATAATCTCCCGGAG
GABRB2-seq-Int6be	GTATACATCAGTTTGTCCAGG	CCAAAGTTCAGTGGGAATC
GABRB2-seq-Int6b	CTGCAGATGGACTTGGGAAG	CACCAATGCACTCCAGCC
GABRB2-seq-Int6bg	GATTCCTAAACAAGACAGTG	ACCTCTTTCAAAACTGGATC
GABRB2-seq-Int6bh	CACTTGTCTATTTCATGGC	GCACATTAAGTAACCCTATTAG
GABRB2-seq-Int6bi	GCAGATATCTTAACCTCTCC	TAAAGAGACTTGTCCAGAAC
GABRB2-seq-Int6bj	GGAGTGAGATTTGGTGGGAG	CCAGGAGTAACATGTGCAAG
GABRB2-seq-Int6bk	TTTGTGTATCTGCCATTCAG	ATGGATGAAGCTGGAAACC
GABRB2-seq-Int6bl	CCCATTTATGAGTGAGAACATG	GGTATGGGTAAAGACTTCATG
GABRB2-seq-Int6bm	GCCTGTTCACTCTGATGAAAG	TGCCTGTTCACTCTGATGAAAG
GABRB2-seq-Int6bn	GCCTGTTCACTCTGATGAAAG	AGATTTGGGAAGCTGCAGCTG
GABRB2-seq-Int6bo	ATTCAGGAGGTATGTATGGGC	AACTATGATTGCACCACTGC
GABRB2-seq-Int6bp	CAGTCTGAAAGAATGCTGATGG	CACTGGACACTCCTGTTACC
GABRB2-seq-Int6bq	TAAGCTCATCCTGTACTCTTCC	GCATCTGACACACACACACAC
GABRB2-seq-Int6br	CTACTGCCATCCTTATACCC	TGGGTGTGGATTCTGGTAGG
GABRB2-seq-Int6bs	CTCTCAGCTAATGAACTTGC	GTTTTAGAAAGGACACAGGAG
GABRB2-seq-Int6bt	TTGTCTCCAATCAGGCTTC	GAGTTATAGGTGCTTTGGAC
GABRB2-seq-Int6bu	TTTATGTCAGCTTGCCAATG	GACTATAGTGGATAAATTCAGG
GABRB2-seq-Int6bv	CCTAAAACAGATGTCCACCC	CCATAAACATCTTCAGTAGG
GABRB2-seq-Int6bw	CAAAATGGCAAAGCTTCCATC	CCAAACTATACTGGCCCTC
GABRB2-seq-Int6bx	TGTAGAGATCTTCATGAAGGG	TAGTCCTGGGTCTAACTGTG
GABRB2-seq-Int6by	CTTTAGAGCAGAGTTTCACTC	ACCTTGACCACTACCTTTCC
GABRB2-seq-Int6bz	GGATAAGAGCAACGTTCCTG	AGATCTATTGTCCCACATTG
GABRB2-seq-Int6ca	CCATTCTCAGCAATTTTCAAGC	TCTAATAAGGGGTTAATGTCC
GABRB2-seq-Int6cb	TCTGATGATTAGTGAGATTGAGC	CTATCCCATTAGTTCTGTCCC
GABRB2-seq-Int6cc	GGATTGTTACTTCTATTTCTGG	AATTGTCCCAGTTTGCAGATG
GABRB2-seq-Int6cd	GTGGCACCTTTTGGATTTTCTG	TAGCCCACAAATGGTGGAGC
GABRB2-seq-Int6ce	AGGCTGCAAGGAAATCCTGG	CCAACTTGTCTTCCTCTGTAG
GABRB2-seq-Int6c	CATCAAAGGATAACCAAATGAG	AGAAGTCCTACCTAGGTTCC
GABRB2-seq-Int6cg	CTACAGCCCATGTATAGCAG	CATGCATTGCCTCATCTAATCG
GABRB2-seq-Int6ch	GACTATGGGACTGAATTTCAG	ATGGTCCAGGTGATATGCAC
GABRB2-seq-Int6ci	GAATGGGTTAAGATGGAGAG	TTTCCACAGCACTCTAGTTG
GABRB2-seq-Int6cj	GCTATAGAGAAGGACCGGAC	CCTGATGTTGATGACCGGTTG

GABRB2-seq-Int6ck	TAAGTTCTGTGATACATGTGCAG	AGAGATGGGGCTAGGTCC
GABRB2-seq-Int6cl	AAATCCAATGCCAGAGTGCAG	AAGGTGATCTGGTGGCTG
GABRB2-seq-Int6cm	TGTGTCAGTAGATTAACTGC	CCACAGTTCCTCAGAACC
GABRB2-seq-Int6cn	TCAATCATGCTGCTTCCCTG	CTAGTGGGAAATTAAGGTGAC
GABRB2-seq-Int6co	TGCCAATTGGTTATGATGAC	CTTTGGATATTCTCTGGAGG
GABRB2-seq-Int6cp	GAAAGCTGTAAAAGTCAGGTC	GCCTCTTGCTCTTACCATAG
GABRB2-seq-Int6cq	GAGTTCTCTTTATTGGGTGTTTC	AGTCTATTAAGCACTTCACTTCC
GABRB2-seq-Int6cr	TAACTTGCAGTATCAGGGCTC	TGAACAGTAGGCAAATTCTCAG
GABRB2-seq-Int6cs	CAGACTGAGTATTTTGTCACC	GTAGGATACCTTCAATTGGAG
GABRB2-seq-Int6ct	GGCTACTCTTATGAATAAAGGG	GTCAGAAGAGAAGTGATTACG
GABRB2-seq-Int6cu	CTACTCCTCAAAGTGATTGG	TCCCTTAACACTTTTCCATG
GABRB2-seq-In7a	GTGTGTGTATGTTTCCTTTC	TGCCTGTGGGGTTTATTTC
GABRB2-seq-Int7b	TTTGCTAGACACTCATTTAGG	GCCTCATAGAGTGAATCAAAC
GABRB2-seq-Ex8	TATGACCAGAACATGCAGCC	CTTAGGTAAAGAAGCACAGATAC
GABRB2-seq-Int8a	CATGTCAGTATTAGTTAACCG	GTAATAGCCCTTTTGTGATCC
GABRB2-seq-Int8b	GCAATGTGGATACCTGGGTG	GACAGCTATAAAGAGGACACG
GABRB2-seq-Int8c	GATACAGGATCATCCGTAGTTG	AAAGAATGCCAGTTAAGAGG
GABRB2-seq-Int8d	CCACTCCAAATTCCCAGATG	TGCCAGGGATGAGGTTAGG
GABRB2-seq-Ex9	CCAGCAGTCCCCATAAGTC	CATCCAAATCTTGTGTCATGAG
GABRB2-seq-Int9a	AACCAAGCCCAGGGAAGAG	GAGAGCTGGCTCTACTGGG
GABRB2-seq-Int9b	GGAGAAAGCATGGGCAACAC	TTTGAAGTCCTATGAGATAGC
GABRB2-seq-Int9c	ATTATGGTTCTCTCAGGCTGTC	AGGAATTAACCCTGTTTCACAG
GABRB2-seq-Int9d	GAGTCATGGTTTAGAACAATGG	GATTATAAGTGCTTGATCTTGG
GABRB2-seq-Int9e	GTTAGAGGAGGTCCACATTGC	ATTCAGTAGACCTCATGTTGG
GABRB2-seq-Int9	AAAATCCCAGCTACCTGCAC	GAAGACAAAGTAGAGAATGTGC
GABRB2-seq-Ex10	CCCATTCTCTGTTAGCCTTTATC	GAGCCTGCCAATACAATACC
GABRB2-seq-Int10a	CTTTTGTTGGTTGGCTGAG	GATAGTCTGAGAGCACAGAC
GABRB2-seq-Int10b	CACCTACTATGTCTGTGCTC	TCTTCTCAGCAGTTGTATAAG
GABRB2-seq-Int10c	GGGTTCATATTCAAGCATTGTC	CTACCACACTCTCAGATGGC
GABRB2-seq-Int10d	TGGATTTGCTAGGTTCAAAG	AGTCTTACTCTGTCATCC
GABRB2-seq-Int10e	CTGGCCAACATGGTGAAACC	CCATGTTGCCCAGGCTTGTC
GABRB2-seq-Int10	CACTTCTTCTCCGGGCACAG	TCTCCCTTGTAGGTCATTCAC
GABRB2-seq-Int10g	AGGATGTAATGGGTTCCAAG	GTATACCACAAAAACGTGGGAC
GABRB2-seq-Int10h	CATAGATTGCCAACCTAGAGTC	CTAATTTAACCACACAACTGAGG
GABRB2-seq-Int10i	GTACAAGTCACGTGTGTCAAAG	GTTCACCAGTTAAGTGGGTCC
GABRB2-seq-Int10j	CTTCCAAATGATTTAGATGGTC	TCTTTTCTTCCCTTCCATTAC
GABRB2-seq-Int10k	GGGAATGAACTGCAAAAGAC	CTGATAAGATTGGTAAGAGAG
GABRB2-seq-Int10l	GGATTAGATGAAACACGGCAG	CAATAACTCATCGTTCAAGATGC
GABRB2-seq-Int10m	CTCTGTATGTTTTCTGGAATAGC	TAACTTCAGGTGATCCACC
GABRB2-seq-Int10n	CAAGATAATTACAACCCCATG	GGTTACTGGATGAGATCTTG

GABRB2-seq-Int10o	GGAGTCTCAGGCTTCAGGTC	CATCAATGACTTCCTAGGTGTC
GABRB2-seq-Int10p	ATCACTAGATATACAGAAGGTAG	GATTCCCAAAGCTGTTCATGG
GABRB2-seq-Int10q	GGAAGGAAGGGAGATGGCAG	CTTGGATTCTGGGCTTTCTCC
GABRB2-seq-Int10r	AACTAATCCTCTTGGAGACC	GGTCAAGCTGTTATGTCCTG
GABRB2-seq-Int10s	GGTCACAGAGCAGTTATATAAC	TAAGCTGCAAGTTGAGTTGGG
GABRB2-seq-Int10t	CACAGCTAGTTAGTGGCAAACC	GCCTTACCCTAGACCTACTG
GABRB2-seq-Int10u	GCTTCTTCATCTTTAACATGC	GTTTCACCGTGTTAGCCAAG
GABRB2-seq-Int10v	GACAAAACCTAGGGGACTTAC	CCATGACACAAGCTTACCTATG
GABRB2-seq-Int10w	CGTGCAGGTTTGTTTCATAG	GCTTATACACTGTTGAAAGGAG
GABRB2-seq-Int10x	GTAATAGGATTGCTGGGTCAAG	GACACATATACCAATGGAGCAG
GABRB2-seq-Int10y	TAGCCAGTTATCCTAGCAAC	GTTCATGCCATTCTCCTGCC
GABRB2-seq-Int10z	CGAGACCATCCTGGCCAAC	CATTAGAATGTGGTCACTGCG
GABRB2-seq-Int10aa	TGCTACTTTCAAAAGACCATC	GCCAAAACAGAGATCTTGAACTG
GABRB2-seq-Int10ab	TAGATCCAATGCTAGTTGAGC	ATCTTCCAAAATCTCTGACAGTG
GABRB2-seq-Int10ac	GGGAAGATTTGGATGACAACC	CCAAAGTGCTGGGACTAC
GABRB2-seq-Int10ad	TCATTCTGCACTTCCTGAGGG	GCCTGCTTAATTTTTTGTGGC
GABRB2-seq-Int10ae	GAATTTGATGCTGTGAGAG	ATAACACTAACCTCAGAGATG
GABRB2-seq-Int10a	GAGTTAGCAGAAGATTCCAC	GCCAGACTATGAGAATGGAG
GABRB2-seq-Int10ag	TACAACTTACCTAGCTTCG	GCTTCCCTTCCTGCAAC
GABRB2-seq-Int10ah	ACAAGGACATCTCTCCAGAG	CTATTCTCTGCAGATAAATTGCC
GABRB2-seq-Int10ai	GTCAGGAATGGACTTGGGAG	CTTGAAAGCCCTATAAACAGG
GABRB2-seq-Int10aj	GGCCAATTGCTAACTATCCC	TTTGGAGGCCTACTACACC
GABRB2-seq-Int10ak	CTTCAGTTTCCTGCCCTGTC	CTGACTTGCTGATGCTAAGC
GABRB2-seq-Int10al	GGCTTAAGAAAGAAAGAGAG	ACACAGAGAGGTTAAGACAC
GABRB2-seq-Int10am	CATGCCTCAGTATGAAAAAGC	GATATTAGCCCATTGTCAGATG
GABRB2-seq-Int10an	CTAAAGAGCTTCTGCACAGC	CTTCCCTGTGTCTATGTGTTC
GABRB2-seq-Int10ao	AGTCCATATCCCTTGCAAGG	TGGGGTTGGAGTCAGAGTC
GABRB2-seq-Int10ap	CCTATTCTTAGTCCACAGAAG	TTCCCAGAGGTTATAGGTG
GABRB2-seq-Int10aq	CCACTCCCAGCATTTCTAC	CCTAGACAGAACAAGTTTTC
GABRB2-seq-Int10ar	CATATATCATCAGAGCCTATCAC	TAAGCTGTGCTAGGCATGC
GABRB2-seq-Ex11	GGCAAGTACTTGATAGTTACCC	CTGTACAACTGGTTTGAGGAGG
GABRB2-seq-3'UTRa	GATGTGAATGCCATAGATCGG	ACCACCTAGGAAATAGCCG
GABRB2-seq-3'UTRb	CAGGAAACCGATTTGGCCAG	CTTGATTAGGCAGATTGCAAC
GABRB2-seq-3'UTRc	GCATTCATTGGTTTGGGGCTTG	GATTGCCTAGGAGAAAGTGTG
GABRB2-seq-3'UTRd	CTGAGTATCATCTTTGGGCAG	ACCTACTAGCGGTCATGTAC
GABRB2-seq-3'UTRe	GATAGCCAGCAATTAAGGTG	ATGCTCAGTACCCTGTATTAC
GABRB2-seq-3'UTR	GGTCTTACAACATATCTACATG	TAAATTACTCTGCTGGGGGC
GABRB2-seq-3'UTRg	CCGTTTATGGCTTAAGTTGATGTG	CTTCAATTTCCAAAAACCTAGTAGC
GABRB2-seq-3'UTRh	CCAGAAAGGAGTGAACAG	CCCAGAGATTTTAATAGTCC
GABRB2-seq-3'UTRi	GCATGGGTGTTTAAATAGTTCTG	CAATTACGCCTCAGAACTGGC

Table A2.3, Variants identified in *GABRA1, GABRA6, GABRG2 and GABRB2* by Sanger sequencing.

Genomic location (GRCh37/hg19)	Ref	Alt	Type	Gene	Consequence	Zygosity	mRNA/protein change	rsID	MAF (1000g)	CADD PHRED
160768974	ATAC	-	DEL	*GABRB2*	Intronic	Hom		rs200288126	0.02034 GnomAD	0.459
160769451	G	A	SNP	*GABRB2*	Intronic	Hom		rs2910305	0.465855	0.583
160772290	G	C	SNP	*GABRB2*	Intronic	Hom		rs2194159	0.311701	1.127
160779758	A	C	SNP	*GABRB2*	Intronic	Hom		rs2910303	0.225639	3.219
160784052	G	A	SNP	*GABRB2*	Intronic	Hom		rs2962401	0.333267	0.179
160791732	T	A	SNP	*GABRB2*	Intronic	Hom		rs2910298	0.0517173	3.177
160793680	T	C	SNP	*GABRB2*	Intronic	Hom		rs1820092	0.252995	0.319
160795065	T	G	SNP	*GABRB2*	Intronic	Hom		rs2910295	0.333666	1.018
160795777	A	G	SNP	*GABRB2*	Intronic	Hom		rs2910294	0.333267	1.057
160795783	G	C	SNP	*GABRB2*	Intronic	Hom		rs2962399	0.30012	6.751
160802879	A	T	SNP	*GABRB2*	Intronic	Het		rs1345736	0.467053	0.785
160803417	A	G	SNP	*GABRB2*	Intronic	Hom		rs989349	0.0069888	0.425
160803585	C	T	SNP	*GABRB2*	Intronic	Hom		rs989350	0.0521166	5.315
160805755	-	G	INS	*GABRB2*	Intronic	Hom		rs34147050	0.32528	1.288
160806655	A	T	SNP	*GABRB2*	Intronic	Hom		rs1422952	0.0515176	0.965
160807733	T	C	SNP	*GABRB2*	Intronic	Hom		rs1159170	0.0513179	0.043
160808802	C	T	SNP	*GABRB2*	Intronic	Hom		rs1363697	0.32528	1.188
160810188	T	C	SNP	*GABRB2*	Intronic	Hom		rs1422953	0.0523163	1.466
160812482	C	T	SNP	*GABRB2*	Intronic	Hom		rs2910291	0.240216	3.167
160812952	T	G	SNP	*GABRB2*	Intronic	Het		rs2910290	0.321685	0.382
160814956	T	A	SNP	*GABRB2*	Intronic	Hom		rs2910289	0.253195	11.6
160815339	A	G	SNP	*GABRB2*	Intronic	Hom		rs2910288	0.253195	1.153
160816830	C	A	SNP	*GABRB2*	Intronic	Het		rs2910287	0.468051	0.092
160817367	T	G	SNP	*GABRB2*	Intronic	Hom		rs2910286	0.74348 GnomAD	1.072
160818567	T	A	SNP	*GABRB2*	Intronic	Hom		rs1422954	0.249002	5.962
160822394	A	G	SNP	*GABRB2*	Intronic	Het		rs2962398	0.468051	5.695
160822497	G	A	SNP	*GABRB2*	Intronic	Hom		rs2962397	0.319888	0.04
160823121	C	T	SNP	*GABRB2*	Intronic	Hom		rs3111046	0.0517173	0.855
160823692	C	T	SNP	*GABRB2*	Intronic	Hom		rs2962396	0.329673	0.356
160825020	GATA	-	DEL	*GABRB2*	Intronic	Hom		rs67936445	0.25 Genome DK	3.339
160825121	A	G	SNP	*GABRB2*	Intronic	Het		rs2910285	0.467452	3.658
160825685	G	T	SNP	*GABRB2*	Intronic	Hom		rs3111047	0.156749	0.079
160826843	T	C	SNP	*GABRB2*	Intronic	Hom		rs2962395	0.252995	1.83
160827390	T	C	SNP	*GABRB2*	Intronic	Hom		rs1363698	0.0519169	15.83
160827494	T	A	SNP	*GABRB2*	Intronic	Hom		rs1862426	0.000998403	5.728
160828583	A	G	SNP	*GABRB2*	Intronic	Het		rs2910284	0.468251	8.582

160830279	G	C	SNP	*GABRB2*	Intronic	Hom		rs2962394	0.301518	2.442
160832211	G	A	SNP	*GABRB2*	Intronic	Hom		rs2910283	0.300519	0.252
160833604	T	G	SNP	*GABRB2*	Intronic	Hom		rs918528	0.329073	3.707
160833871	A	G	SNP	*GABRB2*	Intronic	Hom		rs2962393	0.0523163	0.192
160834217	G	A	SNP	*GABRB2*	Intronic	Hom		rs1422950	0.679130 TOPMED	0.536
160834260	G	A	SNP	*GABRB2*	Intronic	Hom		rs1422951	0.328874	1.958
160835617	C	T	SNP	*GABRB2*	Intronic	Hom		rs2421923	0.0517173	2.039
160836620	T	G	SNP	*GABRB2*	Intronic	Hom		rs1946247	0.117812	2.268
160841762	C	A	SNP	*GABRB2*	Intronic	Hom		rs1363701	0.328874	0.725
160842989	T	C	SNP	*GABRB2*	Intronic	Hom		rs2962423	0.00139776	0.333
160843561	A	T	SNP	*GABRB2*	Intronic	Hom		rs2910304	0.99648 GnomAD	0.006
160860375	G	A	SNP	*GABRB2*	Intronic	Het				1.697
160881766	G	A	SNP	*GABRB2*	Intronic	Het		rs146198359	0.00279553	1.415
160888862	T	A	SNP	*GABRB2*	Intronic	Het		rs149477433	0.00279553	4.223
160889984	G	T	SNP	*GABRB2*	Intronic	Het		rs2962420	0.439696	0.002
160893600	T	C	SNP	*GABRB2*	Intronic	Hom		rs2962417	0.0071885	3.252
160894631	T	C	SNP	*GABRB2*	Intronic	Het		rs7724146	0.170727	6.618
160895592	T	C	SNP	*GABRB2*	Intronic	Hom		rs2962416	0.0071885	3.995
160896516	-	AA	INS	*GABRB2*	Intronic	Hom		rs34246966	0.7468	1.125
160896531	A	C	SNP	*GABRB2*	Intronic	Het		rs5008265	0.439097	1.274
160897482	T	C	SNP	*GABRB2*	Intronic	Het		rs532076409	0.000199681	2.544
160901573	C	T	SNP	*GABRB2*	Intronic	Hom		rs2962415	0.00838658	2.747
160901622	A	G	SNP	*GABRB2*	Intronic	Hom		rs2964777	0.00399361	1.436
160901981	T	A	SNP	*GABRB2*	Intronic	Hom		rs2962414	0.00838658	4.488
160902783	C	T	SNP	*GABRB2*	Intronic	Hom		rs2964776	0.253794	0.599
160904852	C	T	SNP	*GABRB2*	Intronic	Hom		rs2962410	0.0081869	1.377
160905188	G	T	SNP	*GABRB2*	Intronic	Het		rs2964775	0.438898	3.135
160905418	T	A	SNP	*GABRB2*	Intronic	Hom		rs2964774	0.0081869	0.021
160908195	A	G	SNP	*GABRB2*	Intronic	Het		rs2962407	0.431709	5.187
160910101	T	A	SNP	*GABRB2*	Intronic	Hom		rs2962406	0.28115	3.006
160910212	G	A	SNP	*GABRB2*	Intronic	Hom		rs2964773	0.178514	1.736
160910659	T	C	SNP	*GABRB2*	Intronic	Hom		rs2964772	0.105631	2.291
160919571	-	TGTG	INS	*GABRB2*	Intronic	Hom		rs111553488	0.399361	1.065
160935847	C	T	SNP	*GABRB2*	Intronic	Het		rs145763914	0.01993	1.098
160938919	A	T	SNP	*GABRB2*	Intronic	Hom		rs2964770	0.00419329	7.153
160941288	T	G	SNP	*GABRB2*	Intronic	Hom		rs7716156	0.00479233	7.324
160943423	A	G	SNP	*GABRB2*	Intronic	Hom		rs4489099	0.00479233	1.621
160944304	G	A	SNP	*GABRB2*	Intronic	Hom		rs10068375	0.00279553	1.507
160946002	C	T	SNP	*GABRB2*	Intronic	Hom		rs6879106	0.00479233	14.68
160946845	G	A	SNP	*GABRB2*	Intronic	Hom		rs10074319	0.00479233	3.622
160953031	C	T	SNP	*GABRB2*	Intronic	Hom		rs4426954	0.443291	8.428

160956665	A	G	SNP	*GABRB2*	Intronic	Hom		rs10077266	0.00439297	3.933
160964798	C	T	SNP	*GABRB2*	Intronic	Hom		rs2045568	0.413139	0.355
161112693	C	T	SNP	*GABRA6*	5'UTR	Het	NM_000811.3:c.-303C>T	rs3811995	0.396765	17.62
161113646	A	C	SNP	*GABRA6*	Intronic	Hom		rs6556556	0.0341454	1.296
161114522	T	C	SNP	*GABRA6*	Intronic	Het		rs4454083	0.261981	15.05
161115640	T	A	SNP	*GABRA6*	Intronic	Hom		rs6556557	0.033746	0.562
161116942	A	G	SNP	*GABRA6*	Intronic	Het		rs3811992	0.398163	8.152
161119071	G	T	SNP	*GABRA6*	Synonymous	Hom	NM_000811.3:c.951G>T, NP_000802.2:p.Ala317=	rs12522663	0.0317492	9.401
161119125	C	G	SNP	*GABRA6*	Synonymous	Het	NM_000811.3:c.1005C>G, NP_000802.2:p.Ala335=	rs13184586	0.401158	6.239
161120230	C	T	SNP	*GABRA6*	Intronic	Het		rs57113136	0.475439	0.316
161120411	A	C	SNP	*GABRA6*	Intronic	Het		rs13171954	0.401558	0.864
161120821	T	C	SNP	*GABRA6*	Intronic	Het		rs13190034	0.475639	0.343
161121999	C	T	SNP	*GABRA6*	Intronic	Het		rs6879618	0.223642	19.31
161122848	G	C	SNP	*GABRA6*	Intronic	Het		rs6883829	0.47524	5.946
161123514	C	T	SNP	*GABRA6*	Intronic	Het				6.12
161124464	T	C	SNP	*GABRA6*	Intronic	Het		rs13172914	0.402955	11.12
161125325	T	C	SNP	*GABRA6*	Intronic	Het		rs6556558	0.475439	4.288
161125546	G	T	SNP	*GABRA6*	Intronic	Het		rs6556559	0.401158	0.407
161125700	A	G	SNP	*GABRA6*	Intronic	Hom		rs6882370	0.0397364	6.086
161127111	A	G	SNP	*GABRA6*	Intronic	Hom		rs11959228	0.240016	6.122
161127234	G	A	SNP	*GABRA6*	Intronic	Het		rs11949158	0.255192	12.88
161127749	A	G	SNP	*GABRA6*	Intronic	Het		rs9688198	0.255192	5.279
161127822	C	T	SNP	*GABRA6*	Intronic	Het		rs11956731	0.254992	1.311
161128761	C	G	SNP	*GABRA6*	Synonymous	Hom	NM_000811.3:c.1344C>G, NP_000802.2:p.Val448=	rs4277944	0.96803 GnomAD	1.089
161128914	C	G	SNP	*GABRA6*	3'UTR	Het	NM_000811.3:c.*135C>T	rs3219151	0.4301	4
161274224	A	G	SNP	*GABRA1*	5'UTR	Hom	NM_000806.5:c.-441A>G	rs11576001	0.41271	14.78
161275413	C	T	SNP	*GABRA1*	5'UTR	Het	NM_000806.5:c.-31C>T	rs4608967	0.2216	0.514
161276539	T	C	SNP	*GABRA1*	Intronic	Het			0.16	10.39
161279937	C	T	SNP	*GABRA1*	Intronic	Hom		rs10476365	0.0111821	4.733
161280643	A	G	SNP	*GABRA1*	Intronic	Het		rs4340950	0.189097	10.57
161284782	A	C	SNP	*GABRA1*	Intronic	Het		rs4367330	0.407348	6.431
161286802	G	A	SNP	*GABRA1*	Intronic	Het		rs60768803	0.22484	1.194
161287847	G	A	SNP	*GABRA1*	Intronic	Het		rs10068980	0.33726	0.52
161289948	G	C	SNP	*GABRA1*	Intronic	Het		rs10476366	0.221556	3.765
161292038	A	G	SNP	*GABRA1*	Intronic	Het		rs7734447	0.490216	5.145
161293079	C	T	SNP	*GABRA1*	Intronic	Het		rs56873080	0.1877	2.444
161296230	G	A	SNP	*GABRA1*	Intronic	Het		rs7735530	0.225839	0.778
161297386	G	A	SNP	*GABRA1*	Intronic	Het		rs4364428	0.221046	0.023

161297795	G	A	SNP	*GABRA1*	Intronic	Het		rs6866984	0.2565	1.388
161299525	TTTC	-	DEL	*GABRA1*	Intronic	Het		rs75318793	0.2202	3.251
161299920	A	G	SNP	*GABRA1*	Intronic	Het		rs10065822	0.220248	1.08
161299989	C	T	SNP	*GABRA1*	Intronic	Het		rs10058095	0.220248	5.796
161300806	G	A	SNP	*GABRA1*	Intronic	Het		rs6894357	0.220847	6.102
161301452	-	C	INS	*GABRA1*	Intronic	Het		rs111834922	0.495607	1.045
161302728	G	A	SNP	*GABRA1*	Intronic	Het		rs4260711	0.494609	0.999
161305709	A	G	SNP	*GABRA1*	Intronic	Het		rs6556562	0.220447	0.143
161306450	G	C	SNP	*GABRA1*	Intronic	Het		rs11959044	0.220447	0.863
161306618	C	T	SNP	*GABRA1*	Intronic	Het		rs11952776	0.435703	1.478
161306771	T	C	SNP	*GABRA1*	Intronic	Het		rs10042696	0.402356	0.441
161308645	C	T	SNP	*GABRA1*	Intronic	Het		rs7701394	0.494209	4.963
161309303	A	G	SNP	*GABRA1*	Intronic	Het		rs4921199	0.220048	0.706
161312913	C	T	SNP	*GABRA1*	Intronic	Het		rs7732641	0.401358	6.629
161313159	T	C	SNP	*GABRA1*	Intronic	Het		rs4554269	0.435903	1.28
161313175	A	G	SNP	*GABRA1*	Intronic	Het		rs4259206	0.494808	5.402
161314125	A	G	SNP	*GABRA1*	Intronic	Het		rs1350372	0.401358	4.185
161314238	T	G	SNP	*GABRA1*	Intronic	Het		rs1350376	0.401158	14.13
161315096	A	C	SNP	*GABRA1*	Intronic	Het		rs966137	0.0207668	3.556
161315268	G	T	SNP	*GABRA1*	Intronic	Het		rs75585382	0.123003	2.354
161315769	T	C	SNP	*GABRA1*	Intronic	Hom		rs1037715	0.401558	4.564
161316596	A	G	SNP	*GABRA1*	Intronic	Het		rs12189424	0.221645	3.389
161317196	T	C	SNP	*GABRA1*	Intronic	Het		rs7702392	0.402157	4.016
161318132	A	G	SNP	*GABRA1*	Intronic	Het		rs10051659	0.596	2.803
161318629	G	T	SNP	*GABRA1*	Intronic	Het		rs1379547	0.221446	2.316
161319314	T	C	SNP	*GABRA1*	Intronic	Het		rs1157122	0.402157	4.835
161319595	T	C	SNP	*GABRA1*	Intronic	Het		rs12716387	0.402157	4.434
161321006	G	A	SNP	*GABRA1*	Intronic	Het		rs6867604	0.221446	1.093
161321468	A	C	SNP	*GABRA1*	Intronic	Het		rs6894517	0.436502	1.982
161322256	C	G	SNP	*GABRA1*	Intronic	Het		rs17545383	0.221446	2.41
161322889	G	A	SNP	*GABRA1*	Intronic	Het		rs2279020	0.362	0.367
161323241	G	C	SNP	*GABRA1*	Intronic	Het		rs2279021	0.1741	16.76
161324898	A	G	SNP	*GABRA1*	3'UTR	Het	NM_000806.5:c.*470A>G	rs2290732	0.5805	14.83
161325934	A	C	SNP	*GABRA1*	3'UTR	Het	NM_000806.5:c.*1506A>C	rs998754	0.419529	14.83
161498737	T	C	SNP	*GABRG2*	Intronic	Het		rs183294	0.39357	1.587
161500239	T	C	SNP	*GABRG2*	Intronic	Het		rs62384027	0.0864617	8.665
161501261	T	C	SNP	*GABRG2*	Intronic	Het		rs116155059	0.0125799	19.86
161510796	T	C	SNP	*GABRG2*	Intronic	Het		rs209350	0.243011	3.678
161511854	A	C	SNP	*GABRG2*	Intronic	Het		rs209351	0.10623	0.299
161513707	G	A	SNP	*GABRG2*	Intronic	Het		rs61009928	0.128594	4.557
161513901	G	A	SNP	*GABRG2*	Intronic	Het		rs539352915	0.00219649	0.006

161516403	G	A	SNP	*GABRG2*	Intronic	Het		rs72821329	0.119209	2.313
161517059	C	T	SNP	*GABRG2*	Intronic	Het		rs209353	0.498602	1.3
161522318	G	A	SNP	*GABRG2*	Intronic	Hom		rs2963170	0.000798722	5.584
161527253	A	G	SNP	*GABRG2*	Intronic	Het		rs977412	0.103834	9.856
161529369	C	T	SNP	*GABRG2*	Intronic	Hom		rs211036	0.342851	0.054
161529571	A	G	SNP	*GABRG2*	Intronic	Hom		rs211035	0.178714	0.054
161529720	A	G	SNP	*GABRG2*	Intronic	Hom		rs211034	0.278554	4.66
161529808	C	T	SNP	*GABRG2*	Intronic	Hom		rs211033	0.279553	1.291
161530385	T	C	SNP	*GABRG2*	Intronic	Hom		rs211032	0.279153	7.243
161530417	C	T	SNP	*GABRG2*	Intronic	Hom		rs211031	0.179912	5.38
161530735	C	T	SNP	*GABRG2*	Intronic	Hom		rs211030	0.279353	4.001
161531224	A	-	DEL	*GABRG2*	Intronic	Hom				5.328
161531460	T	A	SNP	*GABRG2*	Intronic	Hom		rs188379	0.175719	6.535
161531769	-	TTTAT	INS	*GABRG2*	Intronic	Hom		rs142766177	0.444888	6.111
161532821	-	ACACAC	INS	*GABRG2*	Intronic	Hom		rs149564929	NR	3.366
161533066	C	T	SNP	*GABRG2*	Intronic	Hom		rs211029	0.445088	5.508
161533448	G	C	SNP	*GABRG2*	Intronic	Hom		rs210991	0.198283	6.738
161533496	C	T	SNP	*GABRG2*	Intronic	Hom		rs210990	0.453874	0.739
161533712	C	G	SNP	*GABRG2*	Intronic	Hom		rs210989	0.188898	0.255
161533923	-	GC	INS	*GABRG2*	Intronic	Hom		rs3079257	0.178315	1.551
161533925	T	C	SNP	*GABRG2*	Intronic	Hom		rs59148473	0.178315	1.551
161533976	C	T	SNP	*GABRG2*	Intronic	Hom		rs2223090	0.179912	1.752
161534146	G	A	SNP	*GABRG2*	Intronic	Hom		rs184135	0.179912	1.741
161535030	G	A	SNP	*GABRG2*	Intronic	Hom		rs2103475	0.478634	1.453
161535135	T	C	SNP	*GABRG2*	Intronic	Hom		rs210988	0.176318	0.041
161535755	C	T	SNP	*GABRG2*	Intronic	Hom		rs210987	0.496406	3.573
161539134	G	C	SNP	*GABRG2*	Intronic	Hom		rs210983	0.278754	0.356
161541035	T	A	SNP	*GABRG2*	Intronic	Hom		rs169793	0.441693	0.892
161542811	G	A	SNP	*GABRG2*	Intronic	Hom		rs398690	0.445088	0.327
161545211	GT	-	DEL	*GABRG2*	Intronic	Hom		rs3079259	0.332867	0.407
161561803	G	T	SNP	*GABRG2*	Intronic	Het		rs2422106	0.402157	0.645
161562004	C	T	SNP	*GABRG2*	Intronic	Het		rs721719	0.170927	1.644
161564446	A	C	SNP	*GABRG2*	Intronic	Hom		rs2964616	0.0081869	1.921
161575653	G	A	SNP	*GABRG2*	Intronic	Het		rs211015	0.472244	0.714
161577753	A	G	SNP	*GABRG2*	Intronic	Het		rs145030721	0.00479233	0.22
161580983	C	T	SNP	*GABRG2*	3'UTR	Het	NM_000816.3:c.*609=	rs418210	0.379393	8.656
161581035	T	A	SNP	*GABRG2*	3'UTR	Het	NM_000816.3:c.*661=	rs424740	0.485423	12.69
161582494	AT	-	DEL	*GABRG2*	Intronic	Het		rs34705786	0.3023	14.61

Table A2.4. Heterozygous rare or novel coding whole genome variants (C1 minus B1)

Chr	Position	Ref	Change	Type	Gene name	Trancript_ID	Effect	Old AA/ New AA	Codon Num	RS_ID	MAF (1000g)
1	1635011	A	G	SNP	*CDK11A*	NM_024011	Missense	V/A	655		
1	2537722	T	A	SNP	*MMEL1*	NM_033467	Missense	N/Y	239	rs533316401	0.001997
1	11918519	C	T	SNP	*NPPB*	NM_002521	Missense	R/H	47	rs5229	0.003794
1	33490060	T	C	SNP	*AK2*	NM_001199199	Missense	M/V	68	rs548856916	0.000599
1	38230775	A	T	SNP	*EPHA10*	NM_001099439	Start gained: TTG, 5'UTR: 37 bases from TSS			rs369625502	0.001597
1	48260414	T	G	SNP	*LOC388630*	NM_001194986	Missense	T/P	278		
1	93649545	A	G	SNP	*CCDC18*	NM_206886	Missense	S/G	49	rs537114111	0.002596
1	103471838	T	C	SNP	*COL11A1*	NM_001190709	Missense	T/A	534	rs560019202	0.000399
1	110086039	C	T	SNP	*GPR61*	NM_031936	Missense	S/L	132	rs571875540	0.0002
1	115262365	T	A	SNP	*CSDE1*	NM_007158	Splice site acceptor				
1	120054176	G	A	SNP	*HSD3B1*	NM_000862	Missense	E/K	66	rs587718863	0.001398
1	151263572	A	C	SNP	*ZNF687*	NM_020832	Missense	T/P	1201		
1	152191082	C	G	SNP	*HRNR*	NM_001009931	Missense	S/T	1008	rs139947063	0.002396
1	152277184	T	C	SNP	*FLG*	NM_002016	Missense	H/R	3393	rs146234375	0.000399
1	152770589	A	G	SNP	*LCE1D*	NM_178352	Missense	S/G	107	rs11485496	0.000799
1	156206155	C	T	SNP	*PMF1*	NM_001199654	Missense	R/W	151	rs140410413	0.004193
1	156206155	C	T	SNP	*PMF1-BGLAP*	NM_001199662	Missense	R/W	149	rs140410413	0.004193
1	158813875	C	T	SNP	*MNDA*	NM_002432	Missense	T/I	178	rs148142374	0.004393
1	159824697	G	T	SNP	*C1orf204*	NM_001134233	Missense	P/H	61	rs563047828	0.0002
1	159824697	G	T	SNP	*VSIG8*	NM_001013661	Missense	P/H	364	rs563047828	0.0002
1	162343861	C	T	SNP	*C1orf111*	NM_182581	Missense	G/S	255	rs142892903	0.003994
1	162602365	C	T	SNP	*DDR2*	NM_006182	Start gained: ATG, 5'UTR: 246 bases from TSS				
1	165634302	T	C	SNP	*ALDH9A1*	NM_000696	Missense	Y/C	472	rs112689633	0.000399
1	179820453	G	A	SNP	*TOR1AIP2*	NM_001199260	Missense	A/V	27	rs145359130	0.0002
1	202302657	C	T	SNP	*UBE2T*	NM_014176	Missense	R/Q	69		
1	211846971	C	T	SNP	*NEK2*	NM_001204182	Missense	V/I	137	rs151049149	0.0002
1	214819386	A	G	SNP	*CENPF*	NM_016343	Missense	N/S	2158	rs376037863	0.001198
1	216462702	C	T	SNP	*USH2A*	NM_206933	Missense	D/N	631	rs552400144	0.0002
1	247040572	T	C	SNP	*AHCTF1*	NM_015446	Missense	Q/R	907	rs554869556	0.001398
2	74588664	C	G	SNP	*DCTN1*	NM_001190836	Missense	E/Q	1225	rs146083590	0.000599
2	152320756	C	A	SNP	*RIF1*	NM_001177663	Missense	N/K	1574	rs576707750	0.001198
2	152362733	C	T	SNP	*NEB*	NM_004543	Missense	G/R	6153		
2	167263018	A	T	SNP	*SCN7A*	NM_002976	Missense	M/K	1374		
2	168101751	*	+T	INS	*XIRP2*	NM_152381	Frame shift	-/?	1283		

2	179416845	C	G	SNP	*TTN*	NM_001256850	Missense	S/T	28620		
2	191865831	C	T	SNP	*STAT1*	NM_007315	Missense	S/N	144		
2	200322490	G	A	SNP	*SATB2*	NM_001172509	Start gained: ATG, 5'UTR: 487 bases from TSS			rs561544098	0.000599
2	200790179	A	T	SNP	*C2orf69*	NM_153689	Missense	E/V	243		
2	202352363	C	T	SNP	*ALS2CR11*	NM_001168221	Missense	R/Q	1812	rs34138250	0.003195
2	216964675	G	A	SNP	*TMEM169*	NM_001142310	Missense	V/I	102	rs147230933	0.002796
2	220348366	G	A	SNP	*SPEG*	NM_005876	Missense	G/S	2061		
2	220408050	G	T	SNP	*CHPF*	NM_024536	Missense	R/S	71	rs540846380	0.000998
2	223158834	G	T	SNP	*PAX3*	NM_000438	Stop gained	S/*	213	rs201251689	0.000998
2	228767773	C	A	SNP	*WDR69*	NM_178821	Missense	T/K	199	rs145956341	0.004193
2	237374160	A	T	SNP	*IQCA1*	NM_024726	Missense	F/Y	305		
2	238271906	G	A	SNP	*COL6A3*	NM_004369	Missense	A/V	2018	rs200239695	0.000998
2	241835216	C	T	SNP	*C2orf54*	NM_001085437	Missense	A/T	67	rs532509457	0.000399
3	12046236	T	C	SNP	*SYN2*	NM_003178	Splice site donor				
3	38927644	G	A	SNP	*SCN11A*	NM_014139	Missense	P/L	974		
3	48637986	G	T	SNP	*UQCRC1*	NM_003365	Missense	T/N	381		
3	49928930	G	A	SNP	*MST1R*	NM_001244937	Missense	L/F	1097		
3	58104626	G	T	SNP	*FLNB*	NM_001164317	Missense	G/C	925	rs139875974	0.0002
3	65425561	*	+CTG	INS	*MAGI1*	NM_004742	Codon change plus codon insertion	Q/HR	421	rs142043619	0.003395
3	77657071	C	G	SNP	*ROBO2*	NM_002942	Missense	L/V	1087		
3	100413651	G	A	SNP	*GPR128*	NM_032787	Missense	V/M	734	rs533038587	0.004193
3	111888137	A	T	SNP	*SLC9C1*	NM_183061	Missense	F/L	986	rs144423530	0.003594
3	122283015	C	T	SNP	*PARP9*	NM_001146104	Start gained: CTG, 5'UTR: 146 bases from TSS				
3	130381059	T	G	SNP	*COL6A6*	NM_001102608	Missense	L/V	2137	rs557527574	0.0002
3	135914662	C	T	SNP	*MSL2*	NM_018133	Start gained: CTG, 5'UTR: 707 bases from TSS			rs371044189	0.000998
3	140281713	G	A	SNP	*CLSTN2*	NM_022131	Missense	R/H	758		
3	182740302	C	T	SNP	*MCCC1*	NM_020166	Missense	S/N	591	rs569721834	0.002995
3	193080218	A	C	SNP	*ATP13A5*	NM_198505	Missense	I/S	163	rs138006161	0.002995
3	195508661	A	G	SNP	*MUC4*	NM_018406	Missense	S/P	3264	rs543282847	0.002596
3	195510011	C	T	SNP	*MUC4*	NM_018406	Missense	A/T	2814		
4	3076701	A	C	SNP	*HTT*	NM_002111	Missense	Q/P	50		
4	40356515	T	G	SNP	*CHRNA9*	NM_017581	Missense	L/W	473	rs371532896	0.000799
4	145567428	A	G	SNP	*HHIP*	NM_022475	Start gained: CTG, 5'UTR: 400 bases from TSS				
4	185646168	G	A	SNP	*MLF1IP*	NM_024629	Missense	S/F	88	rs561625231	0.0002
5	428112	C	T	SNP	*AHRR*	NM_001242412	Missense	A/V	304	rs527453593	0.0002
5	32312901	G	C	SNP	*MTMR12*	NM_001040446	Missense	A/G	15		
5	41929852	T	C	SNP	*FBXO4*	NM_012176	Missense	M/T	160	rs542026963	0.0002

5	54523012	G	T	SNP	*MCI*	NM_001190787	Start gained: CTG, 5'UTR: 46 bases from TSS			rs563249538	0.000799
5	68473370	C	A	SNP	*CCNB1*	NM_031966	Missense	A/D	405	rs532762503	0.001398
5	72313079	T	C	SNP	*FCHO2*	NM_001146032	Missense	V/A	214	rs143520046	0.001198
5	74869578	C	T	SNP	*POLK*	NM_016218	Missense	H/Y	142	rs543440102	0.000399
5	94749757	C	T	SNP	*FAM81B*	NM_152548	Missense	R/C	134	rs76962324	0.004393
5	134060780	*	+AATG	INS	*SEC24A*	NM_021982	Frame shift	-/N?	1093		
5	137515500	C	T	SNP	*KIF20A*	NM_005733	Missense	S/F	44	rs150704301	0.002596
5	139422557	C	T	SNP	*NRG2*	NM_001184935	Missense	S/N	33	rs545945148	0.000799
5	141242861	G	A	SNP	*PCDH1*	NM_032420	Missense	T/M	1012	rs200143790	0.002396
5	149242784	T	A	SNP	*PDE6A*	NM_000440	Missense	I/F	802		
5	150696625	C	T	SNP	*SLC36A2*	NM_181776	Missense	R/H	402	rs151035012	0.0002
5	150858889	A	T	SNP	*SLC36A1*	NM_078483	Missense	Q/L	333	rs140884226	0.000998
5	170814839	G	T	SNP	*NPM1*	NM_001037738	Start gained: CTG, 5'UTR: 114 bases from TSS			rs561702061	0.000998
5	178506427	G	A	SNP	*ZNF354C*	NM_014594	Missense	E/K	332	rs200761213	0.002396
6	44103070	G	C	SNP	*TMEM63B*	NM_018426	Missense	R/P	82	rs371238478	0.003994
6	44270199	G	A	SNP	*AARS2*	NM_020745	Missense	R/W	806	rs200653667	0.001198
6	89910970	C	T	SNP	*GABRR1*	NM_001256703	Missense	R/Q	46	rs146686910	0.001198
6	119332572	C	T	SNP	*FAM184A*	NM_024581	Missense	D/N	519	rs541431761	0.000399
6	129802568	G	A	SNP	*LAMA2*	NM_000426	Missense	R/Q	2578	rs530288620	0.0002
6	138428303	C	T	SNP	*PERP*	NM_022121	Missense	G/R	59	rs146568280	0.000799
6	149773790	G	A	SNP	*ZC3H12D*	NM_207360	Missense	P/L	250	rs369123617	0.000599
6	150921127	C	T	SNP	*PLEKHG1*	NM_001029884	Start gained: CTG, 5'UTR: 182 bases from TSS				
6	152786487	G	A	SNP	*SYNE1*	NM_033071	Missense	S/F	620	rs140135976	0.002796
7	195594	G	A	SNP	*FAM20C*	NM_020223	Missense	G/R	216	rs61734970	0.003395
7	6449794	A	G	SNP	*DAGLB*	NM_001142936	Missense	I/T	467	rs139753251	0.002796
7	20689672	A	G	SNP	*ABCB5*	NM_001163941	Missense	K/E	412	rs576787803	0.002596
7	30705195	C	T	SNP	*CRHR2*	NM_001202482	Missense	V/I	125	rs201038501	0.001997
7	44147041	C	T	SNP	*AEBP1*	NM_001129	Missense	A/V	200	rs139352566	0.001997
7	100349928	T	C	SNP	*ZAN*	NM_003386	Missense	S/P	734		
7	100350330	C	A	SNP	*ZAN*	NM_003386	Missense	P/T	868		
7	100639275	C	G	SNP	*MUC12*	NM_001164462	Missense	P/A	1811	rs111685361	0.002596
7	100647774	C	G	SNP	*MUC12*	NM_001164462	Missense	P/A	4644	rs111307353	0.004792
7	107706899	A	G	SNP	*LAMB4*	NM_007356	Missense	C/R	865	rs575894937	0.001198
7	141952094	C	T	SNP	*PRSS58*	NM_001001317	Missense	A/T	225	rs572297660	0.001398
8	3889511	T	C	SNP	*CSMD1*	NM_033225	Missense	I/V	176		
8	8998354	G	A	SNP	*PPP1R3B*	NM_001201329	Missense	P/S	270	rs201747156	0.000399
8	10467590	*	+CCTCTAACTG CACCCCCTCTT	INS	*RP1L1*	NM_178857	Stop gained	- /GLKQ	1340		

			CTTGCAGCCCTTCTTCTGTTTTAGTCC					KKGCKKRGCS*R			
8	23225708	C	A	SNP	*LOXL2*	NM_002318	Missense	V/L	53	rs138503568	0.002796
8	26218551	A	G	SNP	*PPP2R2A*	NM_002717	Missense	N/S	174	rs571776040	0.0002
8	38385849	G	A	SNP	*C8orf86*	NM_207412	Missense	H/Y	103	rs551114589	0.000399
8	42128899	G	C	SNP	*IKBKB*	NM_001190720	Missense	G/A	4	rs201028246	0.000599
8	48689449	G	A	SNP	*PRKDC*	NM_001081640	Missense	T/M	4015	rs56123237	0.001398
8	48771194	A	G	SNP	*PRKDC*	NM_001081640	Missense	I/T	2116	rs200510022	0.000399
8	48771413	T	C	SNP	*PRKDC*	NM_001081640	Missense	K/R	2114	rs201300612	0.000599
8	77895523	C	T	SNP	*PEX2*	NM_001172086	Missense	E/K	298	rs544763390	0.000399
8	86049708	T	C	SNP	*LRRCC1*	NM_033402	Missense	L/P	780	rs147274148	0.000998
8	92375732	A	T	SNP	*SLC26A7*	NM_134266	Missense	E/V	485	rs140569478	0.002396
8	121174734	A	G	SNP	*COL14A1*	NM_021110	Missense	D/G	92	rs554596714	0.003395
8	124266337	G	A	SNP	*ZHX1*	NM_001017926	Missense	T/I	617	rs530080011	0.000599
8	139767402	G	A	SNP	*COL22A1*	NM_152888	Missense	R/W	677	rs180963656	0.002796
8	144620854	G	A	SNP	*ZC3H3*	NM_015117	Missense	A/V	228	rs140144642	0.0002
8	145745747	A	T	SNP	*LRRC14*	NM_014665	Missense	Q/L	152	rs146976820	0.000998
9	36217445	C	T	SNP	*GNE*	NM_001190383	Missense	V/M	622	rs121908627	0.003794
9	100116936	T	G	SNP	*C9orf174*	NM_020893	Missense	L/V	1074		
9	117849385	C	T	SNP	*TNC*	NM_002160	Missense	D/N	209	rs143622444	0.0002
9	125391780	G	A	SNP	*OR1B1*	NM_001004450	Missense	P/L	12	rs543380400	0.0002
9	125796915	A	G	SNP	*GPR21*	NM_005294	Missense	T/A	24	rs535225883	0.003594
9	131196778	C	T	SNP	*CERCAM*	NM_016174	Missense	T/M	474	rs571686461	0.0002
9	134353917	A	G	SNP	*PRRC2B*	NM_013318	Missense	M/V	1557	rs371639999	0.000599
9	136131622	*	-T	DEL	*ABO*	NM_020469	Frame shift	-/-	165	rs563704490	0.003195
10	5790389	C	G	SNP	*FAM208B*	NM_017782	Missense	H/D	1669		
10	13322990	T	C	SNP	*PHYH*	NM_001037537	Missense	S/G	217		
10	16919081	G	A	SNP	*CUBN*	NM_001081	Missense	T/M	2974	rs374477671	0.000998
10	27524061	A	G	SNP	*ACBD5*	NM_145698	Missense	S/P	88	rs199700938	0.001198
10	35897387	C	T	SNP	*GJD4*	NM_153368	Stop gained	R/*	316	rs370286081	0.003395
10	50738878	G	A	SNP	*ERCC6*	NM_000124	Missense	T/M	144	rs149382642	0.000998
10	51748543	C	T	SNP	*AGAP6*	NM_001077665	Missense	S/L	23	rs544067964	0.000599
10	52226673	T	C	SNP	*SGMS1*	NM_147156	Start gained: CTG, 5'UTR: 475 bases from TSS				
10	64574169	T	C	SNP	*EGR2*	NM_000399	Missense	S/G	77	rs554815622	0.000399
10	70406268	A	G	SNP	*TET1*	NM_030625	Missense	D/G	1261		
10	70742489	G	C	SNP	*DDX21*	NM_004728	Missense	R/P	758		
10	74034791	G	C	SNP	*DDIT4*	NM_019058	Missense	D/H	182		
10	81841616	A	T	SNP	*C10orf57*	NM_025125	Missense	Q/L	36	rs557363346	0.0002
10	88917817	A	G	SNP	*FAM35A*	NM_019054	Missense	H/R	513	rs376872876	0.000599

10	118969016	*	+T	INS	*KCNK18*	NM_181840	Frame shift	-/?	121	rs541915908	0.003195
10	127519206	T	G	SNP	*BCCIP*	NM_078469	Missense	L/V	133		
10	127569378	G	C	SNP	*DHX32*	NM_018180	Missense	L/V	6	rs567672999	0.000998
10	129904735	C	T	SNP	*MKI67*	NM_001145966	Missense	G/D	1430		
11	598481	G	T	SNP	*PHRF1*	NM_020901	Missense	A/S	335	rs182144037	0.003195
11	1016919	C	T	SNP	*MUC6*	NM_005961	Missense	R/K	1961	rs536595477	0.000799
11	3800485	T	C	SNP	*NUP98*	NM_016320	Missense	T/A	27	rs532996300	0.000799
11	7846682	C	T	SNP	*OR5P3*	NM_153445	Missense	V/M	280	rs200018758	0.004193
11	19256330	T	C	SNP	*E2F8*	NM_024680	Missense	M/V	243	rs541607823	0.001198
11	46792478	G	A	SNP	*CKAP5*	NM_001008938	Missense	L/F	1016	rs367607182	0.001997
11	46893155	C	T	SNP	*LRP4*	NM_002334	Missense	R/Q	1538	rs140495790	0.002196
11	47193350	A	G	SNP	*ARFGAP2*	NM_001242832	Missense	L/P	197	rs575856978	0.003395
11	48285558	C	T	SNP	*OR4X1*	NM_001004726	Missense	A/V	49	rs140161944	0.004992
11	58604609	T	C	SNP	*GLYATL2*	NM_145016	Missense	T/A	119		
11	59562908	C	G	SNP	*STX3*	NM_001178040	Missense	R/G	247	rs575668968	0.0002
11	60711303	C	T	SNP	*SLC15A3*	NM_016582	Missense	R/H	285	rs570961645	0.004992
11	61017160	A	G	SNP	*PGA5*	NM_014224	Missense	T/A	265	rs470947	0.001198
11	64981653	G	A	SNP	*SLC22A20*	NM_001004326	Missense	A/T	104	rs371972039	0.000998
11	83170965	T	C	SNP	*DLG2*	NM_001142702	Missense	I/V	301	rs533405452	0.000399
11	95657063	G	A	SNP	*MTMR2*	NM_016156	Missense	P/L	19	rs574213477	0.000799
11	111796873	T	C	SNP	*C11orf52*	NM_080659	Missense	F/L	108	rs587635751	0.001997
11	130275586	C	G	SNP	*ADAMTS8*	NM_007037	Missense	S/T	846		
12	11461534	*	- CCTCCTTGTGG GGGTGGTCTTT CTGGCTTTCCT GGAGGAGGTG GGGTACCTTG GGACTGGTTT	DEL	*PRB4*	NM_002723	Codon change plus codon deletion	GNQSQ GTPPP PGKPE RPPPQ GG/G	107		
12	11506094	*	-G	DEL	*PRB1*	NM_005039	Frame shift	-/-	315	rs576051236	0.001198
12	11508441	G	A	SNP	*PRB1*	NM_005039	Missense	A/V	16	rs530990598	0.001198
12	18719939	A	G	SNP	*PIK3C2G*	NM_004570	Missense	D/G	1279	rs563891575	0.004593
12	48105506	T	C	SNP	*ENDOU*	NM_001172439	Missense	Y/C	342	rs546409433	0.000399
12	52994888	C	T	SNP	*KRT72*	NM_001146225	Missense	E/K	117		
12	53241902	G	A	SNP	*KRT78*	NM_173352	Missense	R/W	130	rs138275245	0.0002
12	53566176	G	A	SNP	*CSAD*	NM_015989	Missense	H/Y	127		
12	54963333	G	A	SNP	*PDE1B*	NM_000924	Missense	M/I	138	rs552827981	0.000799
12	57423001	C	T	SNP	*MYO1A*	NM_005379	Missense	E/K	974	rs370708976	0.002196
12	65672579	C	T	SNP	*MSRB3*	NM_198080	Missense	L/F	11	rs200778091	0.001597
12	82871118	A	C	SNP	*C12orf26*	NM_032230	Missense	Y/S	569	rs574493780	0.0002
12	95879687	A	G	SNP	*METAP2*	NM_006838	Missense	I/V	120	rs143687752	0.000998

12	97331118	A	G	SNP	*NEDD1*	NM_001135176	Missense	Q/R	355	rs368698486	0.003395
12	119942926	G	A	SNP	*CCDC60*	NM_178499	Missense	R/Q	234	rs146647643	0.000399
12	122064852	A	C	SNP	*ORAI1*	NM_032790	Missense	N/H	69	rs557887137	0.001198
12	122217430	T	G	SNP	*RHOF*	NM_019034	Missense	K/Q	204	rs145959118	0.001597
12	129189651	C	T	SNP	*TMEM132C*	NM_001136103	Missense	T/M	713	rs561572424	0.0002
12	132624314	C	T	SNP	*DDX51*	NM_175066	Missense	E/K	614	rs61756267	0.001398
12	133311055	G	A	SNP	*ANKLE2*	NM_015114	Missense	L/F	603	rs201330179	0.0002
12	133780554	T	C	SNP	*ZNF268*	NM_003415	Missense	F/S	761	rs142121386	0.000998
13	28626716	C	T	SNP	*FLT3*	NM_004119	Missense	V/M	194	rs146030737	0.001997
13	53421707	T	G	SNP	*PCDH8*	NM_002590	Missense	T/P	289		
13	100622838	*	+CGG	INS	*ZIC5*	NM_033132	Codon change plus codon insertion	P/PR	364	rs532287066	0.002196
13	103390332	C	T	SNP	*CCDC168*	NM_001146197	Missense	E/K	4239	rs200872789	0.003594
14	20585947	T	C	SNP	*OR4K17*	NM_001004715	Missense	C/R	128	rs559109216	0.000998
14	24808496	T	G	SNP	*RIPK3*	NM_006871	Missense	S/R	66		
14	31582503	T	C	SNP	*HECTD1*	NM_015382	Missense	Y/C	2015		
14	39716424	C	T	SNP	*MIA2*	NM_054024	Missense	P/S	216	rs79728805	0.001597
14	55878506	A	G	SNP	*ATG14*	NM_014924	Missense	L/P	12	rs563928479	0.002196
14	69061297	C	T	SNP	*RAD51B*	NM_133509	Missense	H/Y	378	rs561626209	0.000998
14	78174470	G	C	SNP	*SLIRP*	NM_031210	Missense	A/P	6	rs147560423	0.001997
14	92537355	*	+CTGCTGCTGC TGCTGCTGCTG	INS	*ATXN3*	NM_004993	Codon change plus codon insertion	Q/HSSS SSSR	305		
14	94394909	G	C	SNP	*FAM181A*	NM_001207071	Missense	G/A	93	rs533878313	0.002796
14	105407344	C	T	SNP	*AHNAK2*	NM_138420	Missense	G/E	4815	rs368977213	0.000799
14	105414119	G	A	SNP	*AHNAK2*	NM_138420	Missense	P/S	2557		
14	105414133	G	A	SNP	*AHNAK2*	NM_138420	Missense	P/L	2552		
15	45406921	C	T	SNP	*DUOXA2*	NM_207581	Missense	L/F	40	rs545507872	0.002995
15	64041905	G	C	SNP	*HERC1*	NM_003922	Missense	A/G	663	rs137926425	0.002196
15	74836290	*	+CAG	INS	*ARID3B*	NM_006465	Codon insertion	-/Q	5	rs574173186	0.004992
15	89198722	G	A	SNP	*ISG20*	NM_002201	Missense	R/Q	169	rs370675181	0.003395
16	1447265	*	- CCCCGTGCTTC ACACTGGGGC AGGGCGTGGA CCTGGGGATG AGGAGGTGTC AGGGGGACAC AGAGGACTTG GCTCCCCGCCC CCCAC	DEL	*UNKL*	NM_001193388	Frame shift	-/-	245		
16	1447265	*	- CCCCGTGCTTC ACACTGGGGC	DEL	*UNKL*	NM_001193388	Splice site acceptor				

			AGGGCGTGGACCTGGGGATGAGGAGGTGTCAGGGGGACACAGAGGACTTGGCTCCCCGCCCCCCAC								
16	2027615	C	T	SNP	*TBL3*	NM_006453	Missense	R/W	615	rs574487347	0.000599
16	2137891	G	C	SNP	*TSC2*	NM_000548	Missense	V/L	1673	rs45490993	0.000599
16	3071577	C	G	SNP	*TNFRSF12A*	NM_016639	Missense	P/R	74	rs181233313	0.000599
16	3350028	T	C	SNP	*TIGD7*	NM_033208	Missense	K/R	196	rs571664375	0.001997
16	21663909	G	A	SNP	*IGSF6*	NM_005849	Missense	A/V	4	rs565354015	0.001398
16	55903593	C	T	SNP	*CES5A*	NM_001143685	Missense	A/T	161	rs552972383	0.0002
16	58550386	G	A	SNP	*SETD6*	NM_001160305	Missense	E/K	161	rs147467868	0.000599
16	70557398	C	T	SNP	*COG4*	NM_001195139	Missense	V/M	17		
16	71571086	G	A	SNP	*CHST4*	NM_005769	Missense	R/H	169	rs375170120	0.004792
16	84189286	C	G	SNP	*DNAAF1*	NM_178452	Missense	L/V	225		
16	88643673	C	G	SNP	*ZC3H18*	NM_144604	Missense	L/V	48	rs371740351	0.001198
16	89211778	G	C	SNP	*ACSF3*	NM_001127214	Missense	E/D	490	rs147538370	0.000399
16	89985866	G	A	SNP	*MC1R*	NM_002386	Missense	R/Q	67	rs34090186	0.001398
17	1637492	G	A	SNP	*WDR81*	NM_001163673	Missense	V/I	518	rs143022530	0.001797
17	5085970	G	A	SNP	*ZNF594*	NM_032530	Missense	R/C	528	rs181984106	0.002396
17	7836474	G	T	SNP	*CNTROB*	NM_001037144	Missense	G/V	26	rs372319018	0.002995
17	8701676	A	T	SNP	*MFSD6L*	NM_152599	Missense	F/I	255	rs372787695	0.000998
17	10315826	G	A	SNP	*MYH8*	NM_002472	Missense	A/V	426	rs150633264	0.000399
17	26958522	C	T	SNP	*KIAA0100*	NM_014680	Missense	R/Q	1425		
17	27185829	T	C	SNP	*ERAL1*	NM_005702	Missense	V/A	316	rs577237073	0.002196
17	34869259	C	T	SNP	*MYO19*	NM_001033580	Missense	A/T	303	rs556944760	0.001198
17	34937924	A	G	SNP	*GGNBP2*	NM_024835	Missense	T/A	391	rs555095991	0.000399
17	36717904	*	+GCGGCCCCGAGTAGGGGCTGT	INS	*SRCIN1*	NM_025248	Codon insertion	-/TAPTRGR	499		
17	39197304	T	C	SNP	*KRTAP1-1*	NM_030967	Missense	I/V	116		
17	40462547	G	C	SNP	*STAT5A*	NM_003152	Missense	D/H	749		
17	40729675	C	T	SNP	*PSMC3IP*	NM_001256014	Start gained: CTG, 5'UTR: 250 bases from TSS			rs202144890	0.000599
17	42254204	A	G	SNP	*ASB16*	NM_080863	Missense	E/G	223	rs200625323	0.001997
17	72745310	G	T	SNP	*SLC9A3R1*	NM_004252	Stop gained	E/*	109	rs537924938	0.001398
17	78064186	*	-AGCACGTGCATGAACAACACAGGACACACA	DEL	*CCDC40*	NM_001243342	Stop lost	TARA*/?	1027		

			CAGCACGTGC ATGAACAACA CAGGACACAC ACA								
18	7016514	C	T	SNP	*LAMA1*	NM_005559	Missense	A/T	989	rs181556627	0.000599
18	19075727	C	G	SNP	*GREB1L*	NM_001142966	Missense	L/V	1043		
18	29598762	G	T	SNP	*RNF125*	NM_017831	Start gained: CTG, 5'UTR: 65 bases from TSS			rs566672198	0.002596
19	4511283	C	T	SNP	*PLIN4*	NM_001080400	Missense	A/T	883	rs80238130	0.001597
19	7125518	C	T	SNP	*INSR*	NM_000208	Missense	V/M	1012	rs1799816	0.004992
19	9076546	G	A	SNP	*MUC16*	NM_024690	Missense	R/C	3634	rs191804178	0.001997
19	12256396	G	T	SNP	*ZNF625*	NM_145233	Missense	Q/K	279		
19	15483804	G	T	SNP	*AKAP8*	NM_005858	Missense	P/H	240	rs374389542	0.002995
19	17003939	T	C	SNP	*CPAMD8*	NM_015692	Missense	N/D	1927	rs182109236	0.002995
19	19824930	T	C	SNP	*ZNF14*	NM_021030	Missense	D/G	54	rs543600452	0.004193
19	21132822	A	G	SNP	*ZNF85*	NM_001256171	Missense	K/R	531	rs145468338	0.000799
19	22586287	A	T	SNP	*ZNF98*	NM_001098626	Missense	L/I	20	rs2957819	0.0002
19	31769992	G	A	SNP	*TSHZ3*	NM_020856	Missense	T/M	236	rs557330136	0.0002
19	36018037	C	T	SNP	*SBSN*	NM_001166034	Missense	A/T	383	rs561422483	0.002995
19	38877546	A	C	SNP	*GGN*	NM_152657	Missense	V/G	119		
19	46307829	G	C	SNP	*RSPH6A*	NM_030785	Missense	T/S	445	rs558606677	0.000599
19	53116934	*	- CCACACTCATT ACATTTGTAAG GTTTCTCTCCA GTGTGGATTCT CTGATGTTGTG CAAGGTGTGA AATATGATGG AAGACCTTT	DEL	*ZNF83*	NM_018300	Codon change plus codon deletion	GKVFH HISHL AQHQ RIHTG EKPYK CNECG /G	267		
19	54682518	A	C	SNP	*MBOAT7*	NM_001146056	Missense	I/S	259		
19	57286568	A	G	SNP	*ZIM2*	NM_001146326	Missense	C/R	358	rs148139560	0.004992
20	3845239	C	T	SNP	*MAVS*	NM_001206491	Missense	T/I	180	rs558231938	0.0002
20	4229375	G	T	SNP	*ADRA1D*	NM_000678	Missense	A/E	77	rs61759853	0.004393
20	25255313	G	A	SNP	*PYGB*	NM_002862	Missense	G/E	205		
20	32255496	C	A	SNP	*ACTL10*	NM_001024675	Missense	P/T	65	rs538359598	0.000399
20	36152888	C	T	SNP	*BLCAP*	NM_001167823	Start gained: TTG, 5'UTR: 285 bases from TSS			rs149844774	0.001597
20	44505406	C	T	SNP	*ZSWIM3*	NM_080752	Missense	T/M	70	rs573326472	0.000599
20	49366743	G	T	SNP	*PARD6B*	NM_032521	Missense	Q/H	279	rs41283604	0.003395
21	40571063	G	C	SNP	*BRWD1*	NM_018963	Missense	S/C	1760		

21	43557751	*	- GGTGGGTGCG GAGTGGGGTG GGAGGTGCAG GCTGACAGGA AGGTGGGTGT GGAGTGGGGT GGGAGGTGCA GGCTGACAGG AG	DEL	*UMODL1*	NM_173568	Splice site donor				
21	43959746	G	A	SNP	*SLC37A1*	NM_018964	Missense	V/M	159	rs370831528	0.000599
21	45993626	A	T	SNP	*KRTAP10-4*	NM_198687	Start gained: CTG, 5'UTR: 10 bases from TSS			rs371752955	0.0002
21	47831802	C	T	SNP	*PCNT*	NM_006031	Missense	R/W	1939	rs575313866	0.001797
22	17583071	C	T	SNP	*IL17RA*	NM_014339	Missense	A/V	214	rs558799480	0.000399
22	19208911	C	A	SNP	*CLTCL1*	NM_001835	Stop gained	E/*	929	rs546347801	0.0002
22	36537775	C	T	SNP	*APOL3*	NM_145640	Missense	A/T	228	rs142057520	0.004792
22	38119273	G	A	SNP	*TRIOBP*	NM_001039141	Missense	R/Q	237	rs372079149	0.0002
22	45312435	T	C	SNP	*PHF21B*	NM_001242450	Missense	T/A	85	rs371737547	0.000998
22	46762301	G	A	SNP	*CELSR1*	NM_014246	Missense	S/L	2761	rs144039991	0.001398
22	50956614	G	A	SNP	*NCAPH2*	NM_014551	Missense	V/I	185	rs141520277	0.002396
X	70367609	C	T	SNP	*NLGN3*	NM_001166660	Missense	R/W	4	rs376877146	0.002119
X	73960250	A	G	SNP	*KIAA2022*	NM_001008537	Missense	I/T	1381		
X	151900574	T	C	SNP	*MAGEA12*	NM_001166386	Missense	N/S	76	rs369257561	0.00053
X	152483225	*	-T	DEL	*MAGEA1*	NM_004988	Splice site acceptor			rs782577564	0.001854
X	153657081	C	T	SNP	*ATP6AP1*	NM_001183	Stop gained	R/*	15	rs201620814	0.000795

Table A2.5 Heterozygous novel/ rare (MAF<0.005) SNV at 5q34 C1 minus B1

Genomic change GRCh37/hg19	Gene/ Closest gene and distance	mRNA change/ protein change	Consequence	rsID	MAF (1000G)
chr5:g.153558913A>G	*MFAP3-121899, GALNT10-11382*		Intergenic	rs887385914	NA
chr5:g.153568382G>A	*MFAP3-131368, GALNT10-1913*		Upstream: 1913 bases from GALNT10	rs562160602	0.00120
chr5:g.153598033G>C	*GALNT10*	NM_198321:c.159+27448G>C	Intron	rs527304899	0.00260
chr5:g.153599266A>C	*GALNT10*	NM_198321:c.159+28681A>C	Intron	rs539641874	0.00260
chr5:g.153601426C>T	*GALNT10*	NM_198321:c.159+30841C>T	Intron	rs536121175	0.00260
chr5:g.153601759T>C	*GALNT10*	NM_198321:c.159+31174T>C	Intron	rs535695656	0.00020
chr5:g.153601836A>T	*GALNT10*	NM_198321:c.159+31251A>T	Intron	rs556114820	0.00260
chr5:g.153602414A>T	*GALNT10*	NM_198321:c.159+31829A>T	Intron	rs572326204	0.00260
chr5:g.153603947T>A	*GALNT10*	NM_198321:c.159+33362T>A	Intron	rs557958036	0.00260

chr5:g.153605356T>A	*GALNT10*	NM_198321:c.159+34771T>A	Intron	rs796693396	NA
chr5:g.153605371T>C	*GALNT10*	NM_198321.3:c.159+34786T>C	Intron	rs1433617447	NA
chr5:g.153606092A>G	*GALNT10*	NM_198321.4:c.159+35507A>G	Intron	rs578085750	0.00260
chr5:g.153606449A>G	*GALNT10*	NM_198321.3:c.159+35864A>G	Intron	rs543517198	0.00260
chr5:g.153608763A>G	*GALNT10*	NM_198321.3:c.159+38178A>G	Intron	rs541693785	0.00260
chr5:g.153614160G>A	*GALNT10*	NM_198321.3:c.159+43575G>A	Intron	rs563927946	0.00260
chr5:g.153616948G>T	*GALNT10*	NM_198321.3:c.159+46363G>T	Intron	rs549916666	0.00280
chr5:g.153617822C>T	*GALNT10*	NM_198321.3:c.159+47237C>T	Intron	rs529537524	0.00260
chr5:g.153618925T>G	*GALNT10*	NM_198321.3:c.159+48340T>G	Intron	rs549680781	0.00260
chr5:g.153620144A>C	*GALNT10*	NM_198321.3:c.159+49559A>C	Intron	rs534702423	0.00260
chr5:g.153621183T>C	*GALNT10*	NM_198321.3:c.159+50598T>C	Intron	rs571562891	0.00260
chr5:g.153627645G>A	*GALNT10*	NM_198321.3:c.160-46731G>A	Intron	rs573029085	0.00280
chr5:g.153629897G>A	*GALNT10*	NM_198321.3:c.160-44479G>A	Intron	rs564073553	0.00260
chr5:g.153632040G>A	*GALNT10*	NM_198321.3:c.160-42336G>A	Intron	rs563405059	0.00280
chr5:g.153634392C>T	*GALNT10*	NM_198321.3:c.160-39984C>T	Intron	rs529497768	0.00260
chr5:g.153635933G>T	*GALNT10*	NM_198321.3:c.160-38443G>T	Intron	rs539042752	0.00300
chr5:g.153639336A>G	*GALNT10*	NM_198321.3:c.160-35040A>G	Intron	rs557021483	0.00260
chr5:g.153641403A>C	*GALNT10*	NM_198321.4:c.160-32973A>C	Intron	rs535268249	NA
chr5:g.153648808T>C	*GALNT10*	NM_198321.4:c.160-25568T>C	Intron	rs559867144	0.00260
chr5:g.153657913T>C	*GALNT10*	NM_198321.4:c.160-16463T>C	Intron	rs546799926	0.00260
chr5:g.153662288C>T	*GALNT10*	NM_198321.4:c.160-12088C>G	Intron	rs183237694	NA
chr5:g.153666287G>C	*GALNT10*	NM_198321.4:c.160-8089G>C	Intron	rs573697057	0.00260
chr5:g.153667475A>T	*GALNT10*	NM_198321.4:c.160-6901A>T	Intron	rs6871973	NA
chr5:g.153667477A>T	*GALNT10*	NM_198321.3:c.160-6899A>T	Intron	rs910858062	NA
chr5:g.153671900G>C	*GALNT10*	NM_198321.4:c.160-2476G>C	Intron	rs184292339	0.00240
chr5:g.153673488G>A	*GALNT10*	NM_198321.4:c.160-888G>A	Intron	rs181150306	0.00240
chr5:g.153683961A>T	*GALNT10*	NM_198321.4:c.401+6322A>T	Intron	rs571475030	0.00260
chr5:g.153690771A>G	*GALNT10*	NM_198321.4:c.401+13132A>G	Intron	rs141277658	0.00459
chr5:g.153693707C>T	*GALNT10*	NM_198321.4:c.402-15425C>T	Intron	rs545092303	0.00260
chr5:g.153698634A>G	*GALNT10*	NM_198321.4:c.402-10498A>G	Intron	rs146832332	0.00459
chr5:g.153700813A>C	*GALNT10*	NM_198321.4:c.402-8319A>C	Intron	rs192993965	0.00459
chr5:g.153749290G>A	*GALNT10*	NM_198321.4:c.569-6547G>A	Intron	rs370025694	0.00240
chr5:g.153759857G>A	*GALNT10*	NM_198321.4:c.755-151G>A	Intron	rs554163168	0.00240
chr5:g.153767382C>A	*GALNT10*	NM_198321.4:c.1056+1392C>A	Intron	rs7710748	0.00160
chr5:g.153777436A>G	*GALNT10*	NM_198321.4:c.1057-6228A>G	Intron	rs4516909	NA
chr5:g.153784806G>C	*GALNT10*	NM_198321.4:c.1164+1035G>C	Intron	rs3776990	NA
chr5:g.153893914T>C	*HAND1-36090, MIR3141-81658*		Intergenic	rs373852532	0.00220
chr5:g.154006015A>G	*MIR3141-30383, MIR1303-59321*		Intergenic	rs6580096	NA
chr5:g.154007238A>G	*MIR3141-31606, MIR1303-58098*		Intergenic	rs56349590	NA
chr5:g.154023501G>A	*MIR3141-47869, MIR1303-41835*		Intergenic	rs557860959	0.00200
chr5:g.154023504C>T	*MIR3141-47872, MIR1303-41832*		Intergenic	rs577698296	0.00200

chr5:g.154027755C>T	*MIR3141-52123, MIR1303-37581*		Intergenic	rs967826622	NA
chr5:g.154029739A>G	*MIR3141-54107, MIR1303-35597*		Intergenic	rs909296266	NA
chr5:g.154039435C>T	*MIR3141-63803, MIR1303-25901*		Intergenic	rs80021095	NA
chr5:g.154178251G>A	*LARP1*	NM_015315.4:c.1147-900G>A	Intron	rs988433875	NA
chr5:g.154216843T>G	*FAXDC2*	NM_032385.5:c.48+848A>C	Intron	rs548587475	0.00060
chr5:g.154220113G>A	*FAXDC2*	NM_032385.4:c.1-2375C>T	Intron	rs933953537	NA
chr5:g.154256057G>A	*CNOT8*	NM_001301083:c.*1058G>A	3'UTR	rs959527222	NA
chr5:g.154257561G>A	*CNOT8-1209, GEMIN5-9415*		Downstream: 1209 bases from CNOT8	rs1280876739	NA
chr5:g.154257567C>T	*CNOT8-1215, GEMIN5-9409*		Downstream: 1215 bases from CNOT8		
chr5:g.154257568A>G	*CNOT8-1216, GEMIN5-9408*		Downstream: 1216 bases from CNOT8		
chr5:g.154257575A>C	*CNOT8-1223, GEMIN5-9401*		Downstream: 1223 bases from CNOT8		
chr5:g.154257578A>C	*CNOT8-1226, GEMIN5-9398*		Downstream: 1226 bases from CNOT8		
chr5:g.154403425C>T	*KIF4B-5733, SGCD-1350342*		Intergenic	rs934888217	NA
chr5:g.154461219G>C	*KIF4B-63527, SGCD-1292548*		Intergenic	rs34021636	NA
chr5:g.154469187G>A	*KIF4B-71495, SGCD-1284580*		Intergenic	rs185503701	0.00040
chr5:g.154485074G>A	*KIF4B-87382, SGCD-1268693*		Intergenic	rs1019505237	NA
chr5:g.154505263C>T	*KIF4B-107571, SGCD-1248504*		Intergenic	rs913632939	NA
chr5:g.154527802G>A	*KIF4B-130110, SGCD-1225965*		Intergenic	rs10515721	NA
chr5:g.154529813G>T	*KIF4B-132121, SGCD-1223954*		Intergenic	rs4958788	NA
chr5:g.154540532T>C	*KIF4B-142840, SGCD-1213235*		Intergenic	rs199876736	NA
chr5:g.154540599A>G	*KIF4B-142907, SGCD-1213168*		Intergenic	rs201206487	NA
chr5:g.154541247T>C	*KIF4B-143555, SGCD-1212520*		Intergenic	rs36161379	NA
chr5:g.154541362G>A	*KIF4B-143670, SGCD-1212405*		Intergenic	rs146676936	NA
chr5:g.154550111G>A	*KIF4B-152419, SGCD-1203656*		Intergenic	rs200167992	NA
chr5:g.154551187C>T	*KIF4B-153495, SGCD-1202580*		Intergenic	rs188861629	0.00060
chr5:g.154551768G>A	*KIF4B-154076, SGCD-1203161*		Intergenic	rs10477120	NA
chr5:g.154553224C>T	*KIF4B-155532, SGCD-1200543*		Intergenic	rs1317591453	NA
chr5:g.154553586G>C	*KIF4B-155894, SGCD-1200181*		Intergenic	rs10055952	NA
chr5:g.154553805G>C	*KIF4B-156113, SGCD-1199962*		Intergenic	rs79347230	NA
chr5:g.154618390G>A	*KIF4B-220698, SGCD-1135377*		Intergenic	rs576489001	0.00020
chr5:g.154645349C>T	*KIF4B-247657, SGCD-1108418*		Intergenic	rs539078412	0.00020
chr5:g.154660438G>T	*KIF4B-262746, SGCD-1093329*		Intergenic	rs1437265933	NA
chr5:g.154703390A>G	*KIF4B-305698, SGCD-1050377*		Intergenic	rs923707785	NA
chr5:g.154741767C>T	*KIF4B-344075, SGCD-1012000*		Intergenic	rs867620077	NA
chr5:g.154776009A>C	*KIF4B-378317, SGCD-977758*		Intergenic	rs17117406	NA
chr5:g.154779262A>G	*KIF4B-381570, SGCD-974505*		Intergenic		
chr5:g.154780946C>T	*KIF4B-383254, SGCD-972821*		Intergenic	rs1467991901	NA
chr5:g.154808705C>G	*KIF4B-411013, SGCD-945062*		Intergenic	rs1366580	NA
chr5:g.154826032T>A	*KIF4B-428340, SGCD-927735*		Intergenic	rs4958406	NA
chr5:g.154827837G>T	*KIF4B-430145, SGCD-925930*		Intergenic	rs4958813	NA
chr5:g.154831633T>A	*KIF4B-433941, SGCD-922134*		Intergenic	rs444796	NA

chr5:g.154831780G>A	*KIF4B-434088, SGCD-921987*		Intergenic	rs1452595411	NA
chr5:g.154836637T>C	*KIF4B-438945, SGCD-917130*		Intergenic		
chr5:g.154941337G>A	*KIF4B-543645, SGCD-812430*		Intergenic		
chr5:g.155081559G>A	*KIF4B-683867, SGCD-672208*		Intergenic	rs569301615	0.00040
chr5:g.155199458G>A	*KIF4B-801766, SGCD-554309*		Intergenic	rs565445872	0.00140
chr5:g.155202954G>T	*KIF4B-805262, SGCD-550813*		Intergenic	rs62382310	NA
chr5:g.155232523G>A	*KIF4B-834831, SGCD-521244*		Intergenic	rs572753903	0.00140
chr5:g.155234835A>G	*KIF4B-837143, SGCD-518932*		Intergenic	rs558657624	0.00140
chr5:g.155245848C>A	*KIF4B-848156, SGCD-507919*		Intergenic	rs2619723	NA
chr5:g.155286887C>T	*KIF4B-889195, SGCD-466880*		Intergenic	rs546140646	0.00140
chr5:g.155296187T>A	*KIF4B-898495, SGCD-457580*		Intergenic		
chr5:g.155308463A>G	*KIF4B-910771, SGCD-445304*		Intergenic	rs72797605	0.00240
chr5:g.155309756T>G	*KIF4B-912064, SGCD-444011*		Intergenic	rs72797609	0.00419
chr5:g.155315211C>A	*KIF4B-917519, SGCD-438556*		Intergenic	rs72797614	NA
chr5:g.155329291C>T	*KIF4B-931599, SGCD-424476*		Intergenic		
chr5:g.155369332G>C	*KIF4B-971640, SGCD-384435*		Intergenic	rs74461277	NA
chr5:g.155369336G>C	*KIF4B-971644, SGCD-384431*		Intergenic	rs112124160	NA
chr5:g.155369340G>C	*KIF4B-971648, SGCD-384427*		Intergenic	rs62380668	NA
chr5:g.155612862G>T	*KIF4B-1215170, SGCD-140905*		Intergenic	rs12234031	NA
chr5:g.155612864G>T	*KIF4B-1215172, SGCD-140903*		Intergenic	rs1460449662	NA
chr5:g.155612866G>T	*KIF4B-1215174, SGCD-140901*		Intergenic	rs1375952509	NA
chr5:g.155612868G>T	*KIF4B-1215176, SGCD-140899*		Intergenic	rs1278365947	NA
chr5:g.155612870G>T	*KIF4B-1215178, SGCD-140897*		Intergenic	rs1218611266	NA
chr5:g.155615478C>T	*KIF4B-1217786, SGCD-138289*		Intergenic	rs182010764	0.00319
chr5:g.155641857A>G	*KIF4B-1244165, SGCD-111910*		Intergenic	rs552281104	0.00020
chr5:g.155662082T>G	*KIF4B-1264390, SGCD-91685*		Intergenic	rs187081451	0.00300
chr5:g.155709381T>A	*KIF4B-1311689, SGCD-44386*		Intergenic	rs191528759	0.00300
chr5:g.155737475G>A	*KIF4B-1339783, SGCD-16292*		Intergenic	rs987870982	NA
chr5:g.155744163A>G	*KIF4B-1346471, SGCD-9604*		Intergenic	rs10463088	NA
chr5:g.155806419C>A	*SGCD*	NM_000337.5:c.192+34732C>A	Intron	rs182667129	0.00080
chr5:g.155831187G>A	*SGCD*	NM_000337.5:c.192+59500G>A	Intron	rs62380757	NA
chr5:g.155864467C>T	*SGCD*	NM_000337.5:c.193-71144C>T	Intron	rs17053557	NA
chr5:g.155909277G>A	*SGCD*	NM_000337.5:c.193-26334G>A	Intron	rs1476672685	NA
chr5:g.155920130A>G	*SGCD*	NM_000337.5:c.193-15481A>G	Intron	rs572953056	0.00100
chr5:g.155966070G>A	*SGCD*	NM_000337.5:c.294+30358G>A	Intron	rs182560226	0.00040
chr5:g.156044232C>T	*SGCD*	NM_000337.5:c.502+22171C>T	Intron	rs542015108	0.00040
chr5:g.156088579C>T	*SGCD*	NM_000337.5:c.575+14033C>T	Intron	rs200096709	NA
chr5:g.156088895T>C	*SGCD*	NM_000337.5:c.575+14349T>C	Intron	rs373889967	NA
chr5:g.156092306C>T	*SGCD*	NM_000337.5:c.575+17760C>T	Intron	rs376138762	0.00020
chr5:g.156096057C>T	*SGCD*	NM_000337.5:c.575+21511C>T	Intron	rs543889875	0.00060
chr5:g.156145126T>C	*SGCD*	NM_000337.5:c.576-39466T>C	Intron	rs557771871	0.00020

chr5:g.156225861C>G	*SGCD-31063, PPP1R2P3-51636*		Intergenic	rs993835441	NA
chr5:g.156253009G>T	*SGCD-58211, PPP1R2P3-24488*		Intergenic		
chr5:g.156291001G>A	*PPP1R2P3-11462, TIMD4-55292*		Intergenic	rs6884957	NA
chr5:g.156291803G>T	*PPP1R2P3-12264, TIMD4-54490*		Intergenic	rs17053988	NA
chr5:g.156383453G>A	*TIMD4*	NM_138379.2:c.59-1686C>T	intron	rs544684502	NA
chr5:g.156426022G>A	*TIMD4-35756, HAVCR1-30402*		Intergenic	rs72805137	NA
chr5:g.156482489T>G	*HAVCR1*	NM_012206.3:c.102A>C, NP_001166864.1:p.Leu34=	Synonymous	rs202050012	NA
chr5:g.156551664A>G	*HAVCR2-15416, MED7-13787*		Intergenic		
chr5:g.156654872G>A	*ITK*	NM_005546.3:c.648-434G>A	Intron	rs546812934	0.00020
chr5:g.156726657C>G	*CYFIP2*	NM_014376.3:c.388-1066C>G	Intron		
chr5:g.156808560T>G	*CYFIP2*	NM_014376.3:c.3040-1038T>G	Intron	rs62387477	NA
chr5:g.156808562A>G	*CYFIP2*	NM_014376.3:c.3040-1036A>G	Intron	rs1416044956	NA
chr5:g.156836988A>C	*LOC102724404*	NR_136204.1:n.94-33474T>G	Intron	rs569078124	0.00020
chr5:g.156847036G>A	*LOC102724404*	NR_136204.1:n.93+39980C>T	Intron	rs62387540	NA
chr5:g.156855370A>C	*LOC102724404*	NR_136204.1:n.93+31646T>G	Intron	rs567314607	0.00080
chr5:g.156889979C>A	*NIPAL4*	NM_001099287.1:c.224-123C>A	intron	rs568504531	0.00020
chr5:g.156889980G>A	*NIPAL4*	NM_001099287.1:c.224-122G>A	Intron	rs537809300	0.00020
chr5:g.156913340C>T	*ADAM19*	NM_033274.4:c.2550+1933G>A	intron		
chr5:g.157033520C>T	*ADAM19-30689, SOX30-19167*		Intergenic	rs542613817	0.00140
chr5:g.157042404T>C	*ADAM19-39573, SOX30-10283*		Intergenic	rs7717811	NA
chr5:g.157046823G>A	*ADAM19-43992, SOX30-5864*		Intergenic	rs373746569	0.00300
chr5:g.157046824C>A	*ADAM19-43993, SOX30-5863*		Intergenic	rs375820473	0.00300
chr5:g.157057221C>T	*SOX30*	NM_178424.1:c.1881-3492G>A	Intron		
chr5:g.157070731T>C	*SOX30*	NM_178424.1:c.1387+2914A>G	Intron	rs376901964	0.00060
chr5:g.157074259C>T	*SOX30*	NM_178424.1:c.1208-435G>A	Intron	rs376449866	0.00080
chr5:g.157079143C>T	*SOX30*	NM_178424.1:c.-57G>A	5'UTR	rs369841660	0.00200
chr5:g.157098892A>G	*C5orf52*	NM_001145132.1:c.212+58A>G	Intron	rs372789928	0.00200
chr5:g.157119234G>C	*C5orf52-12072, THG1L-39163*		Intergenic	rs370793496	NA
chr5:g.157122521C>T	*C5orf52-15359, THG1L-35876*		Intergenic	rs558576870	0.00120
chr5:g.157132145G>A	*C5orf52-24983, THG1L-26252*		Intergenic		
chr5:g.157146819T>G	*C5orf52-39657, THG1L-11578*		Intergenic	rs895324764	NA
chr5:g.157146821T>G	*C5orf52-39659, THG1L-11576*		Intergenic	rs1339856593	NA
chr5:g.157146824T>G	*C5orf52-39662, THG1L-11573*		Intergenic	rs995040585	NA
chr5:g.157146827T>G	*C5orf52-39665, THG1L-11570*		Intergenic	rs1027812395	NA
chr5:g.157186414A>G	*LSM11*	NM_173491.3:c.*4142A>G	3'UTR	rs548208300	0.00120
chr5:g.157200801C>G	*LSM11-13084, CLINT1-11950*		Intergenic	rs532351127	0.00100
chr5:g.157210703G>A	*LSM11-22986, CLINT1-2048*		Downstream 2048 bases from CLINT1	rs558294806	0.00100
chr5:g.157236283G>A	*CLINT1*	NM_014666.3:c.695+353C>T	Intron	rs1397075422	NA
chr5:g.157257806A>C	*CLINT1*	NM_014666.3:c.42-13251T>G	Intron	rs3104696	NA
chr5:g.157312749G>A	*CLINT1-26566, LINC02227-434963*		Intergenic		

chr5:g.157323456C>T	*CLINT1-37273, LINC02227-424256*		Intergenic	rs556708481	0.00060
chr5:g.157331711T>G	*CLINT1-45528, LINC02227-416001*		Intergenic		
chr5:g.157336221G>A	*CLINT1-50038, LINC02227-411491*		Intergenic	rs1043603039	NA
chr5:g.157366483T>C	*CLINT1-80300, LINC02227-381229*		Intergenic		
chr5:g.157381826C>A	*CLINT1-95643, LINC02227-365886*		Intergenic	rs10070111	NA
chr5:g.157382314C>A	*CLINT1-96131, LINC02227-365398*		Intergenic	rs979079186	NA
chr5:g.157382997C>T	*CLINT1-96814, LINC02227-364715*		Intergenic	rs370550775	NA
chr5:g.157383001C>T	*CLINT1-96818, LINC02227-364711*		Intergenic	rs796182617	NA
chr5:g.157383005C>T	*CLINT1-96822, LINC02227-364707*		Intergenic	rs142889331	NA
chr5:g.157405602A>T	*CLINT1-119419, LINC02227-342110*		Intergenic	rs12655936	NA
chr5:g.157478347T>C	*CLINT1-192164, LINC02227-269365*		Intergenic	rs548743520	0.00180
chr5:g.157480147G>A	*CLINT1-193964, LINC02227-267565*		Intergenic	rs983760939	NA
chr5:g.157514187C>G	*CLINT1-228004, LINC02227-233525*		Intergenic		
chr5:g.157522188T>C	*CLINT1-236005, LINC02227-225524*		Intergenic	rs568230495	0.00020
chr5:g.157612083G>A	*CLINT1-325900, LINC02227-135629*		Intergenic	rs941745811	NA
chr5:g.157718664G>A	*CLINT1-432481, LINC02227-29048*		Intergenic		
chr5:g.157787457A>G	*LINC02227*	NR_109888.1:n.446+1953T>C	Intron		
chr5:g.157839798T>A	*LINC02227-3017, EBF1-283125*		Intergenic	rs114853220	0.00339
chr5:g.157893360G>C	*LINC02227-56579, EBF1-229563*		Intergenic	rs4704942	NA
chr5:g.157905774C>T	*LINC02227-68993, EBF1-217149*		Intergenic		
chr5:g.157915694T>C	*LINC02227-78913, EBF1-207229*		Intergenic	rs369888159	NA
chr5:g.157937710C>T	*LINC02227-100929, EBF1-185213*		Intergenic	rs184761930	0.00479
chr5:g.158016847C>G	*LINC02227-180066, EBF1-106076*		Intergenic	rs56360586	NA
chr5:g.158088211G>A	*LINC02227-251430, EBF1-34712*		Intergenic	rs188089430	0.00280
chr5:g.158100133C>A	*LINC02227-263352, EBF1-22790*		Intergenic	rs182113227	0.00240
chr5:g.158124682A>G	*EBF1*	NM_024007.4:c.*1437T>C	3'UTR		
chr5:g.158138668G>T	*EBF1*	NM_024007.4:c.1549+494C>A	Intron	rs1430067936	NA
chr5:g.158186924C>A	*EBF1*	NM_024007.4:c.1036+17497G>T	Intron		
chr5:g.158216636C>T	*EBF1*	NM_024007.4:c.909+6717G>A	Intron	rs1444970891	NA
chr5:g.158232080T>C	*EBF1*	NM_024007.4:c.779-8597A>G	Intron		
chr5:g.158249275C>T	*EBF1*	NM_024007.4:c.778+909G>A	Intron	rs9313792	NA
chr5:g.158256737T>A	*EBF1*	NM_024007.4:c.637-6412A>T	Intron		
chr5:g.158275302C>T	*EBF1*	NM_024007.4:c.555-8184G>A	Intron	rs6885117	NA
chr5:g.158282011C>T	*EBF1*	NM_024007.4:c.555-14893G>A	Intron	rs75233936	NA
chr5:g.158305560G>A	*EBF1*	NM_024007.4:c.555-38442C>T	Intron	rs904936137	NA
chr5:g.158328657A>T	*EBF1*	NM_024007.4:c.555-61539T>A	Intron	rs981292759	NA
chr5:g.158352322G>A	*EBF1*	NM_024007.4:c.555-85204C>T	Intron	rs1242826769	NA
chr5:g.158387880C>T	*EBF1*	NM_024007.4:c.554+112524G>A	Intron	rs940117282	NA
chr5:g.158474094T>C	*EBF1*	NM_024007.4:c.554+26310A>G	Intron	rs940083921	NA
chr5:g.158540123C>T	*LINC02202*	NR_109890.1:n.496+1117C>T	Intron		
chr5:g.158556861G>C	*LINC02202-12375, RNF145-27556*		Intergenic	rs80337959	NA

chr5:g.158620538T>C	*RNF145*	NM_144726.2:c.377+1186A>G	Intron		
chr5:g.158627469G>A	*RNF145*	NM_144726.2:c.268+2973C>T	Intron	rs112026617	0.00399
chr5:g.158659129G>A	*LINC01932*	NR_134261.1:n.105+2804G>A	Intron	rs117713770	0.00419
chr5:g.158667860C>G	*LINC01932*	NR_134261.1:n.106-3808C>G	Intron		
chr5:g.158695031A>G	*UBLCP1*	NM_145049.4:c.-46-847A>G	Intron	rs78412473	0.00359
chr5:g.158774698C>G	*LOC285626*	NR_037889.1:n.835+85C>G	Intron	rs138525382	NA
chr5:g.158799971C>A	*LOC285626-10129, LINC01845-75593*		Intergenic		
chr5:g.158818857C>T	*LOC285626-29015, LINC01845-56707*		Intergenic	rs1211266043	NA
chr5:g.158879286C>T	*LINC01845*	NR_027110.1:n.1111+304G>A	Intron	rs1040460208	NA
chr5:g.158880340G>T	*LINC01845*	NR_027110.1:n.899+57C>A	Intron	rs4921229	NA
chr5:g.158992025G>A	*LINC01845-98741, LINC01847-211757*		Intergenic	rs1014097412	NA
chr5:g.158997558C>T	*LINC01845-104274, LINC01847-206224*		Intergenic	rs575097540	0.00220
chr5:g.159000670G>T	*LINC01845-107386, LINC01847-203112*		Intergenic	rs955808718	NA
chr5:g.159000672C>T	*LINC01845-107388, LINC01847-203110*		Intergenic	rs1350749369	NA
chr5:g.159026711A>G	*LINC01845-133427, LINC01847-177071*		Intergenic	rs12019308	NA
chr5:g.159027369G>A	*LINC01845-134085, LINC01847-176413*		Intergenic	rs7708583	NA
chr5:g.159036818G>A	*LINC01845-143534, LINC01847-166964*		Intergenic	rs7732774	NA
chr5:g.159048548C>G	*LINC01845-155264, LINC01847-155234*		Intergenic	rs551775979	0.00240
chr5:g.159054927C>A	*LINC01845-161643, LINC01847-148855*		Intergenic	rs371292854	0.00499
chr5:g.159055848A>G	*LINC01845-162564, LINC01847-147934*		Intergenic	rs576386711	0.00120
chr5:g.159061262G>A	*LINC01845-167978, LINC01847-142520*		Intergenic	rs12153266	NA
chr5:g.159110169A>G	*LINC01845-216885, LINC01847-93613*		Intergenic	rs557875890	0.00359
chr5:g.159115640A>C	*LINC01845-222356, LINC01847-88142*		Intergenic	rs74684835	NA
chr5:g.159144464T>C	*LINC01845-251180, LINC01847-59318*		Intergenic	rs377307848	0.00100
chr5:g.159206050G>A	*LINC01847*	NR_109891.1:n.3692-356C>T	Intron		
chr5:g.159236366A>T	*LINC01847*	NR_109891.1:n.3692-30672T>A	Intron	rs533741243	0.00040
chr5:g.159262961G>A	*LINC01847*	NR_109891.1:n.3691+30918C>T	Intron	rs10036026	NA
chr5:g.159308316C>G	*LINC01847-9925, ADRA1B-35424*		Intergenic	rs202026796	NA
chr5:g.159318016A>C	*LINC01847-19625, ADRA1B-25724*		Intergenic	rs897899602	NA
chr5:g.159359417A>G	*ADRA1B*	NM_000679.3:c.949+14556A>G	Intron	rs546121198	0.00060
chr5:g.159366287T>A	*ADRA1B*	NM_000679.3:c.949+21426T>A	Intron	rs9313832	NA
chr5:g.159366321T>C	*ADRA1B*	NM_000679.3:c.949+21460T>C	Intron	rs1156598780	NA
chr5:g.159366323T>C	*ADRA1B*	NM_000679.3:c.949+21462T>C	Intron	rs372472643	NA
chr5:g.159366325T>C	*ADRA1B*	NM_000679.3:c.949+21464T>C	Intron	rs377003381	NA
chr5:g.159464802T>A	*TTC1*	NM_003314.2:c.504+992T>A	Intron	rs977115735	NA
chr5:g.159472456A>C	*TTC1*	NM_003314.2:c.541+2280A>C	Intron	rs902955887	NA
chr5:g.159486354T>C	*TTC1*	NM_003314.2:c.746-5585T>C	Intron	rs4623188	NA
chr5:g.159502182T>C	*PWWP2A-707*		Downstream 707 bases from PWWP2A	rs4446508	NA
chr5:g.159503427A>G	*PWWP2A*	NM_052927.2:c.*1735T>C	3'UTR	rs569368848	0.00040
chr5:g.159524941G>T	*PWWP2A*	NM_052927.2:c.585-3869C>A	Intron	rs538691849	0.00479
chr5:g.159536115T>C	*PWWP2A*	NM_052927.2:c.584+9697A>G	Intron	rs2546987	NA

chr5:g.159589508C>T	*PWWP2A-43056, FABP6-24866*		Intergenic	rs766717991	NA
chr5:g.159611444G>A	*PWWP2A-64992, FABP6-2930*		Upstream 2930 bases from FABP6	rs79664684	NA
chr5:g.159637492G>A	*FABP6*	NM_001040442.1:c.52-3251G>A	Intron	rs545304023	0.00479
chr5:g.159731191G>T	*CCNJL*	NM_024565.6:c.66+7674C>A	Intron	rs12658716	NA
chr5:g.159753756T>C	*CCNJL-14149, C1QTNF2-21019*	NR_131769.1:n.149-11244A>G	Intron	rs547305069	0.00040
chr5:g.159820391C>T	*ZBED8*	NM_022090.4:c.*322G>A	3'UTR		
chr5:g.159823728G>A	*ZBED8*	NM_022090.4:c.-49-1182C>T	Intron	rs949971500	NA
chr5:g.159823729C>A	*ZBED8*	NM_022090.4:c.-49-1183G>T	Intron	rs954381333	NA
chr5:g.159823734C>A	*ZBED8*	NM_022090.4:c.-49-1188G>T	Intron	rs1009039398	NA
chr5:g.159928484C>G	*MIR3142HG-14051, ATP10B-61643*		Intergenic	rs2431102	NA
chr5:g.159993433C>T	*ATP10B*	NM_025153.2:c.3939-526G>A	Intron	rs994067133	NA
chr5:g.159995330T>C	*ATP10B*	NM_025153.2:c.3938+1173A>G	Intron	rs7704156	NA
chr5:g.160025107C>T	*ATP10B*	NM_025153.2:c.3564+670G>A	Intron	rs544141299	0.00220
chr5:g.160093744G>A	*ATP10B*	NM_025153.2:c.675+3726C>T	Intron	rs187844509	0.00300
chr5:g.160093745C>A	*ATP10B*	NM_025153.2:c.675+3725G>T	Intron	rs540344061	0.00300
chr5:g.160152885C>A	*ATP10B*	NM_025153.2:c.-330-8844G>T	Intron	rs73798530	NA
chr5:g.160170908C>A	*ATP10B*	NM_025153.2:c.-330-26867G>T	Intron		
chr5:g.160255400G>A	*ATP10B*	NM_025153.2:c.-576+23548C>T	Intron	rs56055937	NA
chr5:g.160265114T>C	*ATP10B*	NM_025153.2:c.-576+13834A>G	Intron	rs541948582	0.00020
chr5:g.160267509T>A	*ATP10B*	NM_025153.2:c.-576+11439A>T	Intron		
chr5:g.160415929C>A	*LINC02159-50296, GABRB2-299507*		Intergenic	rs3914603	NA
chr5:g.160479240G>C	*LINC02159-113607, GABRB2-236196*		Intergenic		
chr5:g.160543292G>A	*LINC02159-177659, GABRB2-172144*		Intergenic	rs1273305415	NA
chr5:g.160543296G>A	*LINC02159-177663, GABRB2-172140*		Intergenic	rs974231178	NA
chr5:g.160607085G>A	*LINC02159-241452, GABRB2-108351*		Intergenic	rs694676	NA
chr5:g.160618455G>A	*LINC02159-252822, GABRB2-96981*		Intergenic	rs371455770	NA
chr5:g.160643217C>T	*LINC02159-277584, GABRB2-72219*		Intergenic	rs188609885	NA
chr5:g.160666277C>T	*LINC02159-300644, GABRB2-49159*		Intergenic	rs1398749348	NA
chr5:g.160678482C>T	*LINC02159-312849, GABRB2-36954*		Intergenic	rs151215990	0.00399
chr5:g.160860375G>A	*GABRB2*	NM_021911.2:c.459-22312C>T	Intron	rs1052904066	NA
chr5:g.160881766G>A	*GABRB2*	NM_021911.2:c.458+4864C>T	Intron	rs146198359	0.00280
chr5:g.160888862T>A	*GABRB2*	NM_021911.2:c.238-2012A>T	Intron	rs149477433	0.00280
chr5:g.160935847C>T	*GABRB2*	NM_021911.2:c.237+36386G>A	Intron	rs145763914	NA
chr5:g.161037697C>T	*GABRB2-62567, GABRA6-74961*		Intergenic	rs11948316	NA
chr5:g.161058965C>A	*GABRB2-83835, GABRA6-53693*		Intergenic	rs6880823	NA
chr5:g.161073408T>C	*GABRB2-98278, GABRA6-39250*		Intergenic	rs6885619	NA
chr5:g.161098077G>A	*GABRB2-122947, GABRA6-14581*		Intergenic	rs142775258	NA
chr5:g.161123514C>T	*GABRA6*	NM_000811.2:c.1086+4308C>T	Intron		
chr5:g.161158362C>T	*GABRA6-28764, GABRA1-115835*		Intergenic	rs72817443	NA
chr5:g.161185469A>T	*GABRA6-55871, GABRA1-88728*		Intergenic	rs142243581	0.00260
chr5:g.161194268A>C	*GABRA6-64670, GABRA1-79929*		Intergenic	rs902748329	NA

chr5:g.161263172A>G	*GABRA6-133574, GABRA1-11025*		Intergenic	rs796654891	NA
chr5:g.161297795G>A	*GABRA1*	NM_000806.5:c.256-2328G>A	Intron	rs6866984	NA
chr5:g.161313741C>A	*GABRA1*	NM_000806.5:c.703+4034C>A	Intron	rs3980416	NA
chr5:g.161320054T>A	*GABRA1*	NM_000806.5:c.856+1998T>A	Intron	rs72819329	NA
chr5:g.161341634C>A	*LINC01202*	NR_126372.1:n.326-3795G>T	Intron		
chr5:g.161594382C>G	*GABRG2-11837, CCNG1-1270195*		Intergenic	rs434689	NA
chr5:g.161604731T>C	*GABRG2-22186, CCNG1-1259846*		Intergenic		
chr5:g.161635812A>C	*GABRG2-53267, CCNG1-1228765*		Intergenic	rs1012866	NA
chr5:g.161639740G>A	*GABRG2-57195, CCNG1-1224837*		Intergenic	rs169794	NA
chr5:g.161672988A>G	*GABRG2-90443, CCNG1-1191589*		Intergenic	rs2963176	NA
chr5:g.161703925C>A	*GABRG2-121380, CCNG1-1160652*		Intergenic		
chr5:g.161754621A>T	*GABRG2-172076, CCNG1-1109956*		Intergenic	rs4438914	NA
chr5:g.161788156G>T	*GABRG2-205611, CCNG1-1076421*		Intergenic	rs947556706	NA
chr5:g.161807660T>C	*GABRG2-225115, CCNG1-1056917*		Intergenic	rs60842750	NA
chr5:g.161811698C>T	*GABRG2-229153, CCNG1-1052879*		Intergenic	rs4407660	NA
chr5:g.161825456T>G	*GABRG2-242911, CCNG1-1039121*		Intergenic	rs34223698	NA
chr5:g.161830984G>A	*GABRG2-248439, CCNG1-1033593*		Intergenic	rs890659546	NA
chr5:g.161849261C>T	*GABRG2-266716, CCNG1-1015316*		Intergenic	rs940038645	NA
chr5:g.161850169A>G	*GABRG2-267624, CCNG1-1014408*		Intergenic	rs12514135	NA
chr5:g.161850286G>C	*GABRG2-267741, CCNG1-1014291*		Intergenic	rs12516514	NA
chr5:g.161890544A>T	*GABRG2-307999, CCNG1-974033*		Intergenic	rs371753174	NA
chr5:g.161903978C>A	*GABRG2-321433, CCNG1-960599*		Intergenic	rs7443053	NA
chr5:g.161905912A>T	*GABRG2-323367, CCNG1-958665*		Intergenic	rs372878735	NA
chr5:g.161906075G>A	*GABRG2-323530, CCNG1-958502*		Intergenic	rs78760325	NA
chr5:g.161906175A>C	*GABRG2-323630, CCNG1-958402*		Intergenic	rs13185833	NA
chr5:g.161914205T>C	*GABRG2-331660, CCNG1-950372*		Intergenic	rs553577751	0.00100
chr5:g.161929892A>C	*GABRG2-347347, CCNG1-934685*		Intergenic		
chr5:g.161996708G>T	*GABRG2-414163, CCNG1-867869*		Intergenic		
chr5:g.162000806G>A	*GABRG2-418261, CCNG1-863771*		Intergenic	rs112071915	NA
chr5:g.162042528G>A	*GABRG2-459983, CCNG1-822049*		Intergenic		
chr5:g.162131595G>A	*GABRG2-549050, CCNG1-732982*		Intergenic		
chr5:g.162140827T>C	*GABRG2-558282, CCNG1-723750*		Intergenic	rs201833919	NA
chr5:g.162260100G>A	*GABRG2-677555, CCNG1-604477*		Intergenic	rs10051007	0.00419
chr5:g.162341059C>G	*GABRG2-758514, CCNG1-523518*		Intergenic	rs369613283	NA
chr5:g.162382486C>A	*GABRG2-799941, CCNG1-482091*		Intergenic	rs12651726	NA
chr5:g.162397923A>T	*GABRG2-815378, CCNG1-466654*		Intergenic	rs1824274	NA
chr5:g.162399611T>C	*GABRG2-817066, CCNG1-464966*		Intergenic		
chr5:g.162404172C>T	*GABRG2-821627, CCNG1-460405*		Intergenic	rs191193184	0.00020
chr5:g.162471687A>G	*GABRG2-889142, CCNG1-392890*		Intergenic	rs192899383	0.00020
chr5:g.162486588C>T	*GABRG2-904043, CCNG1-377989*		Intergenic	rs542056066	0.00040
chr5:g.162516773T>C	*GABRG2-934228, CCNG1-347804*		Intergenic	rs1021923907	NA

chr5:g.162517226G>T	*GABRG2-934681, CCNG1-347351*		Intergenic	rs4385227	NA
chr5:g.162523202G>A	*GABRG2-940657, CCNG1-341375*		Intergenic	rs1290779587	NA
chr5:g.162523203C>A	*GABRG2-940658, CCNG1-341374*		Intergenic	rs112510221	NA
chr5:g.162579029G>A	*GABRG2-996484, CCNG1-285548*		Intergenic	rs1489551069	NA
chr5:g.162604944G>C	*GABRG2-1022399, CCNG1-259633*		Intergenic	rs12659321	NA
chr5:g.162621342G>C	*GABRG2-1038797, CCNG1-243235*		Intergenic	rs2913576	NA
chr5:g.162622438A>T	*GABRG2-1039893, CCNG1-242139*		Intergenic	rs7711464	NA
chr5:g.162638398A>C	*GABRG2-1055853, CCNG1-226179*		Intergenic	rs558692058	0.00020
chr5:g.162670285T>C	*GABRG2-1087740, CCNG1-194292*		Intergenic	rs1018772468	NA
chr5:g.162705618T>C	*GABRG2-1123073, CCNG1-158959*		Intergenic	rs550358540	0.00280
chr5:g.162762927G>A	*GABRG2-1180382, CCNG1-101650*		Intergenic	rs546962283	0.00120
chr5:g.162780313G>A	*GABRG2-1197768, CCNG1-84264*		Intergenic	rs571339584	0.00120
chr5:g.162843900C>A	*GABRG2-1261355, CCNG1-20677*		Intergenic	rs577237049	0.00260
chr5:g.162850051A>C	*GABRG2-1267506, CCNG1-14526*		Intergenic	rs778152063	NA
chr5:g.162891894G>A	*HMMR*	NM_001142556.1:c.225+86G>A	Intron	rs3756648	NA
chr5:g.162949962G>A	*MAT2B-3603, LINC02143-925466*		Downstream 3603 bases from MAT2B	rs13358512	NA
chr5:g.162976555A>G	*MAT2B-30196, LINC02143-898873*		Intergenic	rs542622633	0.00060
chr5:g.162979518C>T	*MAT2B-33159, LINC02143-895910*		Intergenic	rs529919155	0.00060
chr5:g.163003572G>T	*MAT2B-57213, LINC02143-871856*		Intergenic	rs571874375	0.00060
chr5:g.163009287T>G	*MAT2B-62928, LINC02143-866141*		Intergenic	rs34829092	NA
chr5:g.163071140G>A	*MAT2B-124781, LINC02143-804288*		Intergenic	rs17062118	NA
chr5:g.163130692G>C	*MAT2B-184333, LINC02143-744736*		Intergenic		
chr5:g.163226193G>A	*MAT2B-279834, LINC02143-649235*		Intergenic		
chr5:g.163276946G>C	*MAT2B-330587, LINC02143-598482*		Intergenic	rs147657664	0.00300
chr5:g.163305581T>C	*MAT2B-359222, LINC02143-569847*		Intergenic	rs555536612	0.00260
chr5:g.163352915C>A	*MAT2B-406556, LINC02143-522513*		Intergenic	rs17052725	NA
chr5:g.163398392G>C	*MAT2B-452033, LINC02143-477036*		Intergenic	rs899582063	NA
chr5:g.163398393T>A	*MAT2B-452034, LINC02143-477035*		Intergenic	rs997028106	NA
chr5:g.163398394G>A	*MAT2B-452035, LINC02143-477034*		Intergenic	rs1032891871	NA
chr5:g.163398406G>A	*MAT2B-452047, LINC02143-477022*		Intergenic	rs958329681	NA
chr5:g.163398407T>G	*MAT2B-452048, LINC02143-477021*		Intergenic	rs1010213092	NA
chr5:g.163416845G>A	*MAT2B-470486, LINC02143-458583*		Intergenic	rs559156988	0.00260
chr5:g.163501442A>T	*MAT2B-555083, LINC02143-373986*		Intergenic	rs576722609	0.00120
chr5:g.163546608G>T	*MAT2B-600249, LINC02143-328820*		Intergenic		
chr5:g.163597505G>A	*MAT2B-651146, LINC02143-277923*		Intergenic	rs564132522	0.00260
chr5:g.163635164T>C	*MAT2B-688805, LINC02143-240264*		Intergenic		
chr5:g.163805546C>G	*MAT2B-859187, LINC02143-69882*		Intergenic	rs4868876	NA
chr5:g.163900319A>C	*LOC102546299*	NR_105065.1:n.190+2845A>C	Intron	rs796614819	NA
chr5:g.163909914G>A	*LOC102546299*	NR_105065.1:n.191-3766G>A	Intron	rs7734642	NA
chr5:g.163925532G>A	*LOC102546299*	NR_105065.1:n.258+11785G>A	Intron	rs76817691	NA
chr5:g.163925918G>C	*LOC102546299*	NR_105065.1:n.258+12171G>C	Intron	rs4998980	NA

chr5:g.163925920C>G	*LOC102546299*	NR_105065.1:n.258+12173C>G	Intron	rs867457180	NA
chr5:g.163932005T>C	*LOC102546299*	NR_105065.1:n.258+18258T>C	Intron	rs1155811	NA
chr5:g.163934480G>A	*LOC102546299*	NR_105065.1:n.258+20733G>A	Intron	rs75187588	NA
chr5:g.163942220G>A	*LOC102546299*	NR_105065.1:n.259-26856G>A	Intron	rs867159397	NA
chr5:g.163963365T>C	*LOC102546299*	NR_105065.1:n.259-5711T>C	Intron	rs9687870	NA
chr5:g.164013582G>A	*LOC102546299-43593*		Intergenic	rs371982689	0.00280
chr5:g.164018570G>A	*LOC102546299-48581*		Intergenic	rs372490719	0.00280
chr5:g.164022279G>A	*LOC102546299-52290*		Intergenic	rs184806747	0.00280
chr5:g.164030558G>A	*LOC102546299-60569*		Intergenic	rs868031945	NA
chr5:g.164135425A>G	*LOC102546299-165436*		Intergenic	rs188265147	0.00280
chr5:g.164137357T>C	*LOC102546299-167368*		Intergenic	rs376733552	0.00280
chr5:g.164147261G>A	*LOC102546299-177272*		Intergenic	rs1391118982	NA
chr5:g.164147262G>A	*LOC102546299-177273*		Intergenic		
chr5:g.164147265C>A	*LOC102546299-177276*		Intergenic	rs934789224	NA
chr5:g.164150430C>G	*LOC102546299-180441*		Intergenic	rs181632014	0.00300
chr5:g.164151498T>G	*LOC102546299-181509*		Intergenic	rs369677330	0.00300
chr5:g.164152087G>A	*LOC102546299-182098*		Intergenic	rs545676309	0.00020
chr5:g.164165042C>T	*LOC102546299-195053*		Intergenic	rs1330347080	NA
chr5:g.164180514C>A	*LOC102546299-210525*		Intergenic	rs151198231	0.00419
chr5:g.164229033C>T	*LOC102546299-259044*		Intergenic	rs184595236	0.00399
chr5:g.164331839A>G	*LOC102546299-361850, LINC01947-2000388*		Intergenic	rs562200843	0.00100
chr5:g.164332650A>G	*LOC102546299-362661, LINC01947-1999577*		Intergenic	rs866459945	NA
chr5:g.164444475G>A	*LOC102546299-474486, LINC01947-1887752*		Intergenic	rs560406148	0.00020
chr5:g.164450234T>C	*LOC102546299-480245, LINC01947-1881993*		Intergenic	rs568096108	0.00020
chr5:g.164494688T>A	*LOC102546299-524699, LINC01947-1837539*		Intergenic	rs867948062	NA
chr5:g.164494725C>G	*LOC102546299-524736, LINC01947-1837502*		Intergenic	rs868112786	NA
chr5:g.164509341A>T	*LOC102546299-539352, LINC01947-1822886*		Intergenic	rs865783127	NA
chr5:g.164509567C>T	*LOC102546299-539578, LINC01947-1822660*		Intergenic	rs62383893	NA
chr5:g.164510257T>A	*LOC102546299-540268, LINC01947-1821970*		Intergenic	rs566791231	0.00020
chr5:g.164559524G>A	*LOC102546299-589535, LINC01947-1772703*		Intergenic	rs569620166	0.00020
chr5:g.164580377C>T	*LOC102546299-610388, LINC01947-1751850*		Intergenic	rs867735570	NA
chr5:g.164584757C>G	*LOC102546299-614768, LINC01947-1747470*		Intergenic	rs561177290	0.00020
chr5:g.164608570C>T	*LOC102546299-638581, LINC01947-1723657*		Intergenic	rs759996066	NA
chr5:g.164662863T>C	*LOC102546299-692874, LINC01947-1669364*		Intergenic	rs537838546	0.00120
chr5:g.164696202G>C	*LOC102546299-726213, LINC01947-1636025*		Intergenic	rs151048399	NA
chr5:g.164791076G>A	*LOC102546299-821087, LINC01947-1541151*		Intergenic	rs1187779826	NA
chr5:g.164841024G>C	*LOC102546299-871035, LINC01947-1491203*		Intergenic	rs12109346	NA
chr5:g.164846440A>G	*LOC102546299-876451, LINC01947-1485787*		Intergenic	rs13157743	NA
chr5:g.164846489G>A	*LOC102546299-876500, LINC01947-1485738*		Intergenic	rs13178861	NA
chr5:g.164848801G>A	*LOC102546299-878812, LINC01947-1483426*		Intergenic	rs868140835	NA
chr5:g.164864308G>A	*LOC102546299-894319, LINC01947-1467919*		Intergenic	rs557750713	0.00020

chr5:g.164870878G>A	*LOC102546299-900889, LINC01947-1461349*		Intergenic	rs35229950	NA
chr5:g.164878011G>T	*LOC102546299-908022, LINC01947-1454216*		Intergenic	rs10045320	NA
chr5:g.164896945A>G	*LOC102546299-926956, LINC01947-1435282*		Intergenic	rs561215567	0.00040
chr5:g.164899684C>T	*LOC102546299-929695, LINC01947-1432543*		Intergenic	rs552373794	0.00020
chr5:g.164920933G>C	*LOC102546299-950944, LINC01947-1411294*		Intergenic	rs528643540	0.00040
chr5:g.165013593A>G	*LOC102546299-1043604, LINC01947-1318634*		Intergenic	rs570876945	0.00020
chr5:g.165015166C>T	*LOC102546299-1045177, LINC01947-1317061*		Intergenic	rs938214647	NA
chr5:g.165051061T>C	*LOC102546299-1081072, LINC01947-1281166*		Intergenic	rs185346773	0.00200
chr5:g.165070641C>T	*LOC102546299-1100652, LINC01947-1261586*		Intergenic	rs192259849	0.00260
chr5:g.165139225G>A	*LOC102546299-1169236, LINC01947-1193002*		Intergenic	rs146162975	0.00260
chr5:g.165168773A>T	*LOC102546299-1198784, LINC01947-1163454*		Intergenic	rs566767411	0.00180
chr5:g.165213859A>G	*LOC102546299-1243870, LINC01947-1118368*		Intergenic	rs2161187	NA
chr5:g.165224633T>G	*LOC102546299-1254644, LINC01947-1107594*		Intergenic	rs895081805	NA
chr5:g.165251769A>G	*LOC102546299-1281780, LINC01947-1080458*		Intergenic	rs574483878	0.00240
chr5:g.165265038G>A	*LOC102546299-1295049, LINC01947-1067189*		Intergenic	rs192908274	0.00319
chr5:g.165276037G>A	*LOC102546299-1306048, LINC01947-1056190*		Intergenic	rs187319374	0.00339
chr5:g.165344913C>G	*LOC102546299-1374924, LINC01947-987314*		Intergenic	rs68055880	NA
chr5:g.165540228A>C	*LOC102546299-1570239, LINC01947-791999*		Intergenic	rs545909997	0.00080
chr5:g.165598858G>C	*LOC102546299-1628869, LINC01947-733369*		Intergenic	rs371223579	0.00240
chr5:g.165769836A>T	*LOC102546299-1799847, LINC01947-562391*		Intergenic	rs369930688	0.00419
chr5:g.165797402G>A	*LOC102546299-1827413, LINC01947-534825*		Intergenic	rs759456284	NA
chr5:g.165812969A>G	*LOC102546299-1842980, LINC01947-519258*		Intergenic		
chr5:g.165875572T>A	*LOC102546299-1905583, LINC01947-456655*		Intergenic	rs540707217	0.00040
chr5:g.165881734G>T	*LOC102546299-1911745, LINC01947-450493*		Intergenic	rs1028446729	NA
chr5:g.165998881A>G	*LOC102546299-2028892, LINC01947-333346*		Intergenic	rs7731799	NA
chr5:g.166004638A>G	*LINC01947-327589*		Intergenic	rs535844027	0.00319
chr5:g.166041313A>G	*LINC01947-290914*		Intergenic	rs138001378	0.00399
chr5:g.166067492C>T	*LINC01947-264735*		Intergenic	rs112127009	NA
chr5:g.166080467A>G	*LINC01947-251760*		Intergenic		
chr5:g.166161360A>G	*LINC01947-170867*		Intergenic		
chr5:g.166236514C>T	*LINC01947-95713*		Intergenic	rs547138388	0.00240
chr5:g.166256654G>A	*LINC01947-75573*		Intergenic	rs539434625	0.00260
chr5:g.166298349C>T	*LINC01947-33878*		Intergenic	rs115785121	0.00180
chr5:g.166339579C>T	*LINC01947*	NR_108020.1:n.409-4563G>A	Intron	rs902921285	NA
chr5:g.166339582C>T	*LINC01947*	NR_108020.1:n.409-4566G>A	Intron	rs1331397321	NA
chr5:g.166341777C>T	*LINC01947*	NR_108020.1:n.409-6761G>A	Intron	rs551414577	0.00419
chr5:g.166368051C>A	*LINC01947-14676, LOC101927908-223888*		Intergenic		
chr5:g.166391217T>A	*LINC01947-37842, LOC101927908-200722*		Intergenic	rs2861939	NA
chr5:g.166523822T>G	*TENM2*	NM_001122679.2:c.-75+23461T>G	Intron	rs374523012	NA
chr5:g.166583268T>C	*TENM2*	NM_001122679.2:c.-75+82907T>C	Intron	rs564701826	0.00160
chr5:g.166629286G>A	*TENM2*	NM_001122679.2:c.-74-82483G>A	Intron	rs573061885	0.00120

chr5:g.166639951T>G	*TENM2*	NM_001122679.2:c.-74-71818T>G	Intron		
chr5:g.166651391G>A	*TENM2*	NM_001122679.2:c.-74-60378G>A	Intron	rs544468296	0.00120
chr5:g.166667906G>A	*TENM2*	NM_001122679.2:c.-74-43863G>A	Intron	rs570866413	0.00140
chr5:g.166704145C>G	*TENM2*	NM_001122679.2:c.-74-7624C>G	Intron		
chr5:g.166707765A>G	*TENM2*	NM_001122679.2:c.-74-4004A>G	Intron	rs10050406	NA
chr5:g.166707769A>G	*TENM2*	NM_001122679.2:c.-74-4000A>G	Intron	rs113716059	NA
chr5:g.166755209G>T	*TENM2*	NM_001122679.1:c.226+43141G>T	Intron	rs371035702	NA
chr5:g.166771194C>T	*TENM2*	NM_001122679.1:c.227-31009C>T	Intron	rs375625896	0.00439
chr5:g.166786920A>G	*TENM2*	NM_001122679.1:c.227-15283A>G	Intron	rs376276543	0.00399
chr5:g.166858967T>C	*TENM2*	NM_001122679.1:c.502+56489T>C	Intron	rs201346516	NA
chr5:g.166872339T>G	*TENM2*	NM_001122679.1:c.502+69861T>G	Intron	rs61285311	NA
chr5:g.166872341T>G	*TENM2*	NM_001122679.1:c.502+69863T>G	Intron	rs7712523	NA
chr5:g.166872344A>G	*TENM2*	NM_001122679.1:c.502+69866A>G	Intron	rs796752439	NA
chr5:g.166872368A>G	*TENM2*	NM_001122679.1:c.502+69890A>G	Intron	rs796464549	NA
chr5:g.166894540T>A	*TENM2*	NM_001122679.1:c.502+92062T>A	Intron	rs1253280737	NA
chr5:g.166913127C>G	*TENM2*	NM_001122679.1:c.502+110649C>G	Intron	rs78380084	NA
chr5:g.166929549T>C	*TENM2*	NM_001122679.1:c.502+127071T>C	Intron	rs528647855	0.00100
chr5:g.167025527T>C	*TENM2*	NM_001122679.1:c.502+223049T>C	Intron	rs576861897	0.00140
chr5:g.167051984T>G	*TENM2*	NM_001122679.1:c.502+249506T>G	Intron	rs1273475085	NA
chr5:g.167067305C>T	*TENM2*	NM_001122679.1:c.503-235686C>T	Intron	rs538786664	0.00459
chr5:g.167122731T>A	*TENM2*	NM_001122679.1:c.503-180260T>A	Intron	rs279411	NA
chr5:g.167171538A>G	*TENM2*	NM_001122679.1:c.503-131453A>G	Intron	rs370407114	0.00140
chr5:g.167190343T>A	*TENM2*	NM_001122679.1:c.503-112648T>A	Intron	rs567947309	0.00439
chr5:g.167212760G>A	*TENM2*	NM_001122679.1:c.503-90231G>A	Intron		
chr5:g.167262318G>T	*TENM2*	NM_001122679.1:c.503-40673G>T	Intron		
chr5:g.167444759C>A	*TENM2*	NM_001122679.1:c.1186+24572C>A	Intron	rs561713205	0.00100
chr5:g.167450967A>C	*TENM2*	NM_001122679.1:c.1187-23465A>C	Intron	rs1862419	NA
chr5:g.167465214A>G	*TENM2*	NM_001122679.1:c.1187-9218A>G	Intron	rs247992	NA
chr5:g.167470906G>C	*TENM2*	NM_001122679.1:c.1187-3526G>C	Intron	rs17069703	NA
chr5:g.167493292T>C	*TENM2*	NM_001122679.1:c.1515+4022T>C	Intron		
chr5:g.167516942A>G	*TENM2*	NM_001122679.1:c.1516-637A>G	Intron	rs550672043	0.00100
chr5:g.167531400C>T	*TENM2*	NM_001122679.1:c.1813+6268C>T	Intron	rs545231293	0.00399
chr5:g.167531551G>A	*TENM2*	NM_001122679.1:c.1813+6419G>A	Intron	rs2617969	NA
chr5:g.167531557G>A	*TENM2*	NM_001122679.1:c.1813+6425G>A	Intron	rs2617970	NA
chr5:g.167550140G>T	*TENM2*	NM_001122679.1:c.2009-1715G>T	Intron	rs2617975	NA
chr5:g.167563228G>T	*TENM2*	NM_001122679.1:c.2395+9284G>T	Intron	rs540583017	0.00060
chr5:g.167612429C>G	*TENM2*	NM_001122679.1:c.2543-4913C>G	Intron	rs544018656	0.00339
chr5:g.167651992G>A	*TENM2*	NM_001122679.1:c.5082-1101G>A	Intron	rs562485313	0.00080
chr5:g.167669633T>C	*TENM2*	NM_001122679.1:c.5494-1792T>C	Intron	rs533760753	0.00240
chr5:g.167699878G>A	*TENM2-8716, WWC1-19187*		Intergenic	rs569538307	0.00080
chr5:g.167746150C>T	*WWC1*	NM_015238.2:c.119+26874C>T	Intron	rs369096913	0.00319

chr5:g.167754793A>G	*WWC1*	NM_015238.2:c.119+35517A>G	Intron	rs557449943	0.00300
chr5:g.167803696C>A	*WWC1*	NM_015238.2:c.229+5158C>A	Intron	rs943874668	NA
chr5:g.167839707C>T	*WWC1*	NM_015238.2:c.942-1646C>T	Intron	rs10044771	NA
chr5:g.167872941G>T	*WWC1*	NM_015238.2:c.2525+1351G>T	Intron	rs6876732	NA
chr5:g.168149800G>A	*SLIT3*	NM_003062.3:c.2411+138C>T	Intron	rs574868805	0.00020
chr5:g.168169533C>G	*SLIT3*	NM_003062.3:c.2270+5774G>C	Intron	rs2915812	NA
chr5:g.168199223C>T	*SLIT3*	NM_003062.3:c.1459+563G>A	Intron	rs531227420	0.00180
chr5:g.168227231C>A	*SLIT3*	NM_003062.3:c.936-4648G>T	Intron	rs551429702	0.00140
chr5:g.168286202C>T	*SLIT3*	NM_003062.3:c.486-14542G>A	Intron		
chr5:g.168331500T>C	*SLIT3*	NM_003062.3:c.414-21159A>G	Intron	rs200487769	NA
chr5:g.168536519T>C	*SLIT3*	NM_003062.3:c.413+83964A>G	Intron	rs190916	NA
chr5:g.168559290C>G	*SLIT3*	NM_003062.3:c.413+61193G>C	Intron	rs1808647	NA

Table A2.6. Heterozygous novel or rare InDels (MAF<0.005) at 5q34 (C1 minus B1)

Genomic Position	Ref	Alt	Gene/ Closest gene and distance	Trancript ID	Consequence	rsID	MAF (1000g)
153546582	AGG	A	*MFAP3-109568, GALNT10-23712*		Intergenic	rs764345928	NA
153546585	AAGGAAGGAAGGAAGGAAG	A	*MFAP3-109571, GALNT10-23693*		Intergenic	rs757672708	NA
153562754	C	CA	*MFAP3-125739, GALNT10-7542*		Intergenic	rs551897904	NA
153568180	A	ATTT	*MFAP3-131165, GALNT10-2116*	NM_198321	Upstream: 2115 bases	rs372408504	NA
153578080	AG	A	*GALNT10*	NM_198321	Intron	rs1209559892	NA
153596479	G	GA	*GALNT10*	NM_198321	Intron	rs35510232	NA
153605355	TC	T	*GALNT10*	NM_198321	Intron	rs143281141	NA
153623661	TG	T	*GALNT10*	NM_198321	Intron	rs557130492	0.00260
153635942	TC	T	*GALNT10*	NM_198321	Intron	rs572710498	0.00220
153658869	T	TG	*GALNT10*	NM_198321	Intron		
153665946	GTT	G	*GALNT10*	NM_198321	Intron	rs755457967	NA
153667969	T	TTGTG	*GALNT10*	NM_198321	Intron	rs1417464839	NA
153669652	AT	A	*GALNT10*	NM_198321	Intron	rs1296337436	NA
153673155	G	GAT	*GALNT10*	NM_198321	Intron	rs1330124824	NA
153674537	TTG	T	*GALNT10*	NM_198321	Intron	rs202182228	NA
153684177	CA	C	*GALNT10*	NM_198321	Intron	rs57928112	NA
153684872	AG	A	*GALNT10*	NM_198321	Intron	rs1404650341	NA
153687533	C	CA	*GALNT10*	NM_198321	Intron	rs563510814	NA
153694072	AT	A	*GALNT10*	NM_198321	Intron	rs369968934	NA
153751124	TAC	T	*GALNT10*	NM_198321	Intron	rs796318005	NA
153772618	CAAA	C	*GALNT10*	NM_198321	Intron	rs139303055	NA
153776687	CAAA	C	*GALNT10*	NM_198321	Intron	rs58303911	NA

153796041	TACACAC	T	*SAP30L-AS1*	NR_037897	Intron	rs141022851	NA
153810354	CACA	C	*SAP30L-AS1*	NR_037897	Intron	rs71946604	NA
153863059	AG	A	*HAND1-5235, MIR3141-112513*		Intergenic	rs36011638	NA
153887524	C	CATAT	*HAND1-29699, MIR3141-88049*		Intergenic	rs70978547	NA
153925980	CTG	C	*HAND1-68156, MIR3141-49591*		Intergenic	rs1417745137	NA
153944620	C	CA	*HAND1-86795, MIR3141-30953*		Intergenic	rs375965753	NA
153963004	G	GAAAGA	*HAND1-105179, MIR3141-12569*		Intergenic	rs1312646589	NA
153965824	C	CAA	*HAND1-107999, MIR3141-9749*		Intergenic	rs771554743	NA
153977655	C	CATGG	*MIR3141-2022, MIR1303-87682*	NR_036094	Upstream: 2023 bases	rs35471198	NA
153977887	CGGAT	C	*MIR3141-2255, MIR1303-87446*	NR_036094	Upstream: 2255 bases	rs148855025	NA
153987440	C	CA	*MIR3141-11807, MIR1303-77897*		Intergenic	rs527402243	NA
154005997	CA	C	*MIR3141-30365, MIR1303-59339*		Intergenic	rs1233577755	NA
154121357	GTTGT	G	*LARP1*	NM_015315	Intron	rs1226338269	NA
154142615	C	CT	*LARP1*	NM_015315	Intron	rs147728579	NA
154206866	CA	C	*FAXDC2*	NM_032385	Intron	rs1324808409	NA
154270028	A	AT	*GEMIN5*	NM_015465	Intron	rs748573084	NA
154314386	ATTCTC	A	*GEMIN5*	NM_015465	Intron		
154369920	T	TCACACA	*MRPL22-20948, KIF4B-23396*		Intergenic	rs148982356	NA
154382390	A	AT	*MRPL22-33418, KIF4B-10926*		Intergenic	rs59397384	NA
154430797	CT	C	*KIF4B-33105, SGCD-1322970*		Intergenic	rs1406183790	NA
154449340	ATGTG	A	*KIF4B-51648, SGCD-1304424*		Intergenic	rs56187278	NA
154495924	C	CAGTTT	*KIF4B-98231, SGCD-1257844*		Intergenic	rs10654977	NA
154552977	A	AGGCATG	*KIF4B-155284, SGCD-1200791*		Intergenic	rs779427621	NA
154645605	A	AT	*KIF4B-247912, SGCD-1108163*		Intergenic		
154772598	C	CAT	*KIF4B-374905, SGCD-981170*		Intergenic		
154794659	A	AG	*KIF4B-396966, SGCD-959109*		Intergenic	rs11410539	NA
154823024	C	CTGTGTGTGTGTGTGTGTGTG	*KIF4B-425331, SGCD-930744*		Intergenic	rs142027746	NA
154831769	ATGTGTGTGTG	A	*KIF4B-434077, SGCD-921989*		Intergenic	rs1307574829	NA
154835976	GTT	G	*KIF4B-438284, SGCD-917790*		Intergenic	rs35235829	NA
154838161	CGTGTGTGTGTGTGTGTGTGT	C	*KIF4B-440469, SGCD-915587*		Intergenic	rs34260514	NA
154916760	TGAGAGAGAGAGAGA	T	*KIF4B-519068, SGCD-836994*		Intergenic	rs139084522	NA
154981372	CTTTCTTTA	C	*KIF4B-583680, SGCD-772388*		Intergenic	rs772066571	NA
155001173	GTT	G	*KIF4B-603481, SGCD-752593*		Intergenic	rs3058427	NA
155110036	G	GTGTGTGTGTGTGTGT	*KIF4B-712343, SGCD-643732*		Intergenic		
155190873	C	CCT	*KIF4B-793180, SGCD-562895*		Intergenic	rs61013514	NA
155195264	GTTTTGTTTTGT	G	*KIF4B-797572, SGCD-558493*		Intergenic	rs372428824	NA
155252752	TG	T	*KIF4B-855060, SGCD-501015*		Intergenic	rs201199660	NA
155265786	CAA	C	*KIF4B-868094, SGCD-487980*		Intergenic	rs869078751	NA
155267599	TTG	T	*KIF4B-869907, SGCD-486167*		Intergenic	rs777137883	NA

155269257	G	GT	*KIF4B-871564, SGCD-484511*		Intergenic	rs112761473	NA
155276457	G	GCGCACACACA	*KIF4B-878764, SGCD-477311*		Intergenic	rs201269068	NA
155369329	G	GTATC	*KIF4B-971636, SGCD-384439*		Intergenic	rs1215503536	NA
155374304	TTGTG	T	*KIF4B-976612, SGCD-379460*		Intergenic	rs768759350	NA
155420348	CT	C	*KIF4B-1022656, SGCD-333419*		Intergenic	rs34221751	NA
155525650	TAC	T	*KIF4B-1127958, SGCD-228116*		Intergenic		
155624483	C	CT	*KIF4B-1226790, SGCD-129285*		Intergenic	rs368116194	NA
155668262	C	CGTGTGTGT	*KIF4B-1270569, SGCD-85506*		Intergenic	rs146077140	NA
155678533	C	CT	*SGCD*	NM_000337	Intron	rs34601103	NA
155714638	T	TTG	*KIF4B-1316945, SGCD-39130*		Intergenic	rs70982002	NA
155747108	ATG	A	*KIF4B-1349416, SGCD-6658*		Intergenic	rs66480953	NA
155761620	C	CTTT	*SGCD*	NM_000337	Intron	rs1461532011	NA
155773789	C	CT	*SGCD*	NM_000337	Intron	rs777104014	NA
155801160	CA	C	*SGCD*	NM_000337	Intron	rs72228538	NA
155802428	T	TA	*SGCD*	NM_000337	Intron	rs56706247	NA
155833801	TTATATATATATATA	T	*SGCD*	NM_000337	Intron	rs200600544	NA
155891227	A	AT	*SGCD*	NM_000337	Intron	rs773294916	NA
155964470	CCA	C	*SGCD*	NM_000337	Intron	rs771089553	NA
155964855	GA	G	*SGCD*	NM_000337	Intron	rs369147661	NA
156124762	C	CGGAT	*SGCD*	NM_000337	Intron	rs773596086	NA
156204745	TA	T	*SGCD-9947, PPP1R2P3-72752*		Intergenic	rs1173394592	NA
156224821	CT	T	*SGCD-30023, PPP1R2P3-52666*		Intergenic		
156284328	C	CT	*PPP1R2P3-4788, TIMD4-61966*	NR_038443	Downstream: 4789 bases	rs537390765	NA
156284674	A	AT	*PPP1R2P3-5134, TIMD4-61620*		Intergenic	rs562804078	NA
156287344	T	TTGTG	*PPP1R2P3-7804, TIMD4-58950*		Intergenic	rs113351181	NA
156292502	GGTTGTTGTT	G	*PPP1R2P3-12963, TIMD4-53783*		Intergenic	rs113904862	NA
156309478	G	GA	*PPP1R2P3-29938, TIMD4-36816*		Intergenic	rs553125286	NA
156314496	G	GGT	*PPP1R2P3-34956, TIMD4-31798*		Intergenic	rs143888570	NA
156324170	GA	G	*PPP1R2P3-44631, TIMD4-22123*		Intergenic	rs11321780	NA
156331170	TCACACACA	T	*PPP1R2P3-51631, TIMD4-15116*		Intergenic	rs57089071	NA
156333850	C	CA	*PPP1R2P3-54310, TIMD4-12444*		Intergenic	rs11381835	NA
156384045	C	CT	*TIMD4*	NM_138379	Intron		
156402621	AT	A	*TIMD4-12355, HAVCR1-53803*		Intergenic	rs376467277	NA
156471380	C	CGAAG	*HAVCR1*	NM_012206	Intron	rs1419578548	NA
156471480	AAAGGAAGGAAGGAAGGAAGGAAGG	A	*HAVCR1*	NM_012206	Intron	rs201451883	NA
156471611	GAAAGAAAGAAAGAAAGAGAAAGAAAGAA	G	*HAVCR1*	NM_012206	Intron		
156558606	ATT	A	*HAVCR2-22358, MED7-6844*		Intergenic	rs148068703	NA
156566741	C	CGGCCGGGCGCGG	*MED7*	NM_004270	Intron		

156566744	G	GCTC	*MED7*	NM_004270	Intron	rs1180168403	NA
156566746	C	CGCCTGTAA	*MED7*	NM_004270	Intron	rs1242633797	NA
156612694	C	CA	*ITK*	NM_005546	Intron	rs59585425	NA
156627116	G	GAA	*ITK*	NM_005546	Intron	rs68138116	NA
156636340	C	CA	*ITK*	NM_005546	Intron	rs34768868	NA
156636939	AT	A	*ITK*	NM_005546	Intron	rs34236243	NA
156682907	CTGTG	C	*ITK-798*	NM_005546	Downstream: 798 bases	rs765600144	
156734660	G	GTGTGTGTGTA	*CYFIP2*	NM_014376	Intron		
156814683	A	AT	*LOC102724404*	NM_014376	Intron		
156818841	A	AT	*LOC102724404*	NM_014376	Intron	rs139659989	NA
156820216	GA	G	*CYFIP2*	NM_014376	3'UTR	rs3052310	NA
156842946	ATT	A	*LOC102724404*		Intergenic	rs1293887939	NA
156849923	GA	G	*LOC102724404*		Intergenic	rs33929023	NA
156850243	CT	C	*LOC102724404*		Intergenic	rs34877929	NA
156873690	C	CAAAAA	*LOC102724404*		Intron	rs375790935	NA
156978417	A	AC	*ADAM19*	NM_033274	Intron		
157027973	A	AT	*ADAM19-25141, SOX30-24715*		Intergenic	rs367759800	NA
157045099	A	AAT	*ADAM19-42267, SOX30-7589*		Intergenic	rs34673325	NA
157055374	TCACACACA	T	*SOX30*	NM_178424	Intron	rs1465326943	NA
157059056	TAAAAAAAAAAAA	T	*SOX30*	NM_178424	Intron	rs1221639219	NA
157069324	CA	C	*SOX30*	NM_178424	Intron	rs11317659	NA
157087436	T	TATAAA	*SOX30*		Intron	rs537098756	NA
157092910	CT	C	*SOX30*		Intergenic	rs67520310	NA
157103870	C	CT	*C5orf52*	NM_001145132	Intron	rs61514085	NA
157104677	AAG	A	*C5orf52*	NM_001145132	Intron	rs546081162	NA
157114849	C	CA	*C5orf52-7686, THG1L-43549*		Intergenic	rs34727617	NA
157140584	CT	C	*C5orf52-33422, THG1L-17813*		Intergenic	rs1309606537	NA
157146792	CAAA	C	*C5orf52-39630, THG1L-11603*		Intergenic	rs1328250285	NA
157158119	AT	A	*THG1L-278*	NM_017872	Upstream: 204 bases	rs1242657763	NA
157179171	CT	C	*LSM11*	NM_173491	Intron	rs1304026589	NA
157257766	C	CCT	*CLINT1*	NM_014666	Intron	rs1236018682	NA
157265131	G	GT	*CLINT1*	NM_014666	Intron	rs199515235	NA
157272562	A	AT	*CLINT1*	NM_014666	Intron		
157282171	G	GC	*CLINT1*	NM_014666	Intron	rs144882378	NA
157282175	G	GT	*CLINT1*	NM_014666	Intron	rs372664515	NA
157290190	G	GT	*CLINT1-4006, LINC02227- 457523*	NM_014666	Upstream: 4007 bases	rs546656582	NA
157324683	CATATATATATATATATATAT	C	*CLINT1-38500, LINC02227-423010*		Intergenic	rs368237550	NA
157324703	T	TCTCTCTCTC	*CLINT1-38519, LINC02227-423010*		Intergenic		
157350426	T	TAC	*CLINT1-64242, LINC02227-397287*		Intergenic	rs774183492	NA
157355259	A	ATT	*CLINT1-69075, LINC02227-392454*		Intergenic	rs563619291	NA

157366168	CTT	C	*CLINT1-79985, LINC02227-381543*		Intergenic	rs771248122	NA
157372106	C	CTTTCT	*CLINT1-85922, LINC02227-375607*		Intergenic	rs143454506	NA
157372715	A	ACCTTCCTT	*CLINT1-86531, LINC02227-374998*		Intergenic	rs1244491207	NA
157382991	C	CT	*CLINT1-96807, LINC02227-364722*		Intergenic		
157382994	C	CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT	*CLINT1-96810, LINC02227-364719*		Intergenic		
157385156	C	CTTTATTTATTTATTTATTTA	*CLINT1-98972, LINC02227-362557*		Intergenic	rs140576849	NA
157390099	CTTT	C	*CLINT1-103916, LINC02227-357611*		Intergenic	rs1330761546	NA
157405599	A	AT	*CLINT1-119415, LINC02227-342114*		Intergenic		
157431624	T	TA	*CLINT1-145440, LINC02227-316089*		Intergenic	rs542658129	NA
157432083	A	AT	*CLINT1-145899, LINC02227-315630*		Intergenic	rs577288754	0.00200
157446958	C	CTTTTATTTTTTTT	*CLINT1-160774, LINC02227-300755*		Intergenic		
157458624	A	ATCTATATC	*CLINT1-172440, LINC02227-289089*		Intergenic	rs11280701	NA
157472364	GA	G	*CLINT1-186181, LINC02227-275348*		Intergenic	rs944950422	NA
157474186	AT	A	*CLINT1-188003, LINC02227-273526*		Intergenic	rs796459408	NA
157575688	CCTTCCCTTCTTCCCTCCTTCCCTT	C	*CLINT1-289505, LINC02227-172001*		Intergenic	rs200440722	NA
157581188	A	AAC	*CLINT1-295004, LINC02227-166525*		Intergenic	rs57378473	NA
157585865	C	CATATATATATATATATATGTGCATATATATATATATATATATATGCACATATAT	*CLINT1-299681, LINC02227-161848*		Intergenic		
157660282	T	TAA	*CLINT1-374098, LINC02227-87431*		Intergenic	rs1414654567	NA
157929037	GTA	G	*LINC02227-92256, EBF1-193885*		Intergenic	rs137912110	NA
157940144	G	GGAGAGAGA	*LINC02227-103362, EBF1-182780*		Intergenic	rs34258670	NA
158041184	A	AAAAC	*LINC02227-204402, EBF1-81740*		Intergenic		
158087718	TTTTTC	T	*LINC02227-250937, EBF1-35201*		Intergenic	rs962080657	NA
158249268	C	CGGAT	*EBF1*	NM_024007	Intron	rs1197459241	NA
158252565	CT	C	*EBF1*	NM_024007	Intron	rs201596659	NA
158252569	AAAT	A	*EBF1*	NM_024007	Intron	rs200208298	NA
158278564	GAGGGGAGGGAAAAAGGGA	G	*EBF1*	NM_024007	Intron	rs200154704	NA

158291232	CAAAAAAAAAAA	C	*EBF1*	NM_024007	Intron	rs564004570	NA
158396609	CA	C	*EBF1*	NM_024007	Intron	rs763360629	NA
158419233	GA	G	*EBF1*	NM_024007	Intron	rs35596220	NA
158433656	TAAAAAA	T	*EBF1*	NM_024007	Intron	rs1177310324	NA
158477161	T	TTCTCTCTC	*EBF1*	NM_024007	Intron		
158544834	T	TAC	*LINC02202-347*		Intron	rs9313800	NA
158553622	GT	G	*LINC02202-9136, RNF145-30795*		Intergenic	rs566776929	0.00060
158656955	C	CA	*LINC01932*		Intron	rs71577369	NA
158840814	C	CAAAAA	*LOC285626-50971, LINC01845-34751*		Intergenic	rs55761945	NA
158953856	ACACACACACACACG	A	*LINC01845-60572, LINC01847-249913*		Intergenic	rs761293082	NA
159025196	GCT	G	*LINC01845-131912, LINC01847-178585*		Intergenic	rs35079774	NA
159031749	GTATATATATATATATATATATATATATATA	G	*LINC01845-138465, LINC01847-172002*		Intergenic	rs10528213	NA
159036018	T	TTGTG	*LINC01845-142733, LINC01847-167765*		Intergenic	rs145558492	NA
159054204	C	CAAAAAAAAAA	*LINC01845-160919, LINC01847-149579*		Intergenic		
159101740	CTCTCTCTGTGTGTGTG	C	*LINC01845-208456, LINC01847-102027*		Intergenic	rs199568972	NA
159118343	CTTTG	C	*LINC01845-225059, LINC01847-85436*		Intergenic	rs149744674	NA
159251356	G	GGAAGGAAGGAAA	*LINC01847*		Intron	rs746597481	NA
159264387	C	CTT	*LINC01847*		Intron	rs57599953	NA
159280990	T	TA	*LINC01847*		Intron	rs776984270	NA
159293303	G	GA	*LINC01847*		Intron	rs34220518	NA
159361825	CA	C	*ADRA1B*	NM_000679	Intron	rs111598190	NA
159366260	G	GGAGAGA	*ADRA1B*	NM_000679	Intron	rs140658519	NA
159370221	GA	G	*ADRA1B*	NM_000679	Intron	rs11304091	NA
159387716	TA	T	*ADRA1B*	NM_000679	Intron	rs1325582274	NA
159435859	AAAAT	A	*TTC1-245*	NM_003314	Upstream: 321 bases	rs202207510	NA
159478090			*TTC1*	NM_003314	Intron	rs5872626	NA
159485417	CT	C	*TTC1*	NM_003314	Intron	rs1474326202	NA
159508247	C	CT	*PWWP2A*	NM_052927	Intron	rs762798184	NA
159512513			*PWWP2A*	NM_052927	Intron	rs5872628	NA
159532782	CA	C	*PWWP2A*	NM_052927	Intron	rs549595565	NA
159542145	CAAAAAAAAAAAAAAA	C	*PWWP2A*	NM_052927	Intron	rs760105134	NA
159551031	C	CA	*PWWP2A*	NM_052927	Upstream: 4579 bases	rs34976521	NA
159573622	C	CAA	*PWWP2A-27169, FABP6-40753*		Intergenic	rs35723602	NA

159583880	TG	T	*PWWP2A-37428, FABP6-30494*		Intergenic	rs540392804	0.00499
159654533	C	CAAA	*FABP6*	NM_001040442	Intron	rs538577462	NA
159663825	CA	C	*FABP6*	NM_001445	Intron	rs375689902	NA
159885705	A	ATGTGTG	*PTTG1-29953, MIR3142HG-9554*		Intergenic	rs113888268	NA
159953880	T	TGA	*MIR3142HG-39446, ATP10B-36248*		Intergenic	rs34104091	NA
159987297	GGT	G	*MIR3142HG-72864, ATP10B-2829*	NM_025153	Downstream: 2830 bases	rs368068125	NA
160035480	ATTTCTAGT	A	*ATP10B*	NM_025153	Intron	rs951323812	NA
160098977	C	CTTT	*ATP10B*	NM_025153	Intron	rs754150310	NA
160193345	AACACACACACACACACACACACACACACA C	A	*ATP10B*	NM_025153	Intron	rs149750950	NA
160209446	TACACAC	T	*ATP10B*	NM_025153	Intron	rs70990746	NA
160248204	C	CA	*ATP10B*	NM_025153	Intron	rs549603478	0.00080
160306545	T	TA	*ATP10B-27325, LINC02159-52242*		Intergenic	rs757034689	NA
160426997	TTATATATATATATATA TATATATATATATATA	T	*LINC02159-61364, GABRB2-288408*		Intergenic	rs149676754	NA
160439829	AT	A	*LINC02159-74196, GABRB2-275607*		Intergenic	rs5872674	NA
160440151	GT	G	*LINC02159-74518, GABRB2-275285*		Intergenic	rs34688231	NA
160445538	TTA	T	*LINC02159-79905, GABRB2-269897*		Intergenic	rs36099890	NA
160507331	GTTTTTTTTTTTTTTTTT TTT	G	*LINC02159-141698, GABRB2-208086*		Intergenic	rs1435243354	NA
160562354	TACACACACAC	T	*LINC02159-196721, GABRB2-153073*		Intergenic	rs1191398739	NA
160569474	T	TA	*LINC02159-203840, GABRB2-145963*		Intergenic	rs372283034	NA
160618454	T	TAA	*LINC02159-252820, GABRB2-96983*		Intergenic		
160671108	ACT	A	*LINC02159-305475, GABRB2-44327*		Intergenic	rs879936615	NA
160671113	T	TAACAA	*LINC02159-305479, GABRB2-44324*		Intergenic	rs372419784	NA
160754049	G	GACACACAC	*GABRB2*	NM_000813	Intron	rs150178203	NA
160763822	C	CA	*GABRB2*	NM_000813	Intron	rs746697698	NA
160812952	CT	C	*GABRB2*	NM_000813	Intron	rs372100746	NA
160812955	GTGTGTGTGTGT	G	*GABRB2*	NM_000813	Intron	rs757411747	NA
160816578	A	AGT	*GABRB2*	NM_000813	Intron	rs10666283	NA
160838821	TTA	T	*GABRB2*	NM_000813	Intron	rs148863496	NA
160867085	AAC	A	*GABRB2*	NM_000813	Intron	rs538131557	NA
160924537	A	ATG	*GABRB2*	NM_000813	Intron	rs71587162	NA
161004861	GA	G	*GABRA6-124737, GABRA1-269330*		Intergenic	rs74732118	NA
161140512	AAAAAAAC	A	*GABRA6-10914, GABRA1-133679*		Intergenic	rs766185489	NA
161150381	C	CTGTGTGTG	*GABRA6-20782, GABRA1-123817*		Intergenic	rs112271024	NA
161168663	CTT	C	*GABRA6-39065, GABRA1-105533*		Intergenic	rs1491076151	NA

161188906	A	AAC	*GABRA6-59307, GABRA1-85292*		Intergenic	rs112277651	NA
161217261	G	GAT	*GABRA6-87662, GABRA1-56937*		Intergenic	rs34019506	NA
161229750	TGATA	T	*GABRA6-100152, GABRA1-44444*		Intergenic	rs750974435	NA
161229797	TAGATAG	T	*GABRA6-100199, GABRA1-44395*		Intergenic	rs779735717	NA
161235737	C	CAT	*GABRA6-106138, GABRA1-38461*		Intergenic	rs56043626	NA
161256609	AAAAAAAAACAAAAAAAAAAAAC	A	*GABRA6-127011, GABRA1-17566*		Intergenic	rs375684226	NA
161263175	AAGG	A	*GABRA6-133577, GABRA1-11020*		Intergenic	rs879601314	NA
161320018	ATAATAATAAT	A	*GABRA1*	NM_000806	Intron	rs755962484	NA
161338836	C	CAT	*LINC01202*		Intron	rs33952585	NA
161345371	TTGTGTGTG	T	*LINC01202*		Intron	rs56863544	NA
161356187	CAA	C	*LINC01202*		Intron	rs201273039	NA
161427343	TGTGTGA	T	*LINC01202*		Intron	rs1281142588	NA
161474740	C	CT	*LINC01202-46537, GABRG2-19909*		Intergenic	rs201377486	NA
161482128	CTT	C	*LINC01202-53926, GABRG2-12519*		Intergenic	rs386694323	NA
161482130	T	TCAC	*LINC01202-53927, GABRG2-12519*		Intergenic	rs778748737	NA
161488749	G	GA	*LINC01202-60546, GABRG2-5900*		Intergenic	rs368255626	NA
161599830	TA	T	*GABRG2-17298*		Intergenic	rs5872750	NA
161605429	CTT	C	*GABRG2-22884, CCNG1-1259147*		Intergenic	rs1491199469	NA
161627500	TAC	T	*GABRG2-44955, CCNG1-1237076*		Intergenic	rs796842323	NA
161658231	AT	A	*GABRG2-75686, CCNG1-1206346*		Intergenic	rs34058252	NA
161688155	T	TTG	*GABRG2-105609, CCNG1-1176423*		Intergenic	rs67229628	NA
161753625	C	CT	*GABRG2-171079, CCNG1-1110953*		Intergenic	rs758473581	NA
161759739	TAA	T	*GABRG2-177194, CCNG1-1104837*		Intergenic	rs766756861	NA
161808554	T	TAGATAGATAGATAGATAGATAGATAGATAGAG	*GABRG2-226008, CCNG1-1056024*		Intergenic	rs139434750	NA
161809017	CATATATAT	C	*GABRG2-226472, CCNG1-1055553*		Intergenic	rs5872765	NA
161819658	CTTTTT	C	*GABRG2-237113, CCNG1-1044915*		Intergenic	rs149172843	NA
161820358	C	CA	*GABRG2-237812, CCNG1-1044220*		Intergenic	rs70994068	NA
161825911	A	AAG	*GABRG2-243365, CCNG1-1038667*		Intergenic	rs71587190	NA
161836920	AATATATATATATAT	A	*GABRG2-254375, CCNG1-1027644*		Intergenic	rs57091620	NA
161850600	A	AAT	*GABRG2-268054, CCNG1-1013978*		Intergenic	rs5872770	NA
161853550	C	CTTTTTTTT	*GABRG2-271004, CCNG1-1011028*		Intergenic	rs754676958	NA
161957748	C	CGACTTTGAAAAGGTACATAATATAAACACGTTAAAAAAACTAAGGGAAATAATG	*GABRG2-375202, CCNG1-906830*		Intergenic		
161986670	G	GA	*GABRG2-404124, CCNG1-877908*		Intergenic		
162013969	C	CA	*GABRG2-431423, CCNG1-850609*		Intergenic	rs11444893	NA
162026196	CTTGAGTTATTTATTATTA	C	*GABRG2-443651, CCNG1-838364*		Intergenic	rs201757655	NA

162183208	C	CT	*GABRG2-600662, CCNG1-681370*		Intergenic	rs756738760	NA
162183546	C	CT	*GABRG2-601000, CCNG1-681032*		Intergenic	rs766963693	NA
162421794	CA	C	*GABRG2-839249, CCNG1-442783*		Intergenic	rs201252781	NA
162424446	G	GAA	*GABRG2-841900, CCNG1-440132*		Intergenic	rs3085099	NA
162429067	CACAG	C	*GABRG2-846522, CCNG1-435507*		Intergenic	rs61453480	NA
162439995	CA	C	*GABRG2-857450, CCNG1-424582*		Intergenic	rs113992109	NA
162497476	GT	T	*GABRG2-914931, CCNG1-367100*		Intergenic	rs35211872	NA
162507128	CAA	C	*GABRG2-924583, CCNG1-357448*		Intergenic	rs1491061605	NA
162568118	CA	C	*GABRG2-985573, CCNG1-296459*		Intergenic	rs113178869	NA
162580174	AATTATTATT	A	*GABRG2-997629, CCNG1-284395*		Intergenic	rs150275776	NA
162587949	G	GA	*GABRG2-1005403, CCNG1-19483*		Intergenic	rs11303435	NA
162607432	G	GA	*GABRG2-1024886, CCNG1-257146*		Intergenic	rs59224969	NA
162615146	C	CAT	*GABRG2-1032600, CCNG1-249432*		Intergenic	rs139860192	NA
162653956	C	CTGTG	*GABRG2-1071410, CCNG1-210622*		Intergenic	rs1362041645	NA
162714208	CAT	C	*GABRG2-1131663, CCNG1-150368*		Intergenic	rs1444862111	NA
162746604	GAAAA	G	*GABRG2-1164059, CCNG1-117970*		Intergenic	rs543314198	0.00280
162750427	C	CT	*GABRG2-1167881, CCNG1-114151*		Intergenic	rs753295109	NA
162757780	C	CA	*GABRG2-1175234, CCNG1-106798*		Intergenic	rs397959909	NA
162884146	G	GTT	*NUDCD2*	NM_145266	Intron	rs375953521	NA
162889834	CAA	C	*HMMR*	NM_012484	Intron	rs747644960	NA
162927317	GTTTC	G	*HMMR-AS1-6253, MAT2B-2750*	NM_182796	Upstream: 2914 bases	rs10544001	NA
162937401	C	CTTT	*MAT2B*	NM_013283	Intron	rs1270371144	NA
162938446	CTTT	C	*MAT2B*	NM_013283	Intron	rs60009583	NA
162978155	T	TCACA	*MAT2B-31795, LINC02143-897274*		Intergenic	rs61431868	NA
162990919	GA	G	*MAT2B-44560, LINC02143-884509*		Intergenic	rs1324388042	NA
163091362	CA	C	*MAT2B-145018, LINC02143-784050*		Intergenic	rs35347532	NA
163091531	C	CAA	*MAT2B-145171, LINC02143-783898*		Intergenic	rs752603877	NA
163126437	C	CTTTTTTTT	*MAT2B-180077, LINC02143-748992*		Intergenic		
163306647	C	CTTT	*MAT2B-360287, LINC02143-568782*		Intergenic	rs1405998935	NA
163349628	C	CAAAA	*MAT2B-403268, LINC02143-525801*		Intergenic	rs10675498	0.00160
163352854	GTACACACACACACAC A	G	*MAT2B-406495, LINC02143-522559*		Intergenic	rs879776232	NA
163393864	T	TTTC	*MAT2B-447504, LINC02143-481565*		Intergenic	rs201501652	NA
163394907	C	CTGTG	*MAT2B-448547, LINC02143-480522*		Intergenic	rs781085008	NA
163416190	A	ATGTG	*MAT2B-469830, LINC02143-459239*		Intergenic	rs755799344	NA
163494702	C	CT	*MAT2B-548342, LINC02143-380727*		Intergenic	rs557975592	NA
163500385	A	AC	*MAT2B-554025, LINC02143-375044*		Intergenic	rs543547863	0.00120
163534994	CA	C	*MAT2B-588635, LINC02143-340434*		Intergenic	rs527987578	0.00060
163545512	GCA	G	*MAT2B-599153, LINC02143-329915*		Intergenic	rs568182794	0.00080
163737426	G	GTGTA	*MAT2B-791066, LINC02143-138003*		Intergenic	rs796561247	NA

163810936	A	AT	*LOC105377703*		Intron	rs35051061	NA
163870243	C	CTGTT	*MAT2B-923883, LINC02143-5186*		Intergenic		
163871035	A	AGT	*MAT2B-924675, LINC02143-4394*		Intergenic	rs879935216	NA
163871036	A	AAG	*MAT2B-924676, LINC02143-4393*		Intergenic	rs879340540	NA
163878933	ATTTTTTTTTTAGTTGTTTTT	A	*LINC02143*		Intron	rs1317984586	NA
163879477	AG	A	*LINC02143*		Intron	rs767340239	NA
163879480	ATCGC	A	*LINC02143*		Intron	rs200244866	NA
163900311	C	CT	*LOC102546299*		Intron	rs201761975	NA
163909189	AT	A	*LOC102546299*		Intron	rs751790840	NA
163955900	CAAAAAAAAAAAAA	C	*LOC102546299*		Intron	rs771723677	NA
163969269	TCA	T	*LOC102546299*		Intron	rs1491350986	NA
163980313	C	CT	*LOC102546299-10323*		Intergenic	rs375116415	NA
164057791	G	GA	*LOC102546299-87801*		Intergenic	rs752898585	NA
164096111	C	CTT	*LOC102546299-126121*		Intergenic	rs1291553868	NA
164110207	C	CT	*LOC102546299-140217*		Intergenic		
164123622	A	AAAAAG	*LOC102546299-153632*		Intergenic	rs776833671	NA
164131972	TTGTGTG	T	*LOC102546299-161983*		Intergenic	rs56841318	NA
164167290	A	AAC	*LOC102546299-197300*		Intergenic	rs570650455	0.00220
164189732	AT	A	*LOC102546299-219743*		Intergenic	rs1272773429	NA
164199946	A	AAAAG	*LOC102546299-229956*		Intergenic	rs549945802	NA
164201713	G	GTTT	*LOC102546299-231723,*		Intergenic	rs35034367	NA
164373756	C	CA	*LOC102546299-403766, LINC01947-1958472*		Intergenic	rs538593501	0.00100
164422429	CACACATACATACATATATATAT	C	*LOC102546299-452440, LINC01947-1909777*		Intergenic	rs142937010	NA
164425032	G	GAAATAAAT	*KIF4B-910613, SGCD-445462*		Intergenic	rs149760125	NA
164428136	TAAA	T	*LOC102546299-458147, LINC01947-1904089*		Intergenic	rs34140648	NA
164440271	T	TACACACAC	*LOC102546299-470281, LINC01947-1891957*		Intergenic	rs373230453	NA
164573454	T	TTGTGTGTG	*LOC102546299-603464, LINC01947-1758774*		Intergenic	rs372065371	NA
164725555	GT	G	*LOC102546299-755566, LINC01947-1606672*		Intergenic	rs537221735	0.00120
164786307	C	CT	*LOC102546299-816317, LINC01947-1545921*		Intergenic		
164789914	CT	C	*LOC102546299-819925, LINC01947-1542313*		Intergenic	rs61699662	NA
164798808	C	CA	*LOC102546299-828818, LINC01947-1533420*		Intergenic	rs769484339	NA
164822439	CTCTCTCTATATA	C	*LOC102546299-852450, LINC01947-1509777*		Intergenic	rs373505886	NA
164836366	G	GA	*LOC102546299-866376, LINC01947-1495862*		Intergenic	rs71580948	NA
164838697	AAAATAAAT	A	*LOC102546299-868708, LINC01947-1493523*		Intergenic	rs60778494	NA
164853712	T	TA	*LOC102546299-883722, LINC01947-1478516*		Intergenic	rs200114247	NA
164853713	T	TA	*LOC102546299-883723, LINC01947-1478515*		Intergenic		
164960612	AG	A	*LOC102546299-990623, LINC01947-1371615*		Intergenic	rs5872940	NA
164964680	CAG	C	*LOC102546299-994691, LINC01947-1367546*		Intergenic	rs777255440	NA
164979741	GTTT	G	*LOC102546299-1009752, LINC01947-1352484*		Intergenic	rs200559996	NA

165237695	C	CAAAACAAA	*LOC102546299-1267705*		Intergenic	rs368928263	NA
165267367	G	GT	*LOC102546299-1297377, LINC01947-1064861*		Intergenic	rs200278261	NA
165268706	A	ATT	*LOC102546299-1298716, LINC01947-1063522*		Intergenic	rs555997131	NA
165300704	TA	T	*LOC102546299-1330715, LINC01947-1031521*		Intergenic	rs10710577	NA
165307705	C	CA	*LOC102546299-1337715, LINC01947-1024523*		Intergenic	rs35160116	NA
165311267	CAA	C	*LOC102546299-1341278, LINC01947-1020959*		Intergenic	rs1284865899	NA
165434156	A	ATT	*LOC102546299-1464166, LINC01947-898072*		Intergenic	rs138638411	NA
165503322	CAAA	C	*LOC102546299-1533333, LINC01947-828903*		Intergenic	rs1237684147	NA
165515747	T	TC	*LOC102546299-1545757, LINC01947-816481*		Intergenic	rs199650708	NA
165553405	TA	T	*LOC102546299-1583416, LINC01947-778822*		Intergenic	rs748276665	NA
165635743	TGTC	T	*LOC102546299-1665754, LINC01947-696482*		Intergenic	rs1458238961	NA
165659739	TACAC	T	*LOC102546299-1689750, LINC01947-672485*		Intergenic	rs199695683	NA
165671654	GA	G	*LOC102546299-1701665, LINC01947--660572*		Intergenic	rs61629867	NA
165676442	C	CAT	*LOC102546299-1706452, LINC01947-655786*		Intergenic	rs10672721	NA
165682277	CAAAAAAAAAA	C	*LOC102546299-1712288, LINC01947-649941*		Intergenic	rs755537086	NA
165761890	CT	C	*LOC102546299-1791901, LINC01947--570336*		Intergenic	rs138593792	NA
165763317	A	AACTCCCT	*LOC102546299-1793327, LINC01947-568911*		Intergenic	rs765202825	NA
165763318	TGGTGGTTGTTTTGG	T	*LOC102546299-1793329, LINC01947-568896*		Intergenic	rs759651081	NA
165797381	CTT	C	*LOC102546299-1827392, LINC01947-534845*		Intergenic	rs796788399	NA
165827595	C	CAA	*LOC102546299-1857605, LINC01947-504633*		Intergenic	rs373715734	NA
165828495	GA	G	*LOC102546299-1858506, LINC01947-503732*		Intergenic	rs371246458	NA
165830937	C	CTTTTCT	*LOC102546299-1860947, LINC01947-501291*		Intergenic	rs112842214	NA
165881991	T	TAA	*LOC102546299-1912001, LINC01947-450237*		Intergenic	rs535077310	NA
165952561	GTTTTTGT	G	*LOC102546299-1982572, LINC01947-379660*		Intergenic	rs142804473	NA
165957854	TACACACACAC	T	*LOC102546299-1987865, LINC01947-374364*		Intergenic	rs1478780605	NA
165979701	TTTTG	T	*LOC102546299-2009712, LINC01947-352523*		Intergenic	rs139825131	NA
165996398	G	GGT	*GABRG2-5020989, TENM2-409673*		Intergenic	rs768463363	NA
166004999	A	AT	*LINC01947-327229*		Intergenic	rs397971959	NA
166015272	C	CA	*LINC01947-316956*		Intergenic		
166020527	G	GT	*LINC01947-311701*		Intergenic		
166023712	CTGTG	C	*LINC01947-308512*		Intergenic	rs148113429	NA
166190690	A	ATT	*LINC01947-141538*		Intergenic	rs560312898	NA
166203999	G	GA	*LINC01947-128229*		Intergenic		
166346544	T	TG	*LINC01947*		Intron	rs57522991	NA
166393766	A	AT	*LINC01947-40390, LOC101927908-198174*		Intergenic		
166405945	C	CTGTG	*LINC01947-52569, LOC101927908-185995*		Intergenic	rs3042089	NA
166471042	A	AGAAG	*LINC01947-117666, LOC101927908-120898*		Intergenic		
166685235	GTA	G	*LOC101927908-89829, TENM2-26607*		Intergenic	rs150607521	NA
166689396	CT	C	*TENM2*	NM_001122679	Intron	rs36035281	NA

166689412	GC	G	*LOC101927908-94006, TENM2-22431*		Intergenic	rs144800532	NA
166691111	CAA	C	*LOC101927908-95705, TENM2-20731*		Intergenic	rs201840688	NA
166703356	TTTTGTGTG	T	*LOC101927908-107950, TENM2-8480*		Intergenic	rs57312851	NA
166706519	CT	C	*TENM2*	NM_001122679	Intron	rs56332524	NA
166707750	GTCTA	G	*LOC101927908-112344, TENM2-4090*	NM_001122679	Upstream: 4093 bases	rs148852752	NA
166708279	A	AT	*LOC101927908-112872, TENM2-3565*	NM_001122679	Upstream: 3564 bases	rs761224474	NA
166780506	G	GT	*TENM2*	NM_001122679	Intron	rs144072908	NA
166806937	C	CAAA	*TENM2*	NM_001122679	Intron	rs10622295	NA
166858953	ATGTATATATATATG	A	*TENM2*	NM_001122679	Intron	rs200951197	NA
166872317	TTATA	T	*TENM2*	NM_001122679	Intron	rs1250687778	NA
166872338	T	TAGGGAGAGAGAG	*TENM2*	NM_001122679	Intron	rs71591182	NA
166872375	A	AGAGTGAGT	*TENM2*	NM_001122679	Intron	rs35699708	NA
166873699	CT	C	*TENM2*	NM_001122679	Intron	rs758982741	NA
166876573	GATAA	G	*TENM2*	NM_001122679	Intron	rs1459693240	NA
166886917	AACAC	A	*TENM2*	NM_001122679	Intron	rs1278844619	NA
166889186	CA	C	*TENM2*	NM_001122679	Intron	rs1276165786	NA
166891250	TAG	T	*TENM2*	NM_001122679	Intron	rs373765222	NA
166902953	G	GA	*TENM2*	NM_001122679	Intron	rs544112255	0.00260
166905994	A	ATGTG	*TENM2*	NM_001122679	Intron	rs199668382	NA
166934507	TGAA	T	*TENM2*	NM_001122679	Intron	rs542883191	0.00140
166937743	GTTGT	G	*TENM2*	NM_001122679	Intron	rs150214139	NA
166942675	G	GTATATATATACACATATATACGTATATA	*TENM2*	NM_001122679	Intron	rs139007961	NA
166959832	G	GTGTA	*TENM2*	NM_001122679	Intron	rs148505748	NA
166977695	TG	T	*TENM2*	NM_001122679	Intron	rs199684611	NA
166977698	TG	T	*TENM2*	NM_001122679	Intron	rs1349509320	NA
166977705	AGTGTGTGTGTGT	A	*TENM2*	NM_001122679	Intron	rs1159247676	NA
166995652	G	GGTGTGT	*TENM2*	NM_001122679	Intron	rs57611545	NA
167000901	G	GAA	*TENM2*	NM_001122679	Intron	rs1253510440	NA
167008927	GA	G	*TENM2*	NM_001122679	Intron	rs546931095	NA
167047555	C	CTT	*TENM2*	NM_001122679	Intron	rs769213109	NA
167067625	C	CAA	*TENM2*	NM_001122679	Intron	rs777939339	NA
167109100	ATCCC	A	*TENM2*	NM_001122679	Intron	rs1210507799	NA
167117927	GTA	G	*TENM2*	NM_001122679	Intron	rs1475532024	NA
167129546	A	ATG	*TENM2*	NM_001122679	Intron	rs765624083	NA

167143890	ATATTTTATTTTATTTTATTTTATTTTATTT	A	*TENM2*	NM_001122679	Intron	rs200661613	NA
167173681	C	CAGAGAGAGAG	*TENM2*	NM_001122679	Intron	rs367714833	NA
167234638	A	ATTTT	*TENM2*	NM_001122679	Intron	rs766988937	NA
167241459	C	CA	*TENM2*	NM_001122679	Intron		
167258491	GTT	G	*TENM2*	NM_001122679	Intron	rs796928868	NA
167258989	GAA	G	*TENM2*	NM_001122679	Intron	rs200330152	NA
167261646	CTTT	C	*TENM2*	NM_001122679	Intron	rs555150513	NA
167289247	ATG	A	*TENM2*	NM_001122679	Intron	rs138403142	NA
167293503	C	CA	*TENM2*	NM_001122679	Intron	rs374953163	NA
167299551	AAAGAGAAAGAAAGAAAGAAAGAAAGAAAG	A	*TENM2*	NM_001122679	Intron	rs762478144	NA
167336916	C	CT	*TENM2*	NM_001122679	Intron		
167357776	T	TA	*TENM2*	NM_001122679	Intron	rs199742305	NA
167358557	GA	G	*TENM2*	NM_001122679	Intron	rs368486009	NA
167410189	GA	G	*TENM2*	NM_001122679	Intron	rs61167885	NA
167427948	CAGG	C	*TENM2*	NM_001122679	Intron	rs3056562	NA
167427952	TCCAATGAGACCTGCATTGGACTGC	T	*TENM2*	NM_001122679	Intron	rs57784685	NA
167430307	GAA	G	*TENM2*	NM_001122679	Intron	rs72452959	NA
167431506	CACGCAT	C	*TENM2*	NM_001122679	Intron	rs1413441561	NA
167431515	GCGCGCGCGCACACA	G	*TENM2*	NM_001122679	Intron		
167431523	G	GCGCACA	*TENM2*	NM_001122679	Intron	rs377379682	NA
167448781	CTT	C	*TENM2*	NM_001122679	Intron	rs10617266	NA
167474890	GGCA	G	*TENM2*	NM_001122679	Intron	rs370697921	NA
167517224	CCACA	C	*TENM2*	NM_001122679	Intron	rs780244391	NA
167527930	TA	T	*TENM2*	NM_001122679	Intron	rs546232452	NA
167529059	TTGTGTG	T	*TENM2*	NM_001122679	Intron	rs5873087	NA
167531009	GGTTT	G	*TENM2*	NM_001122679	Intron	rs67606445	NA
167542312	A	AGGAAGGAAG	*TENM2*	NM_001122679	Intron	rs55957924	NA
167575879	T	TGATA	*TENM2*	NM_001122679	Intron	rs752440206	NA
167582372	CA	C	*TENM2*	NM_001122679	Intron	rs35543060	NA
167582409	AGG	A	*TENM2*	NM_001122679	Intron	rs5873092	NA
167583253	TA	T	*TENM2*	NM_001122679	Intron	rs1265073253	NA
167672578	TA	T	*TENM2*	NM_001122679	Intron	rs35630506	NA
167694074	GAAAAAAAA	G	*TENM2-2912, WWC1-24984*	NM_001122679	Downstream: 2912 bases	rs76152547	NA
167703577	T	TGTGA	*TENM2-12414, WWC1-15489*		Intergenic	rs368554513	NA
167711114	A	AAAAG	*TENM2-19951, WWC1-7952*		Intergenic	rs796952166	NA
167737843	A	AAAAAAAAAAAAC	*WWC1*	NM_015238	Intron	rs371515656	NA
167763577	CA	C	*WWC1*	NM_015238	Intron	rs57339649	NA

167778150	C	CA	*WWC1*	NM_015238	Intron		
167837937	C	CTT	*WWC1*	NM_015238	Intron	rs1356706239	NA
167841880	TCT	T	*WWC1*	NM_015238	Intron	rs35168203	NA
167871868	T	TA	*WWC1*	NM_015238	Intron		
167884146	AT	A	*WWC1*	NM_015238	Intron	rs34063177	NA
167975041	TAC	T	*FBLL1-17402, PANK3-7586*		Intergenic	rs1261046562	NA
168024735	C	CT	*PANK3-18120, SLIT3-64004*		Intergenic		
168031453	C	CAAA	*PANK3-24838, SLIT3-57286*		Intergenic	rs777783928	NA
168284379	C	CTTTTG	*LINC01932*	NM_003062	Intron	rs143735267	NA
168324652	G	GCGCGCGCGCGCGCGCA CACACACACACA	*SLIT3*	NM_003062	Intron		
168327621	CAAACAAAACAAAACA	C	*SLIT3*	NM_003062	Intron	rs141764198	NA
168388864	ATG	A	*SLIT3*	NM_003062	Intron	rs1379803871	NA
168442937	AACACAC	A	*SLIT3*	NM_003062	Intron	rs554924575	NA
168455250	G	GA	*SLIT3*	NM_003062	Intron	rs113331358	NA
168459605	G	GT	*LOC728095*	NM_003062	Intron		
168461918	AGT	A	*LOC728095*	NM_003062	Intron	rs751349769	NA
168471593	TG	T	*SLIT3*	NM_003062	Intron	rs57256659	NA

Table A2.7. Primer sequences for Sanger confirmation of novel or rare variants (MAF<0.005) in NIH34-C1

Amplicon	Forward primer 5' -3'	Reverse primer 5' -3'
CYFIP +208	AGGATCCAACTGGACAACGTG	CTTGGGAGGAGGTCTTGGG
PWWP2A +1735	GTGAGCCCGAAAGTACTGC	CAGCTCAGGAACTGTAGCAGC
EBF1 +3122	ATACACAACTAGGAGCTGCCATC	TGCAGCTTAAATGAAGTCATCTC
CNOT8 +1058	GATGGATTAATGTGAGTAACAGG	GACCATTCATTCACTTGAGTG
EBF1 +1437	TAGGTGACCAAACAAATACGCAC	CTCTGAGGATGCATTTGCCTG
C5ORF54 +322	CCTCTTCTGATTTCATTAACCTC	CTGCCTAGTTGCTATCATGAG
HAVCR1 L34L	CAATAGCTTATAGCGTGTGTCC	GAAATAGTGAGGGCTCATTCTCAG

Table A2.8. Cloning primer sequences for amplification of UTRs for reporter luciferase assay

Amplicon	Primer sequence 5' -3'	Amplicon Size
EBF1-3'UTR-SacI-F	GAGCCGAGCTCAAGAATTGCCTTGAAGAATTG	3196 bp
EBF1-3'UTR-MluI-R	GGCTCACGCGTTGCCATCACAGACAGGAAGTAAC	
CNOT8-3'UTR-SacI-F	GAGCCGAGCTCTGGCGCCAGGCTCTGCAGGGTG	1353 bp
CNOT8-3'UTR-MluI-R	GGCTCACGCGTTGCCAGTGGTTAAAAAGAAATC	
ZBED8-3'UTR-SacI-F	GAGCCGAGCTCGCTGATATACTTTTTTAATG	558 bp
ZBED8-3'UTR-HindIII-R	GGCTCAAGCTTTTAATATATTAGGCTATTTA	
SOX30-5'UTR-NheI-F	GTTAGATGCTAGCCCACTGGGGGATTTTTAAG	941 bp
SOX30-5'UTR-HindIII-R	GGCTCTAAGCTTGGGGGAGGGGGACGCCCCGG	

Table A2.9 Site-directed mutagenesis primer sequence for 3'UTR reporter luciferase assay

Amplicon	Primer sequence 5' -3'
EBF1-SDM+1437T>C-F	CAATATCACAAAAAGATTTTACCGCGTATTTTGCAAAGAAAAAAAG
EBF1-SDM+1437T>C-R	CTTTTTTTCTTTGCAAAATACGCGGTAAAATCTTTTTGTGATATTG
CNOT8-SDM+1058G>A-F	GTAATATCCCACAACTGGGATAGGAAGCTCAGGACTTTTTTC
CNOT8-SDM+1058G>A-R	GAAAAAAGTCCTGAGCTTCCTATCCCAGTTGTGGGATATTAC
ZBED8-SDM+322G>A-F	CTAGGAATTGAGCCAGGAAAAAAGAGAAGATATCCCAG
ZBED8-SDM+322G>A-R	CTGGGATATCTTCTCTTTTTTCCTGGCTCAATTCCTAG

Appendix III for Chapter 3

Table S3.1. Primers used for sub-cloning SOX30 Isoform I

pcDNA3.1(+) → pcDNA3.1(+)-SOX30	
SOX30-NheI-F	AAGCTGGCTAGCGCCACCATGGAGAGAGCCAGACCCG
SOX30-HindIII-R	GGTACCAAGCTTTTATAAATCCCTGAGCAC
p3XFLAG-CMV™-10 → pCMV-3XFLAGSOX30-10	
3XFLAG-SOX30-HindIII-F	GATGACAAGCTTATGGAGAGAGCCAGACCCGAG
3XFLAG-SOX30-EcoRI-R	CGATGAATTCCGCTTATAAATCCCTGAGCACTTTTTC
pIRES2-EGFP → pIRES2-EGFP-SOX30	
SOX30-NheI-F	AAGCTGGCTAGCGCCACCATGGAGAGAGCCAGACCCG
3XFLAG-SOX30-EcoRI-R	CGATGAATTCCGCTTATAAATCCCTGAGCACTTTTTC
pIRES2-EGFP-SOX30 → pIRES2-EGFP-3XFLAGSOX30	
NdeI and EcoRI restriction enzymes used to digest pCMV-3XFLAGSOX30-10 and inserted into pIRES2-EGFP-SOX30.	
pCMV Tag4a → pCMV Tag4a SOX30	
pCMV-tag4A-SOX30-F	CTGCAGGAATTCGCCACCATGGAGAGAGCCAGACCCGAG
pCMV-tag4A-SOX30-R	CTCGAGGTCGACTAAATCCCTGAGCACTTTTTCTTC
pCMV Tag4a SOX30 → pCMV mCherry-Tag4a-SOX30	
mCherry cDNA is excised from vector pIRES-Centrin1-mCherry (Addgene # 64338) using restriction enzymes NheI – EcoRI and inserted into pCMV-Tag4a-SOX30	

Table S3.2. SDM primer sequences for pcDNA3.1(+)-SOX30

Amplicon	Primer sequence 5' -3'
Sox30-P82R-F	GGTGTTGCTGCTACGACAGCCTCAGGCCC
Sox30-P82R-R	GGGCCTGAGGCTGTCGTAGCAGCAACACC
Sox30-P123T-F	GCCACCTCCAGGACCGAGTTGCACC
Sox30-P123T-R	GGTGCAACTCGGTCCTGGAGGTGGC
Sox30-P123S-F	GCCACCTCCAGGTCCGAGTTGCACC
Sox30-P123S-R	GGTGCAACTCGGACCTGGAGGTGGC
Sox30-A228T-F	CGAGGACTGCAGGCTCGGCACGGAGCCCGCGTCC
Sox30-A228T-R	GGACGCGGGCTCCGTGCCGAGCCTGCAGTCCTCG
Sox30-A231P-F	GGCGCGGAGCCCCCGTCCAATGGCC
Sox30-A231P-R	GGCCATTGGACGGGGGCTCCGCGCC
Sox30-P353R-F	CAAGGATCCACCGACGAGCACTAGCCAAAGC
Sox30-P353R-R	GCTTTGGCTAGTGCTCGTCGGTGGATCCTTG
Sox30-W404X-F	CAGAGAGGAATTTCCTGGTTGAGTTTATCAGCCTCGTCCAGG
Sox30-W404X-R	CCCTGGACGAGGCTGATAAACTCAACCAGGAAATTCCTCTCT
Sox30-P564A-F	CAACTTCGACCATCCAAGCTCCTAGGGAGTATTCC
Sox30-P564A-R	GGAATACTCCCTAGGAGCTTGGATGGTCGAAGTTG
Sox30-V571F-F	CCTCCTAGGGAGTATTCCAGCTTTTCCCCTTGTCCCAGAAGTG
Sox30-V571F-R	CACTTCTGGGACAAGGGGAAAAGCTGGAATACTCCCTAGGAGG
Sox30-S611P-F	GGACACCACCAAGATTCCCTTTTCATCACCCTTAC
Sox30-S611P-R	GTAAGGGTGATGAAAAGGGAATCTTGGTGGTGTCC
Sox30-Y638C-F	GGCCTCCCTTTGGCTGTGGAAATTTTCCGAG
Sox30-Y638C-R	CTCGGAAAATTTCCACAGCCAAAGGGAGGCC

Sox30-M645I-F	ATGGAAATTTTCCGAGTTCAATACCAGAATGCCTTAGTTATTATG
Sox30-M645I-R	CATAATAACTAAGGCATTCTGGTATTGAACTCGGAAAATTTCCAT
Sox30-D654N-F	GAATGCCTTAGTTATTATGAAAACAGGTACCCAAAACATGAGG
Sox30-D654N-R	CCTCATGTTTTGGGTACCTGTTTTCATAATAACTAAGGCATTC
Sox30-N667S-F	GAGGGTATCTTTTCAACTTTAAGTAGAGACTATTCTTTTAGAGAC
Sox30-N667S-R	GTCTCTAAAAGAATAGTCTCTACTTAAAGTTGAAAAGATACCCTC
Sox30-N667D-F	GAGGGTATCTTTTCAACTTTAGATAGAGACTATTCTTTTAGAG
Sox30-N667D-R	CTCTAAAAGAATAGTCTCTATCTAAAGTTGAAAAGATACCCTC

Table S3.3. SOX30 cDNA sequencing primers

Amplicon	Forward primer 5' -3'	Reverse primer 5' -3'
SOX30-1	GTCGAGGGCACCTCCTTT	CTCGTCCCCTCGGAAGTAG
SOX30-2	ATGTCAAGGCCAAGAAGCAG	TTTAGTAGGCACTGGTGTCAGG
SOX30-3	GCACCAGCAAGACCTTAGGA	CACTTAGAGGGAATCGTTTTCG
SOX30-4	CCAGCTTGGGTTAGAGTGGA	AGGTTGGATGGTCGAAGTTG
SOX30-5	GCCTTCCCAAACAGACACTC	CCTCATGTTTTGGGTACCTG
SOX30-6	ACTTCCTACCCGGACCTCAC	TTATAAATCCCTGAGCACTTTTTC

Table S3.4. Vector sequencing primers

N-CMV-30 Seq FP	AATGTCGTAATAACCCCGCCCCGTTGACGC
C-CMV-24 Seq RP	TATTAGGACAAGGCTGGTGGGCAC
Universal M13 RP	CAGGAAACAGCTATGAC
M13 (-21) FP	TGTAAAACGACGGCCAGT
pGL3-GLprimer2	CTTTATGTTTTTGGCGTCTTCCA
pGL3-RVprimer3	CTAGCAAAATAGGCTGTCCC
pGL3-RVprimer4	GACGATAGTCATGCCCCGCG
BGH Reverse	TAGAAGGCACAGTCGAGG
CMV Forward	CGCAAATGGGCGGTAGGCGTG

Table S3.5. Chip-PCR and qRT-PCR primers for the SOX30 targets

Gene	Primer sequence 5' – 3'	Product length	Application
BCL2L1-F	GCGATGGAGGAGGAAGCAAG	157 bp	Chip-PCR / qRT-PCR
BCL2L1-R	GGGATGCCGGTAACTCAGC		
BRWD1-F	CGCAGGGATTCGGACCAC	126 bp	Chip-PCR / qRT-PCR
BRWD1-R	GAGGCGAACCTCGCTCAG		
ZSCAN9-F	GTCCGCCTTCGAGGAGAG	123 bp	Chip-PCR / qRT-PCR
ZSCAN9-R	AACTCACGAGACCCGGAAC		
STX16-F	ATGGAGGGAGCCGGAGAAG	111 bp	Chip-PCR / qRT-PCR
STX16-R	TTCCTTTCCTGGACGGCCTG		
MYOD-F	CCTCTTTCGGTCCCTCTTTC	106 bp	Chip-PCR
MYOD-R	ATGGGTAGAGCGGCTGTAGA		
GAPDH-F	TCACCACCATGGAGAAGGCT	256 bp	qRT-PCR
GAPDH-R	AAGCAGTTGGTGGTGCAGGA		

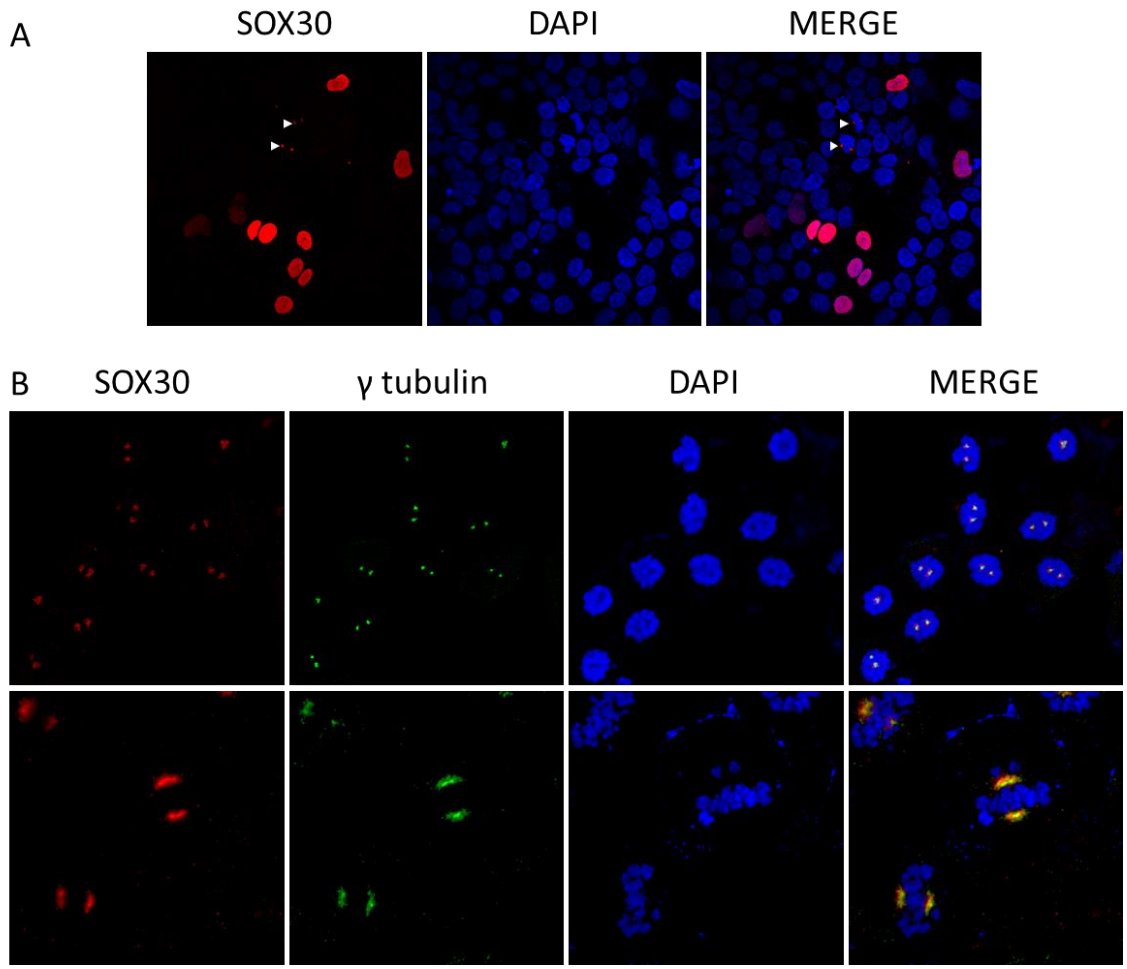
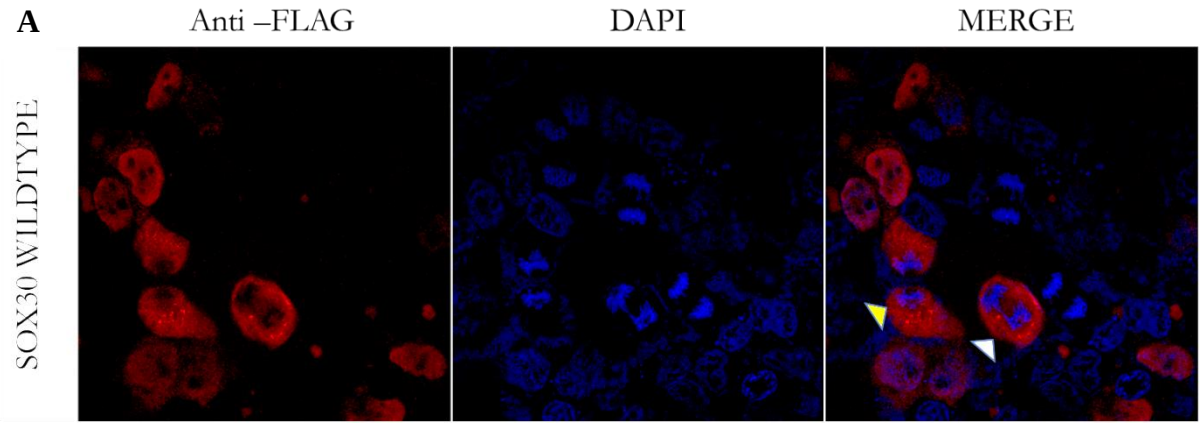


Figure S3.1. (A) SOX30 was detected in fixed HEK293 cells overexpressing SOX30 using anti-SOX30 affinity purified polyclonal antibody at 1:1000 dilution for 1 hour at room temperature. Cells were stained using the Alexafluor 568-conjugated anti-rabbit IgG secondary antibody (red) and counterstained with DAPI (blue). Specific staining was localized to nucleus and in spindle poles of mitotic cells (white arrow. (B) Nocodazole arrested HEK293 cells stained using SOX30 affinity purified polyclonal antibody (red) and gamma tubulin (green) to mark spindle poles indicated SOX30 to colocalize with gamma tubulin that is inconsistent with SOX30's subcellular localization. Upper panel at 63X and lower panel 63x with zoom 4.



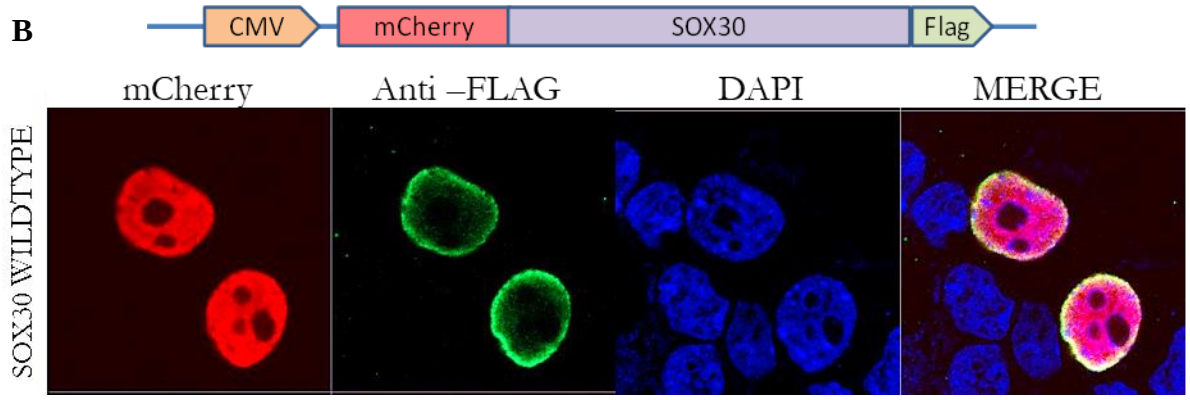
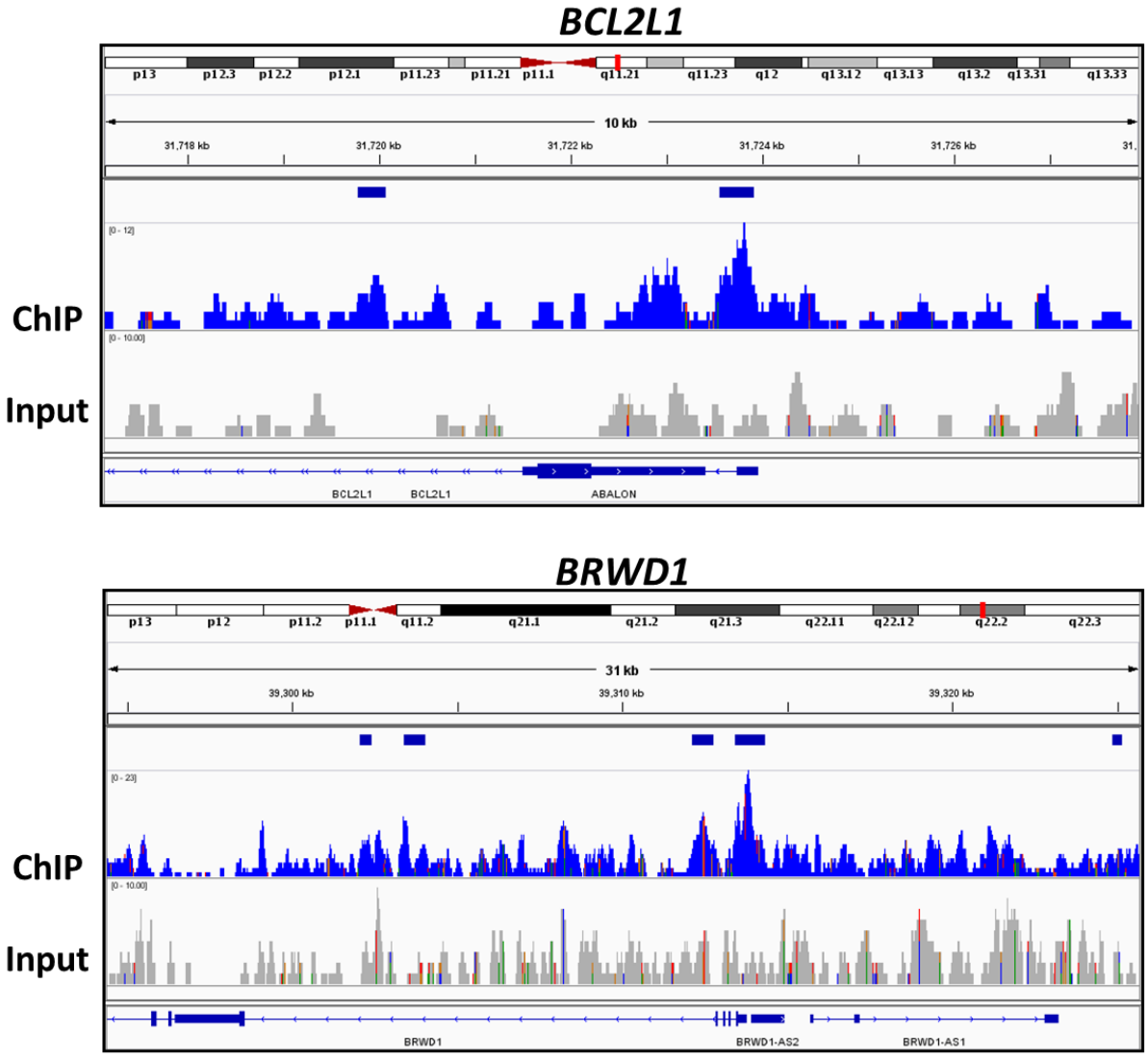


Figure S3.2. (A) Subcellular localization of FLAG tagged SOX30 protein in cultured HEK293 cells. Mitotic cells anaphase (white arrow) and telophase (yellow arrow) exhibit cytoplasmic SOX30 staining. (**B)** Subcellular localization of mCherry tagged SOX30 protein in HEK293, also detected using FLAG antibody (green). Nucleus stained with DAPI (blue).



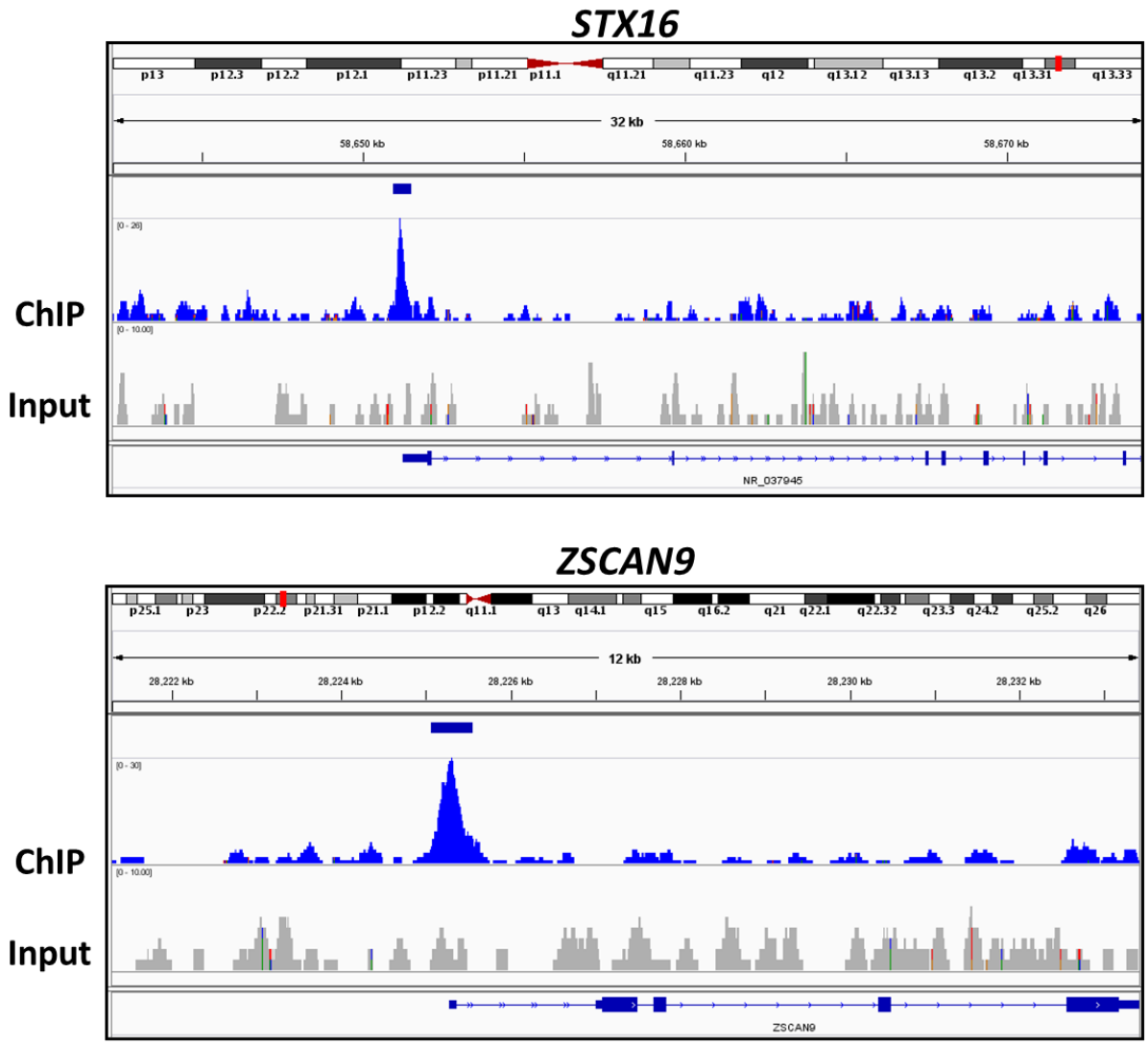


Figure S3.3. Integrative Genomics Viewer (IGV) snapshots depicting SOX30 binding sites for genes *ZSCAN9*, *BCL2L1*, *BRWD1* and *STX16* at 5' UTR or promoter regions of the genes. ChIP peaks tracks (Blue) and input peak tracks (Gray). The RefSeq gene map is presented in blue at the bottom of each panel.

Table S3.6. Sample C Top 500 ChIP-Seq annotated peaks

Sl.No.	Chr	Peak location	Peak ID	Score	Location annotation	Exon/Intron no.	Distance To TSS	Gene
1	**chr5**	**157651738**	**41909**	**129.51863**	**Exon**	**exon 1 of 4**	**19742**	***SOX30***
2	chr21	8420457	28884	72.05733	Distal Intergenic		211984	*MIR3648-1*
3	chr17	26603916	18500	69.45684	Distal Intergenic		690080	*MIR4522*
4	chr1	125180210	3052	53.35207	Distal Intergenic		-3994671	*FAM72B*
5	chr1	143247073	3056	46.43203	Distal Intergenic		452546	*RNVU1-17*
6	chr21	8992949	28888	25.28313	Distal Intergenic		784476	*MIR3648-1*
7	chr1	143263577	3058	21.354	Distal Intergenic		436042	*RNVU1-17*
8	chr21	8214717	28881	20.50926	Intron	intron 2 of 6	6244	*MIR3648-1*
9	chr21	10402816	28965	19.16683	Distal Intergenic		-79922	*TPTE*
10	chr16	34576235	16080	17.97919	Distal Intergenic		-416199	*LINC00273*
11	**chr20**	**58651169**	**28382**	**16.4886**	**Promoter (<=1kb)**		**-84**	***STX16***
12	chr15	20340974	13323	15.78969	Distal Intergenic		200826	*GOLGA6L6*
13	**chr6**	**28225293**	**42838**	**14.68506**	**Promoter (<=1kb)**		**407**	***ZSCAN9***
14	chr21	10704978	29060	14.51823	Distal Intergenic		222240	*TPTE*
15	chr21	10702671	29059	14.44255	Distal Intergenic		219933	*TPTE*
16	chr11	118436188	8548	13.08067	Promoter (<=1kb)		-302	*KMT2A*
17	chr21	8988848	28886	12.79437	Distal Intergenic		780375	*MIR3648-1*
18	chr21	10737076	29076	12.79437	Distal Intergenic		254338	*TPTE*

19	chr21	10689760	29053	12.55031	Distal Intergenic		207022	*TPTE*
20	chr8	37697982	46697	12.5399	Promoter (2-3kb)		2231	*ZNF703*
21	**chr21**	**39313839**	**32865**	**12.43945**	**Promoter (<=1kb)**		**-96**	***BRWD1-IT2***
22	chr21	7259010	28846	12.10423	Distal Intergenic		-485952	*SMIM11B*
23	chr21	7952388	28869	12.01716	Distal Intergenic		207426	*SMIM11B*
24	chr21	9027159	28897	11.59718	Distal Intergenic		786850	*LINC01667*
25	chr21	10754261	29079	11.59718	Distal Intergenic		271523	*TPTE*
26	chr21	41337704	33617	11.49507	Intron	intron 3 of 9	-24239	*MX2*
27	chr21	7918252	28854	11.45431	Distal Intergenic		173290	*SMIM11B*
28	chr12	113300088	10085	11.40801	Intron	intron 16 of 16	8565	*MIR6762*
29	chr16	34581974	16081	11.40447	Distal Intergenic		-421938	*LINC00273*
30	chr21	39022271	32789	11.22124	Distal Intergenic		-44497	*LINC01700*
31	chr21	39117039	32828	11.22124	Distal Intergenic		66812	*PSMG1*
32	chr21	10542594	28983	11.22124	Intron	intron 16 of 31	59856	*TPTE*
33	chr21	10671198	29043	11.02854	Distal Intergenic		188460	*TPTE*
34	chr21	10654594	29035	10.9821	Distal Intergenic		171856	*TPTE*
35	chr21	10730564	29073	10.9323	Distal Intergenic		247826	*TPTE*
36	chr21	37709389	32297	10.83428	Intron	intron 3 of 3	207057	*KCNJ6*
37	chr21	19524551	29817	10.79659	Distal Intergenic		-764548	*MIR548XHG*
38	chr21	41304832	33602	10.78279	Promoter (<=1kb)		620	*FAM3B*
39	chr21	42010891	33875	10.77097	Promoter (<=1kb)		-504	*ZBTB21*
40	chr21	42183797	33928	10.71641	Distal Intergenic		-15892	*ABCG1*
41	chr21	10707756	29062	10.61335	Distal Intergenic		225018	*TPTE*
42	chr1	234532307	4585	10.49227	Promoter (<=1kb)		-528	*LINC01354*
43	chr21	37616185	32264	10.42969	3' UTR		250395	*DYRK1A*
44	chr22	49305357	36754	10.42969	Distal Intergenic		238109	*MIR3667*
45	chr21	9838123	28927	10.42969	Distal Intergenic		-24114	*LINC01667*
46	chr17	10917563	18038	10.42969	Distal Intergenic		-79118	*PIRT*
47	chr21	10379883	28957	10.42969	Distal Intergenic		-102855	*TPTE*
48	chr21	14056972	29185	10.42969	Intron	intron 1 of 2	153919	*LIPI*
49	chr21	38882746	32730	10.42969	Intron	intron 3 of 3	73721	*LOC400867*
50	chr21	40849113	33427	10.3902	Promoter (1-2kb)		-1974	*DSCAM*
51	chr21	10706921	29061	10.35381	Distal Intergenic		224183	*TPTE*
52	chr1	29237275	1283	10.26126	Promoter (<=1kb)		759	*PTPRU*
53	chr21	38595539	32627	10.19838	Intron	intron 1 of 11	66241	*ERG*
54	chr21	9735713	28907	10.15113	Distal Intergenic		78296	*LINC01667*
55	chr21	39919911	33073	10.13599	Intron	intron 2 of 2	52594	*PCP4*
56	chr20	31514579	27246	10.13599	Promoter (<=1kb)		151	*HM13*
57	chr17	63842748	19436	10.13391	Promoter (<=1kb)		192	*SMARCD2*
58	chr17	26881403	18519	10.08321	Distal Intergenic		412593	*MIR4522*
59	chr1	183472039	4057	10.07944	Promoter (<=1kb)		-177	*SMG7*
60	chr21	37257958	32201	10.00298	Intron	intron 1 of 7	9961	*DSCR3*
61	chr1	89525075	2314	10.00298	Promoter (<=1kb)		239	*LRRC8B*
62	chr21	42060053	33888	9.91929	Promoter (2-3kb)		-2906	*UMODL1*
63	chr21	10724630	29071	9.85449	Distal Intergenic		241892	*TPTE*
64	chr21	10721009	29070	9.83936	Distal Intergenic		238271	*TPTE*
65	chr20	28839884	27134	9.7314	Distal Intergenic		-2417780	*DEFB115*
66	chr21	40118859	33146	9.69014	Intron	intron 20 of 32	93572	*MIR4760*
67	chr21	9003211	28890	9.69014	Intron	intron 1 of 1	794738	*MIR3648-1*
68	chr21	40913108	33455	9.67739	Distal Intergenic		-65969	*DSCAM*
69	chr21	17745748	29642	9.67739	Distal Intergenic		46775	*C21orf91-OT1*
70	chr22	34624156	35715	9.67739	Distal Intergenic		-441980	*ISX*
71	chr21	21279582	30017	9.67739	Intron	intron 1 of 17	281267	*NCAM2*
72	chr14	19168404	11535	9.67739	Intron	intron 1 of 1	-187662	*LOC100508046*
73	chr21	35801822	31904	9.67739	Intron	intron 6 of 12	81107	*MIR802*
74	chrX	41084809	50193	9.67739	Promoter (<=1kb)		-826	*USP9X*
75	chr21	8990289	28887	9.60275	Distal Intergenic		781816	*MIR3648-1*
76	chr21	39563702	32923	9.51442	Intron	intron 1 of 4	7260	*B3GALT5*
77	chr17	48614429	18921	9.38781	Promoter (<=1kb)		510	*HOXB8*
78	chr21	10716183	29067	9.35928	Distal Intergenic		233445	*TPTE*
79	chr21	38599449	32628	9.31496	Intron	intron 1 of 11	62331	*ERG*
80	chr22	48319997	36508	9.29406	Distal Intergenic		45633	*MIR3201*
81	chr21	27937904	30663	9.29406	Distal Intergenic		-75459	*LINC00314*
82	chr14	22012795	11650	9.29406	Distal Intergenic		347712	*OR4E2*
83	chr14	20876957	11608	9.29406	Distal Intergenic		-14442	*RNASE3*
84	chr14	20788493	11599	9.29406	Distal Intergenic		7442	*RNASE6*
85	chr15	20276822	13314	9.29406	Distal Intergenic		264978	*GOLGA6L6*
86	chr20	23938076	26905	9.29406	Distal Intergenic		50703	*GGTLC1*
87	chr21	40333602	33230	9.29406	Intron	intron 8 of 32	-49481	*DSCAM-AS1*
88	chr22	47843722	36438	9.29406	Intron	intron 8 of 8	212048	*LOC284930*
89	chr7	64984901	45174	9.29406	Intron	intron 1 of 3	21783	*ZNF117*

90	chr1	175946902	3880	9.29406	Intron	intron 8 of 8	21638	*SCARNA3*
91	chr21	10337215	28954	9.29406	Intron	intron 3 of 3	-145523	*TPTE*
92	chr4	18021265	39506	9.29406	Promoter (<=1kb)		611	*LCORL*
93	chr17	64621487	19444	9.2565	Intron	intron 1 of 18	40581	*SMURF2*
94	chr21	30485243	31005	9.2565	Promoter (2-3kb)		2193	*KRTAP19-2*
95	chr1	211259052	4256	9.25521	Promoter (<=1kb)		675	*RCOR3*
96	chr22	49087268	36705	9.18807	Distal Intergenic		220498	*LINC01310*
97	chr21	39938157	33080	9.18807	Distal Intergenic		70840	*PCP4*
98	chr21	40269457	33205	9.18807	Intron	intron 11 of 32	-57026	*MIR4760*
99	chr15	20342060	13324	9.18311	Distal Intergenic		199740	*GOLGA6L6*
100	chr21	10729146	29072	9.14301	Distal Intergenic		246408	*TPTE*
101	chr21	41660077	33753	9.12042	Distal Intergenic		-19104	*LINC00111*
102	chr21	10414312	28969	9.12042	Distal Intergenic		-68426	*TPTE*
103	chr21	39885777	33048	9.02685	Intron	intron 1 of 2	18460	*PCP4*
104	chr21	40854445	33430	9.0235	Distal Intergenic		-7306	*DSCAM*
105	chr1	151941389	3141	8.99813	Distal Intergenic		-31581	*THEM4*
106	chr21	39479255	32898	8.99813	Intron	intron 3 of 6	32245	*MIR6508*
107	chr16	29807233	15959	8.98115	Promoter (1-2kb)		1127	*MAZ*
108	chr21	10677438	29046	8.91675	Distal Intergenic		194700	*TPTE*
109	chr21	10685765	29051	8.91675	Distal Intergenic		203027	*TPTE*
110	chr21	7948035	28867	8.83186	Distal Intergenic		203073	*SMIM11B*
111	chr21	10719461	29069	8.83065	Distal Intergenic		236723	*TPTE*
112	chr21	28611586	30742	8.79821	Intron	intron 2 of 4	72268	*LINC00161*
113	chrX	132489631	51887	8.7491	Promoter (<=1kb)		337	*MBNL3*
114	chr21	39010865	32784	8.7372	Intron	intron 1 of 2	-33091	*LINC01700*
115	chr21	26844914	30562	8.6893	Promoter (<=1kb)		495	*ADAMTS1*
116	chr21	41704250	33772	8.64372	Distal Intergenic		11525	*LINC00479*
117	chr22	48882075	36664	8.62155	Intron	intron 1 of 6	15305	*LINC01310*
118	chr7	31336542	44436	8.61548	Downstream (<1kb)		4352	*NEUROD6*
119	chr22	48902195	36669	8.60379	Distal Intergenic		35425	*LINC01310*
120	chr9	41229833	48253	8.60379	Distal Intergenic		-3922	*MIR4477A*
121	chr21	22702327	30159	8.60379	Distal Intergenic		-376957	*MIR6130*
122	chr21	38440447	32583	8.60379	Intron	intron 2 of 8	-107026	*LINC01423*
123	chr22	33724152	35490	8.60379	Promoter (<=1kb)		-862	*LARGE-AS1*
124	chr6	87155549	43269	8.60379	Promoter (2-3kb)		2716	*ZNF292*
125	chr21	10681353	29048	8.58075	Distal Intergenic		198615	*TPTE*
126	chr1	46303182	1668	8.55622	Promoter (<=1kb)		426	*LRRC41*
127	chr22	48389707	36525	8.54068	Distal Intergenic		-99753	*FAM19A5*
128	chr6	1990261	42364	8.54068	Intron	intron 4 of 10	255431	*GMDS*
129	chr15	65851865	13775	8.54068	Intron	intron 2 of 2	-59572	*DENND4A*
130	chr21	32932303	31387	8.54068	Intron	intron 1 of 1	-93542	*OLIG2*
131	chr21	9839428	28928	8.53997	Distal Intergenic		-25419	*LINC01667*
132	chr21	10587855	29012	8.53997	Intron	intron 26 of 31	105117	*TPTE*
133	chr21	38084572	32439	8.53997	Intron	intron 2 of 3	36788	*DSCR4*
134	chr21	36320443	32034	8.53997	Promoter (<=1kb)		254	*MORC3*
135	chr2	197515985	25603	8.53997	Promoter (<=1kb)		414	*MOB4*
136	chr21	32553484	31362	8.44008	Distal Intergenic		33889	*TCP10L*
137	chr21	10713139	29065	8.42077	Distal Intergenic		230401	*TPTE*
138	chr21	37676550	32284	8.41972	Intron	intron 3 of 3	239896	*KCNJ6*
139	chr21	46112242	34371	8.39034	Exon	exon 3 of 28	14145	*COL6A2*
140	chr21	40157437	33159	8.39034	Intron	intron 16 of 32	54994	*MIR4760*
141	chr21	41413731	33653	8.38315	Distal Intergenic		-6573	*MX1*
142	chr21	10679865	29047	8.36239	Distal Intergenic		197127	*TPTE*
143	chr1	58396468	1847	8.34265	Intron	intron 3 of 20	150266	*DAB1*
144	chr21	39608887	32941	8.34131	Intron	intron 2 of 4	3934	*B3GALT5-AS1*
145	chr20	61312294	28639	8.32695	Intron	intron 2 of 15	233154	*LINC01718*
146	chr3	51384821	37720	8.29279	Promoter (<=1kb)		-226	*MANF*
147	chr21	41646346	33748	8.27148	Distal Intergenic		-32835	*LINC00111*
148	chr21	34654232	31628	8.27148	Distal Intergenic		-15157	*CLIC6*
149	chr21	42372371	33984	8.2473	3' UTR		-5777	*TFF1*
150	chr21	45555792	34309	8.22773	Distal Intergenic		-11381	*SLC19A1*
151	chr21	45075220	34271	8.22773	Promoter (1-2kb)		1367	*ADARB1*
152	chr7	63522060	45115	8.19277	Distal Intergenic		-122851	*LOC100287834*
153	chr15	22265856	13351	8.19277	Distal Intergenic		-40527	*MIR1268A*
154	chr2	89611394	24545	8.19277	Distal Intergenic		799024	*MIR4436A*
155	chr22	47304540	36334	8.19277	Distal Intergenic		182571	*LINC01644*
156	chr1	143202310	3054	8.19277	Distal Intergenic		497309	*RNVU1-17*
157	chr16	26836175	15818	8.19277	Distal Intergenic		-230532	*C16orf82*
158	chr21	40996658	33481	8.19277	Distal Intergenic		-149519	*DSCAM*
159	chr21	9723744	28901	8.19277	Distal Intergenic		90265	*LINC01667*
160	chr1	144101714	3061	8.19277	Distal Intergenic		-129749	*FAM72C*

161	chr1	68892357	2011	8.19277	Distal Intergenic		-395136	*DEPDC1*
162	chr15	20238915	13307	8.19277	Distal Intergenic		302885	*GOLGA6L6*
163	chr16	27138538	15844	8.19277	Distal Intergenic		-64957	*KDM8*
164	chr4	49615453	39596	8.19277	Distal Intergenic		629206	*CWH43*
165	chr22	11614005	34463	8.19277	Distal Intergenic		-3914153	*OR11H1*
166	chr22	34235487	35620	8.19277	Distal Intergenic		-312646	*LARGE1*
167	chr22	34400014	35661	8.19277	Distal Intergenic		-477173	*LARGE1*
168	chr13	19561107	10620	8.19277	Distal Intergenic		-24345	*TPTE2*
169	chr1	175272017	3858	8.19277	Distal Intergenic		-79018	*KIAA0040*
170	chr1	114892410	2780	8.19277	Intron	intron 15 of 31	37607	*SYCP1*
171	chr16	17165637	15319	8.19277	Intron	intron 5 of 11	31133	*LOC102723692*
172	chr19	55841710	23249	8.19277	Intron	intron 1 of 9	-4910	*NLRP11*
173	chr2	29441016	23946	8.19277	Intron	intron 4 of 28	343311	*CLIP4*
174	chr2	158407129	25430	8.19277	Intron	intron 1 of 13	49624	*CCDC148*
175	chr21	28880608	30783	8.19277	Intron	intron 3 of 5	4763	*N6AMT1*
176	chr21	40756541	33381	8.19277	Intron	intron 1 of 32	90598	*DSCAM*
177	chr8	142507421	47821	8.19277	Intron	intron 17 of 30	57991	*ADGRB1*
178	chr16	13180513	15245	8.19277	Intron	intron 2 of 4	278893	*SHISA9*
179	chr21	35418000	31808	8.19277	Intron	intron 9 of 12	162760	*LOC100506403*
180	chr11	78064100	8144	8.19277	Promoter (<=1kb)		239	*THRSP*
181	chr8	103501552	47447	8.19277	Promoter (<=1kb)		804	*RIMS2*
182	chr22	33546199	35446	8.16264	Intron	intron 7 of 15	-109530	*MIR4764*
183	chr17	15741173	18309	8.15857	3' UTR		8926	*TBC1D26*
184	chr22	47399038	36353	8.15857	Distal Intergenic		88073	*LINC01644*
185	chr3	59646692	37866	8.15857	Distal Intergenic		-596608	*C3orf67*
186	chr21	41046313	33505	8.15857	Distal Intergenic		101820	*LINC00323*
187	chr8	52578705	47019	8.15857	Distal Intergenic		-13198	*ALKAL1*
188	chr15	20344058	13325	8.15857	Distal Intergenic		197742	*GOLGA6L6*
189	chr22	11568715	34445	8.15857	Distal Intergenic		-3959443	*OR11H1*
190	chr21	21747307	30068	8.15857	Promoter (<=1kb)		334	*LINC01425*
191	chr21	38805967	32707	8.12195	Promoter (<=1kb)		660	*ETS2*
192	chr7	158856754	46460	8.12195	Promoter (<=1kb)		176	*WDR60*
193	chr1	3977405	730	8.10923	Distal Intergenic		36840	*LINC01346*
194	chr21	29629736	30885	8.10923	Intron	intron 4 of 4	-118439	*GRIK1-AS1*
195	chr21	10693353	29054	8.05771	Distal Intergenic		210615	*TPTE*
196	chr14	100538994	13116	8.05771	Exon	exon 1 of 2	48419	*BEGAIN*
197	chr16	19114615	15398	8.05771	Promoter (<=1kb)		683	*ITPRIPL2*
198	chr21	41111876	33527	8.05648	Distal Intergenic		36257	*LINC00323*
199	chr6	142272590	43656	8.05648	Distal Intergenic		-29264	*ADGRG6*
200	chr21	10638830	29029	8.05648	Distal Intergenic		156092	*TPTE*
201	chr1	178939300	3950	8.05648	Distal Intergenic		-68248	*ANGPTL1*
202	chr21	41203669	33559	8.05648	Intron	intron 1 of 8	-18430	*PLAC4*
203	chr20	38446643	27424	8.05648	Promoter (<=1kb)		65	*SNHG11*
204	chr10	43408952	5253	8.05648	Promoter (<=1kb)		214	*HNRNPF*
205	chr22	50090022	36902	8.05648	Promoter (<=1kb)		143	*MOV10L1*
206	chr6	7542681	42494	8.05648	Promoter (1-2kb)		1106	*DSP*
207	chr21	16954811	29595	8.03489	Distal Intergenic		364574	*MIR125B2*
208	chr20	56417544	28250	8.03489	Intron	intron 2 of 6	5432	*CASS4*
209	chr22	32496436	35247	8.03489	Intron	intron 8 of 8	21760	*FBXO7*
210	chr19	51032735	22931	8.03489	Promoter (2-3kb)		2495	*KLK12*
211	chr21	25664428	30435	8.01212	Intron	intron 1 of 9	25156	*JAM2*
212	chr21	38695413	32666	8.00437	Distal Intergenic		-33633	*ERG*
213	chr1	16094803	1086	7.967	Distal Intergenic		-21171	*FAM131C*
214	chr22	17199362	34607	7.9529	Intron	intron 4 of 9	22627	*ADA2*
215	chr21	39031612	32795	7.91488	Distal Intergenic		-53838	*LINC01700*
216	chr21	24928954	30363	7.91488	Intron	intron 2 of 3	88404	*LINC01692*
217	chr17	35142296	18754	7.91488	Promoter (<=1kb)		19	*NLE1*
218	chr1	93079218	2370	7.86288	Promoter (<=1kb)		-17	*MTF2*
219	chr22	18151822	34722	7.8562	Promoter (1-2kb)		1923	*USP18*
220	chr11	114144170	8467	7.85177	Intron	intron 2 of 6	84577	*ZBTB16*
221	chr15	92883008	14337	7.85177	Promoter (<=1kb)		301	*LINC01578*
222	chr11	68213726	7911	7.85177	Promoter (<=1kb)		102	*KMT5B*
223	chr11	62337438	7839	7.85177	Promoter (<=1kb)		-10	*ASRGL1*
224	chr21	39663923	32959	7.8491	3' UTR		-51102	*B3GALT5-AS1*
225	chr17	25389914	18499	7.8491	Distal Intergenic		1904082	*MIR4522*
226	chr1	161440684	3542	7.8491	Distal Intergenic		-64746	*FCGR2A*
227	chr21	10329745	28951	7.8491	Intron	intron 3 of 3	-152993	*TPTE*
228	chr21	39796975	33012	7.83429	Intron	intron 8 of 8	51568	*IGSF5*
229	chr21	36540574	32084	7.8021	Intron	intron 1 of 2	35995	*CLDN14*
230	chr22	26882165	35026	7.79837	Distal Intergenic		-21127	*LINC01422*
231	chr16	26821547	15817	7.79837	Distal Intergenic		-245160	*C16orf82*

232	chr21	39998513	33103	7.79837	Distal Intergenic		131196	*PCP4*
233	chr19	29424509	21954	7.79837	Intron	intron 2 of 9	101243	*LOC284395*
234	chr21	41935122	33852	7.79837	Intron	intron 2 of 13	18768	*C2CD2*
235	chr3	93980591	38338	7.79837	Promoter (<=1kb)		452	*ARL13B*
236	chr22	11619637	34465	7.79799	Distal Intergenic		-3908521	*OR11H1*
237	chr19	30226288	22058	7.79091	Promoter (2-3kb)		-2226	*ZNF536*
238	chr5	179413662	42239	7.75301	Distal Intergenic		-68232	*ADAMTS2*
239	chrX	24025111	50002	7.75301	Promoter (2-3kb)		2075	*KLHL15*
240	chr21	7937394	28863	7.74137	Distal Intergenic		192432	*SMIM11B*
241	chr21	40869426	33436	7.74137	Distal Intergenic		-22287	*DSCAM*
242	chr14	34529837	11916	7.74137	Intron	intron 3 of 5	9874	*EAPP*
243	chr21	41213380	33565	7.74137	Intron	intron 1 of 8	-28141	*PLAC4*
244	chr21	17513073	29628	7.72672	Promoter (<=1kb)		691	*CXADR*
245	chr20	62143548	28769	7.70448	Promoter (<=1kb)		-108	*PSMA7*
246	chr22	49494700	36805	7.68517	Intron	intron 1 of 4	48766	*MIR3667*
247	chr20	35988063	27380	7.68517	Intron	intron 7 of 11	-32704	*SCAND1*
248	chr5	1264958	40157	7.68517	Intron	intron 6 of 10	30111	*TERT*
249	chr22	10727323	34403	7.65651	Distal Intergenic		-4800835	*OR11H1*
250	chr21	41736434	33784	7.65651	Downstream (2-3kb)		10407	*MIR6814*
251	chr9	90842478	48676	7.65651	Intron	intron 1 of 13	40691	*SYK*
252	chr9	79573204	48470	7.62977	Promoter (1-2kb)		1431	*TLE4*
253	chr21	17819243	29649	7.62154	Promoter (<=1kb)		143	*C21orf91*
254	chr21	44939969	34257	7.6144	Promoter (<=1kb)		-41	*FAM207A*
255	chr21	39530861	32914	7.60933	Distal Intergenic		-25581	*B3GALT5*
256	chr7	153808310	46238	7.60933	Distal Intergenic		-78787	*DPP6*
257	chr21	38699411	32668	7.60933	Distal Intergenic		-37631	*ERG*
258	chr7	60019554	45059	7.60933	Distal Intergenic		2569377	*ZNF716*
259	chr21	39942940	33082	7.60933	Distal Intergenic		75623	*PCP4*
260	chr21	10663763	29039	7.60933	Distal Intergenic		181025	*TPTE*
261	chr22	11605723	34461	7.60933	Distal Intergenic		-3922435	*OR11H1*
262	chr21	39749487	32991	7.60933	Intron	intron 2 of 8	4080	*IGSF5*
263	chr21	40154388	33157	7.60933	Intron	intron 16 of 32	58043	*MIR4760*
264	chr22	33253631	35383	7.60933	Intron	intron 5 of 5	145102	*LINC01640*
265	chr22	48837461	36655	7.60933	Intron	intron 3 of 3	-29309	*LINC01310*
266	chr22	35475068	35859	7.60907	Distal Intergenic		-65800	*RASD2*
267	chr20	63956024	28840	7.60907	Promoter (<=1kb)		391	*UCKL1*
268	chr1	228103195	4471	7.6035	Promoter (2-3kb)		2216	*C1orf35*
269	chr3	66500248	37987	7.58019	Promoter (1-2kb)		1015	*LRIG1*
270	chr10	114073075	6328	7.5751	Distal Intergenic		29019	*ADRB1*
271	chr21	20990437	29994	7.5751	Distal Intergenic		-7878	*NCAM2*
272	chr6	21589448	42727	7.5751	Distal Intergenic		-3321	*SOX4*
273	chr17	6853196	17800	7.5751	Distal Intergenic		-21435	*TEKT1*
274	chr22	33935562	35543	7.5751	Distal Intergenic		-12721	*LARGE1*
275	chr21	40828787	33418	7.5751	Intron	intron 1 of 32	18352	*DSCAM*
276	chr22	33783478	35507	7.5751	Intron	intron 2 of 15	58464	*LARGE-AS1*
277	chr22	44268588	36104	7.5751	Intron	intron 4 of 4	44263	*KIAA1644*
278	chr19	53048543	23104	7.5751	Intron	intron 1 of 1	3805	*ERVV-2*
279	chr21	38280168	32520	7.5751	Intron	intron 3 of 5	53253	*LINC01423*
280	chr17	21747213	18464	7.56447	Distal Intergenic		54690	*KCNJ18*
281	chr14	101459237	13159	7.56447	Distal Intergenic		17161	*LINC02314*
282	chr7	57389021	45028	7.56447	Distal Intergenic		-15750	*LOC100653233*
283	chr18	27858151	20278	7.56447	Distal Intergenic		-262987	*LOC105372038*
284	chr21	26620992	30538	7.56447	Distal Intergenic		-47708	*CYYR1*
285	chr7	51512702	44832	7.56447	Distal Intergenic		-195884	*COBL*
286	chr17	6171382	17755	7.56447	Distal Intergenic		263817	*AIPL1*
287	chr16	8418930	14884	7.56447	Distal Intergenic		171263	*TMEM114*
288	chr20	25126420	27041	7.56447	Distal Intergenic		22838	*LOC284798*
289	chr14	19866189	11544	7.56447	Distal Intergenic		-10043	*OR4K2*
290	chr9	21727636	47993	7.56447	Distal Intergenic		-74907	*MTAP*
291	chr19	53594883	23152	7.56447	Distal Intergenic		-37173	*DPRX*
292	chr14	52844059	12134	7.56447	Distal Intergenic		-52391	*GNPNAT1*
293	chr21	10735734	29075	7.56447	Distal Intergenic		252996	*TPTE*
294	chr7	56824544	45011	7.56447	Exon	exon 5 of 5	-12098	*LOC401357*
295	chr15	26882383	13511	7.56447	Intron	intron 4 of 10	16020	*GABRA5*
296	chr16	82996891	17375	7.56447	Intron	intron 2 of 13	-6593	*LOC101928417*
297	chr21	41308911	33605	7.56447	Intron	intron 1 of 9	4699	*FAM3B*
298	chr9	134785630	49519	7.56447	Intron	intron 30 of 65	63739	*MIR3689C*
299	chr16	6118845	14740	7.56447	Intron	intron 4 of 6	486378	*MIR8065*
300	chr16	10629558	15058	7.56447	Intron	intron 2 of 2	-48860	*EMP2*
301	chr20	2159896	26079	7.56447	Intron	intron 2 of 3	-47321	*LOC388780*
302	chr21	10555367	28993	7.56447	Intron	intron 17 of 31	72629	*TPTE*

303	chr21	35852346	31917	7.56447	Intron	intron 6 of 12	131631	*MIR802*
304	chr21	38075990	32433	7.56447	Intron	intron 2 of 3	45370	*DSCR4*
305	chr22	48756698	36643	7.56447	Intron	intron 3 of 3	-23597	*MIR4535*
306	chr11	102452788	8351	7.56447	Promoter (<=1kb)		-131	*LOC101928424*
307	chr19	34482439	22390	7.56447	Promoter (<=1kb)		801	*WTIP*
308	chr3	15860269	37145	7.56447	Promoter (<=1kb)		-498	*ANKRD28*
309	chr5	34007785	40857	7.56447	Promoter (<=1kb)		323	*AMACR*
310	chr7	19117312	44314	7.56447	Promoter (<=1kb)		360	*TWIST1*
311	chr7	48088600	44751	7.56447	Promoter (<=1kb)		-28	*UPP1*
312	chr21	17437471	29621	7.56447	Promoter (1-2kb)		-1419	*LINC01549*
313	chr11	34359228	7373	7.56447	Promoter (1-2kb)		-1220	*ABTB2*
314	chr20	20369551	26672	7.56447	Promoter (1-2kb)		1430	*INSM1*
315	chr2	3655891	23597	7.56447	Promoter (2-3kb)		-2304	*ALLC*
316	chr17	21815371	18467	7.56217	Distal Intergenic		122848	*KCNJ18*
317	chr1	228620938	4488	7.56217	Distal Intergenic		7328	*RNA5S9*
318	chr20	30941899	27184	7.56217	Distal Intergenic		-315765	*DEFB115*
319	chr21	40138562	33151	7.56217	Intron	intron 18 of 32	73869	*MIR4760*
320	chr21	38564957	32620	7.56217	Intron	intron 2 of 8	96823	*ERG*
321	chr19	21321192	21467	7.56217	Intron	intron 1 of 15	8233	*ZNF708*
322	chr1	52553278	1761	7.56217	Promoter (<=1kb)		209	*ZCCHC11*
323	chr21	45277347	34279	7.53923	5' UTR		10551	*POFUT2*
324	chr21	42340884	33978	7.53923	Distal Intergenic		10244	*TFF2*
325	chr21	39183209	32851	7.52349	Promoter (<=1kb)		642	*PSMG1*
326	chr21	28825086	30773	7.5212	Distal Intergenic		60285	*N6AMT1*
327	chrX	24149880	50009	7.5212	Promoter (<=1kb)		-226	*ZFX-AS1*
328	chr21	40210650	33183	7.5212	Promoter (1-2kb)		1781	*MIR4760*
329	chr10	132786669	6917	7.51579	Promoter (<=1kb)		-617	*NKX6-2*
330	chr22	35633895	35869	7.46992	Intron	intron 2 of 3	4056	*MB*
331	chr1	37794105	1551	7.46802	Promoter (<=1kb)		303	*MANEAL*
332	chr16	67029845	16939	7.46802	Promoter (<=1kb)		729	*CBFB*
333	chr21	42021877	33876	7.45843	Intron	intron 1 of 1	-11490	*ZBTB21*
334	chr19	50511293	22903	7.45843	Promoter (<=1kb)		60	*JOSD2*
335	chr21	39611501	32942	7.42455	Promoter (1-2kb)		1320	*B3GALT5-AS1*
336	chr9	128162357	49372	7.42455	Promoter (2-3kb)		2097	*C9orf16*
337	chr4	49511470	39587	7.4099	Distal Intergenic		525223	*CWH43*
338	chr21	8996315	28889	7.38609	Distal Intergenic		787842	*MIR3648-1*
339	chr1	143213192	3055	7.38609	Distal Intergenic		486427	*RNVU1-17*
340	chr21	43605050	34113	7.38609	Intron	intron 7 of 8	4939	*MIR6070*
341	chr21	28751127	30762	7.37967	Distal Intergenic		134244	*N6AMT1*
342	chr22	17007668	34573	7.37967	Promoter (<=1kb)		554	*GAB4*
343	chr7	149460908	46116	7.37967	Promoter (<=1kb)		215	*ZNF777*
344	chr21	41766849	33795	7.37967	Promoter (<=1kb)		257	*RIPK4*
345	chr3	39313810	37470	7.36378	Distal Intergenic		-15896	*CCR8*
346	chr21	37801681	32331	7.36378	Intron	intron 2 of 3	114765	*KCNJ6*
347	chr21	10544627	28985	7.36378	Intron	intron 16 of 31	61889	*TPTE*
348	chrX	41333279	50195	7.36378	Promoter (<=1kb)		-69	*DDX3X*
349	chr5	100901162	41501	7.36378	Promoter (2-3kb)		2104	*ST8SIA4*
350	chr21	41492203	33688	7.33998	Intron	intron 3 of 13	38913	*TMPRSS2*
351	chr8	37701308	46699	7.31268	Distal Intergenic		5557	*ZNF703*
352	chr14	99185290	13064	7.31268	Intron	intron 3 of 3	86234	*BCL11B*
353	chr7	144559218	46038	7.31268	Intron	intron 9 of 11	-148991	*NOBOX*
354	chr5	114487	40128	7.31268	Intron	intron 2 of 2	-25771	*PLEKHG4B*
355	chr21	37844032	32347	7.29135	Intron	intron 1 of 3	72414	*KCNJ6*
356	chr21	37727391	32303	7.26225	Intron	intron 2 of 3	189055	*KCNJ6*
357	chr21	38929324	32752	7.26225	Intron	intron 1 of 3	27143	*LOC400867*
358	chr21	42613972	34019	7.26225	Promoter (1-2kb)		1086	*LINC01671*
359	chr7	157472440	46390	7.25872	Intron	intron 2 of 3	6209	*LOC101927914*
360	chr21	41597842	33728	7.24789	Distal Intergenic		-66726	*TMPRSS2*
361	chr21	7934593	28861	7.22467	Distal Intergenic		189631	*SMIM11B*
362	chr21	7949346	28868	7.22467	Distal Intergenic		204384	*SMIM11B*
363	chr1	111384757	2663	7.22467	Distal Intergenic		38469	*PIFO*
364	chr19	33820184	22363	7.22467	Distal Intergenic		24251	*KCTD15*
365	chr21	39092024	32818	7.22467	Distal Intergenic		91827	*PSMG1*
366	chr2	238126488	25909	7.22467	Intron	intron 6 of 8	-12234	*KLHL30*
367	chr21	38743596	32684	7.22467	Intron	intron 2 of 5	3864	*LINC00114*
368	chr21	44552973	34200	7.22467	Promoter (1-2kb)		-1468	*KRTAP10-2*
369	chr17	81396610	19914	7.22467	Promoter (1-2kb)		1135	*BAHCC1*
370	chr21	38757416	32691	7.21246	Distal Intergenic		-9956	*LINC00114*
371	chr6	17743176	42674	7.21246	Distal Intergenic		-36342	*NUP153*
372	chr10	71757603	5574	7.21246	Exon	exon 1 of 5	15895	*VSIR*
373	chr21	35763674	31894	7.21246	Intron	intron 6 of 12	42959	*MIR802*

374	chr7	135384377	45901	7.20488	3' UTR		125750	*CNOT4*
375	chr1	231368172	4525	7.20488	Intron	intron 3 of 4	-30320	*EXOC8*
376	chr21	41497083	33691	7.20488	Intron	intron 2 of 13	34033	*TMPRSS2*
377	chr21	44902769	34251	7.20033	Intron	intron 4 of 14	-18282	*ITGB2-AS1*
378	chr4	49513366	39588	7.16924	Distal Intergenic		527119	*CWH43*
379	chr21	39077939	32813	7.1633	Distal Intergenic		-100165	*LINC01700*
380	chr1	157420223	3351	7.1633	Distal Intergenic		132297	*FCRL5*
381	chr21	35965924	31949	7.1633	Intron	intron 5 of 12	38743	*LOC101928269*
382	chr11	67649864	7907	7.16232	Promoter (<=1kb)		795	*ACY3*
383	chr1	37233820	1529	7.12874	Distal Intergenic		72257	*MIR4255*
384	chr22	48068289	36469	7.12874	Distal Intergenic		-206075	*MIR3201*
385	chr22	48120878	36475	7.12874	Distal Intergenic		-153486	*MIR3201*
386	chr19	29029805	21895	7.12874	Distal Intergenic		27250	*LINC01532*
387	chr21	7291777	28850	7.12874	Distal Intergenic		-453185	*SMIM11B*
388	chr21	7929994	28858	7.12874	Distal Intergenic		185032	*SMIM11B*
389	chr19	15870520	21210	7.12874	Distal Intergenic		-5616	*LINC01835*
390	chr6	41470039	42975	7.12874	Distal Intergenic		-32405	*LINC01276*
391	chr7	136642402	45926	7.12874	Distal Intergenic		-226267	*CHRM2*
392	chr21	28707470	30758	7.12874	Distal Intergenic		168152	*LINC00161*
393	chr19	9086448	20914	7.12874	Distal Intergenic		-6731	*OR1M1*
394	chrX	62781372	50529	7.12874	Distal Intergenic		569972	*SPIN4*
395	chr12	36559899	9369	7.12874	Distal Intergenic		-1756679	*ALG10B*
396	chr16	26715482	15805	7.12874	Distal Intergenic		-351225	*C16orf82*
397	chr21	40912080	33454	7.12874	Distal Intergenic		-64941	*DSCAM*
398	chr20	12050619	26350	7.12874	Distal Intergenic		159896	*BTBD3*
399	chr9	60572237	48258	7.12874	Distal Intergenic		-342137	*SPATA31A7*
400	chr18	79825942	20669	7.12874	Distal Intergenic		-37726	*KCNG2*
401	chr21	30080947	30944	7.12874	Distal Intergenic		85809	*CLDN17*
402	chr1	104336737	2542	7.12874	Distal Intergenic		719926	*AMY2A*
403	chr10	43953820	5282	7.12874	Distal Intergenic		14817	*LINC00841*
404	chr2	30324957	23955	7.12874	Distal Intergenic		-21702	*LOC285043*
405	chr5	179092940	42217	7.12874	Distal Intergenic		32525	*ZNF354C*
406	chr18	7238031	20020	7.12874	Distal Intergenic		6906	*LRRC30*
407	chr21	24790101	30346	7.12874	Distal Intergenic		-50449	*LINC01692*
408	chr21	25120935	30387	7.12874	Distal Intergenic		280385	*LINC01692*
409	chr16	25340532	15725	7.12874	Distal Intergenic		-82601	*ZKSCAN2*
410	chr13	18992554	10596	7.12874	Distal Intergenic		-15705	*LINC00442*
411	chr10	77730793	5667	7.12874	Distal Intergenic		-92198	*KCNMA1*
412	chr16	73264060	17069	7.12874	Distal Intergenic		-119613	*C16orf47*
413	chr17	11174483	18061	7.12874	Distal Intergenic		-66780	*SHISA6*
414	chr21	14765718	29318	7.12874	Distal Intergenic		-106897	*LOC388813*
415	chr14	20177359	11554	7.12874	Distal Intergenic		-19977	*OR11G2*
416	chr9	29992978	48103	7.12874	Distal Intergenic		582034	*LINC01242*
417	chr21	16641131	29567	7.12874	Distal Intergenic		50894	*MIR125B2*
418	chr10	129016257	6695	7.12874	Distal Intergenic		-450927	*MGMT*
419	chr19	19834481	21355	7.12874	Distal Intergenic		-12730	*ZNF506*
420	chr19	23212153	21613	7.12874	Distal Intergenic		38237	*ZNF724*
421	chr3	84103437	38247	7.12874	Distal Intergenic		778242	*LINC00971*
422	chr21	9845238	28929	7.12874	Distal Intergenic		-31229	*LINC01667*
423	chr13	28585802	10827	7.12874	Distal Intergenic		-73302	*POMP*
424	chr17	15294987	18295	7.12874	Distal Intergenic		-29661	*PMP22*
425	chr17	52288713	19071	7.12874	Distal Intergenic		-128696	*CA10*
426	chr19	30878559	22125	7.12874	Distal Intergenic		470988	*TSHZ3*
427	chr10	99282100	6068	7.12874	Distal Intergenic		-46238	*HPSE2*
428	chr1	118750132	2938	7.12874	Distal Intergenic		239424	*TBX15*
429	chr21	10775621	29084	7.12874	Distal Intergenic		292883	*TPTE*
430	chr15	19991222	13298	7.12874	Distal Intergenic		550578	*GOLGA6L6*
431	chr15	20407055	13336	7.12874	Distal Intergenic		134745	*GOLGA6L6*
432	chr19	22860214	21573	7.12874	Distal Intergenic		-76107	*ZNF99*
433	chr22	10723107	34402	7.12874	Distal Intergenic		-4805051	*OR11H1*
434	chrX	50123077	50339	7.12874	Distal Intergenic		77936	*AKAP4*
435	chr22	34015800	35565	7.12874	Distal Intergenic		-92959	*LARGE1*
436	chr22	34270132	35627	7.12874	Distal Intergenic		-347291	*LARGE1*
437	chr2	128433327	25167	7.12874	Distal Intergenic		-114750	*HS6ST1*
438	chr19	24329216	21704	7.12874	Distal Intergenic		295811	*ZNF254*
439	chr20	40523552	27532	7.12874	Distal Intergenic		165688	*MAFB*
440	chr1	97470893	2458	7.12874	Intron	intron 13 of 22	-326028	*DPYD-AS2*
441	chr11	19862607	7219	7.12874	Intron	intron 3 of 37	-102524	*MIR4694*
442	chr11	101927067	8345	7.12874	Intron	intron 2 of 10	-10545	*ANGPTL5*
443	chr14	58005651	12205	7.12874	Intron	intron 6 of 7	-139589	*SLC35F4*
444	chr17	11258942	18069	7.12874	Intron	intron 1 of 5	17679	*SHISA6*

445	chr19	29397701	21946	7.12874	Intron	intron 4 of 9	128051	*LOC284395*
446	chr19	52449914	23041	7.12874	Intron	intron 4 of 4	-3662	*ZNF578*
447	chr21	21352411	30023	7.12874	Intron	intron 8 of 17	354096	*NCAM2*
448	chr21	24483446	30316	7.12874	Intron	intron 4 of 5	-162069	*LINC01689*
449	chr21	40181953	33171	7.12874	Intron	intron 14 of 32	30478	*MIR4760*
450	chr22	48698310	36619	7.12874	Intron	intron 2 of 3	-81985	*MIR4535*
451	chr4	5132990	39208	7.12874	Intron	intron 1 of 11	81548	*STK32B*
452	chr6	2837569	42382	7.12874	Intron	intron 4 of 6	4437	*SERPINB1*
453	chrX	114670179	51533	7.12874	Intron	intron 2 of 6	17554	*MIR1264*
454	chr9	69179720	48303	7.12874	Intron	intron 2 of 21	58456	*TJP2*
455	chr1	157862044	3382	7.12874	Intron	intron 1 of 3	36212	*CD5L*
456	chr10	78545063	5689	7.12874	Intron	intron 2 of 2	277753	*LINC00595*
457	chr11	134459893	8830	7.12874	Intron	intron 1 of 1	23420	*LOC283177*
458	chr14	102263207	13185	7.12874	Intron	intron 4 of 11	41993	*MOK*
459	chr16	6267410	14753	7.12874	Intron	intron 4 of 6	634943	*MIR8065*
460	chr16	18007051	15379	7.12874	Intron	intron 1 of 2	-536170	*XYLT1*
461	chr16	26311697	15786	7.12874	Intron	intron 1 of 1	286460	*MIR548W*
462	chr17	10316409	17993	7.12874	Intron	intron 1 of 1	56721	*MYH13*
463	chr17	35553711	18773	7.12874	Intron	intron 1 of 1	4387	*SLFN14*
464	chr19	23093613	21605	7.12874	Intron	intron 1 of 3	18403	*ZNF730*
465	chr19	27774647	21766	7.12874	Intron	intron 1 of 2	-18816	*LOC101927151*
466	chr20	52298482	28067	7.12874	Intron	intron 2 of 6	87837	*LINC01524*
467	chr21	24501126	30319	7.12874	Intron	intron 1 of 2	-179749	*LINC01689*
468	chr21	38493701	32600	7.12874	Intron	intron 1 of 8	-160280	*LINC01423*
469	chr3	42966459	37572	7.12874	Intron	intron 5 of 5	-12808	*FAM198A*
470	chr4	9850494	39377	7.12874	Intron	intron 1 of 3	68814	*DRD5*
471	chr7	51145750	44818	7.12874	Intron	intron 3 of 11	171068	*COBL*
472	chr9	17051783	47914	7.12874	Intron	intron 1 of 1	-83199	*CNTLN*
473	chr9	88639398	48627	7.12874	Intron	intron 2 of 3	12762	*LOC286238*
474	chr9	110759998	49027	7.12874	Intron	intron 7 of 14	91227	*MUSK*
475	chr19	52956362	23084	7.12874	Intron	intron 1 of 7	6406	*ZNF320*
476	chr2	111884656	24820	7.12874	Promoter (<=1kb)		34	*ANAPC1*
477	chr15	101295067	14514	7.12874	Promoter (<=1kb)		215	*SNRPA1*
478	chr17	27295847	18526	7.12874	Promoter (1-2kb)		1771	*WSB1*
479	chr21	30503311	31007	7.12874	Promoter (1-2kb)		-1194	*KRTAP19-5*
480	chrX	4630604	49703	7.12874	Promoter (2-3kb)		2968	*LOC101928201*
481	chr21	39045738	32800	7.1202	Distal Intergenic		-67964	*LINC01700*
482	chr21	39748448	32990	7.1202	Intron	intron 2 of 8	3041	*IGSF5*
483	chr20	8132853	26261	7.1202	Intron	intron 1 of 27	18400	*PLCB1*
484	chr6	83067938	43240	7.1202	Promoter (<=1kb)		272	*DOPEY1*
485	chr11	45665101	7557	7.1202	Promoter (<=1kb)		521	*CHST1*
486	chr12	1691256	8870	7.1202	Promoter (2-3kb)		2682	*ADIPOR2*
487	chr21	14113974	29194	7.11475	Intron	intron 2 of 2	96917	*LIPI*
488	chr20	3409346	26110	7.11475	Promoter (1-2kb)		-1721	*C20orf194*
489	chr21	38749316	32688	7.11475	Promoter (1-2kb)		-1856	*LINC00114*
490	chr20	31178789	27198	7.0979	Distal Intergenic		-78875	*DEFB115*
491	chr21	39952149	33086	7.0979	Distal Intergenic		84832	*PCP4*
492	chr16	5225741	14663	7.0979	Distal Intergenic		-14061	*RBFOX1*
493	chr2	119218034	24916	7.0979	Distal Intergenic		-5797	*STEAP3*
494	chr7	153055045	46203	7.0979	Distal Intergenic		295296	*ACTR3B*
495	chr20	24451725	26949	7.0979	Distal Intergenic		-17474	*SYNDIG1*
496	chr9	89329144	48642	7.0979	Exon	exon 3 of 3	10638	*SECISBP2*
497	chr20	43017770	27709	7.0979	Intron	intron 1 of 30	172200	*PTPRT*
498	chr21	40184489	33173	7.0979	Intron	intron 14 of 32	27942	*MIR4760*
499	chr17	27549494	18541	7.0979	Intron	intron 1 of 3	-73870	*LGALS9*
500	chr20	48921584	27978	7.0979	Promoter (<=1kb)		-306	*ARFGEF2*

Table S3.7. Sample D= Top 500 ChIP-Seq annotated peaks

Sl.No.	**Chr**	**Peak location**	**Peak ID**	**Score**	**Location annotation**	**Exon/Intron no.**	**Distance To TSS**	**Gene**
1	**chr5**	**157651788**	**47506**	**615.807**	**Exon**	**exon 1 of 4**	**19692**	***SOX30***
2	chr17	63917006	23815	139.4801	Promoter (1-2kb)		1832	*CSHL1*
3	chr1	143246990	3580	66.36189	Distal Intergenic		452629	*RNVU1-17*
4	chr1	143263474	3583	61.20729	Distal Intergenic		436145	*RNVU1-17*
5	chr21	8420433	34898	53.26097	Distal Intergenic		211960	*MIR3648-1*
6	chr1	125180211	3577	52.44954	Distal Intergenic		-3994672	*FAM72B*
7	chr16	34588195	20389	51.91314	Distal Intergenic		-428159	*LINC00273*
8	chr4	112587558	44479	28.04184	3' UTR		-49406	*LARP7*
9	chr17	26603915	22769	25.49575	Distal Intergenic		690081	*MIR4522*
10	chr4	112619066	44481	21.96072	3' UTR		-17898	*LARP7*

11	**chr6**	**28225300**	**48846**	**21.33079**	**Promoter (<=1kb)**		**414**	***ZSCAN9***
12	chr1	119508140	3532	21.0505	Promoter (<=1kb)		942	*HSD3B1*
13	chr1	93079166	2881	19.35048	Promoter (<=1kb)		-69	*MTF2*
14	chr8	42052108	54731	17.22757	Promoter (<=1kb)		-118	*KAT6A*
15	chr21	39348911	37040	15.94507	Promoter (<=1kb)		736	*HMGN1*
16	chr21	25607611	35723	14.60862	Promoter (<=1kb)		-94	*MRPL39*
17	chr13	76886322	14981	14.22222	Promoter (<=1kb)		78	*KCTD12*
18	chr1	151941415	3722	14.14528	Distal Intergenic		-31607	*THEM4*
19	chr3	23917340	40435	13.95163	Promoter (<=1kb)		795	*RPL15*
20	chr21	10722807	35018	13.57245	Distal Intergenic		240069	*TPTE*
21	chr21	15730340	35202	13.50362	Promoter (<=1kb)		315	*USP25*
22	chr21	14098344	35077	13.29831	Intron	intron 2 of 2	112547	*LIPI*
23	chr11	68213680	10127	13.19857	Promoter (<=1kb)		148	*KMT5B*
24	chr4	190108190	45040	13.04487	Distal Intergenic		43688	*DBET*
25	chr21	37006849	36685	12.96088	Promoter (<=1kb)		699	*RIPPLY3*
26	chr21	20998227	35496	12.63525	Promoter (<=1kb)		-88	*NCAM2*
27	chr17	26881386	22788	12.58126	Distal Intergenic		412610	*MIR4522*
28	chr12	109052553	13200	12.47831	Promoter (1-2kb)		1399	*USP30-AS1*
29	chr3	156817078	42935	12.46425	Promoter (<=1kb)		-16	*LINC00886*
30	**chr20**	**58651160**	**34562**	**12.46425**	**Promoter (<=1kb)**		**-93**	***STX16***
31	chr21	15064984	35141	12.35689	Promoter (<=1kb)		16	*NRIP1*
32	chr21	44873332	37821	12.19589	Promoter (<=1kb)		571	*PTTG1IP*
33	chr21	41585411	37479	12.19589	Distal Intergenic		-54295	*TMPRSS2*
34	chr5	167573601	47648	12.17025	Intron	intron 2 of 28	288802	*TENM2*
35	chr1	63323321	2274	12.13318	Promoter (<=1kb)		280	*FOXD3*
36	chr3	169663519	43019	11.97609	Promoter (<=1kb)		99	*MECOM*
37	chr21	8214724	34894	11.97609	Intron	intron 2 of 6	6251	*MIR3648-1*
38	chr22	10730223	37963	11.97609	Distal Intergenic		-4797935	*OR11H1*
39	chr22	38455796	39187	11.85689	Promoter (<=1kb)		-597	*KCNJ4*
40	chr21	10729219	35021	11.77936	Distal Intergenic		246481	*TPTE*
41	chr15	69452879	18070	11.63656	Promoter (<=1kb)		95	*RPLP1*
42	chr21	38749355	36930	11.60788	Promoter (1-2kb)		-1895	*LINC00114*
43	chr16	70251125	21154	11.42215	Promoter (<=1kb)		805	*EXOSC6*
44	chr21	17466860	35299	11.33154	Distal Intergenic		27970	*LINC01549*
45	chr21	41741339	37512	11.31627	Exon	exon 9 of 9	5502	*MIR6814*
46	chr10	41883800	6706	11.31627	Distal Intergenic		-591743	*LINC00839*
47	chr16	67873038	21074	11.29978	Promoter (<=1kb)		15	*EDC4*
48	chr16	30957837	20295	11.27004	Promoter (<=1kb)		543	*SETD1A*
49	chr21	10716212	35016	11.27004	Distal Intergenic		233474	*TPTE*
50	chr21	43659558	37714	11.25284	Promoter (<=1kb)		10	*RRP1B*
51	chr10	104329047	7922	11.08924	Intron	intron 2 of 2	5678	*LOC101927472*
52	chr11	120336221	11044	11.06827	Promoter (<=1kb)		-693	*ARHGEF12*
53	chr1	218285281	5232	10.98323	Promoter (<=1kb)		-6	*RRP15*
54	chr14	18661372	15442	10.96645	Distal Intergenic		-27738	*LINC02297*
55	chr2	207711865	31905	10.83689	Promoter (<=1kb)		325	*CCNYL1*
56	chr16	53131578	20704	10.83689	Intron	intron 1 of 38	76545	*CHD9*
57	chr2	234364497	32272	10.83689	Distal Intergenic		132556	*ARL4C*
58	chr20	62467	32509	10.83689	Distal Intergenic		-24783	*DEFB125*
59	chr10	77671224	7287	10.83689	Distal Intergenic		-32629	*KCNMA1*
60	chr1	234478577	5683	10.80342	Promoter (<=1kb)		526	*TARBP1*
61	chr1	99970037	3024	10.68852	Promoter (<=1kb)		248	*SLC35A3*
62	chr1	11478896	848	10.68852	Promoter (<=1kb)		-270	*DISP3*
63	chr15	98647627	18886	10.65561	Promoter (1-2kb)		-1344	*IGF1R*
64	chr21	37367981	36736	10.58337	Promoter (2-3kb)		2191	*DYRK1A*
65	chr9	132670640	58523	10.49888	Promoter (<=1kb)		-239	*DDX31*
66	chr21	31654294	36144	10.49888	Exon	exon 3 of 3	-5328	*SOD1*
67	chr21	34792449	36412	10.43371	3' UTR		46692	*LINC01426*
68	chr13	91348282	15237	10.4294	Promoter (<=1kb)		462	*MIR17HG*
69	chr16	21599586	19863	10.3397	Promoter (2-3kb)		2368	*METTL9*
70	chr2	174336254	31549	10.3397	Promoter (1-2kb)		1308	*SP9*
71	chr20	25067389	33306	10.3397	Distal Intergenic		-8409	*ACSS1*
72	chr13	40789503	14487	10.3225	Promoter (<=1kb)		91	*SLC25A15*
73	chr1	89633037	2780	10.31746	Promoter (<=1kb)		-35	*LRRC8C*
74	chr15	99565889	18918	10.27195	Promoter (<=1kb)		472	*MEF2A*
75	chr13	62021159	14737	10.14375	Intron	intron 1 of 1	8389	*LINC00358*
76	chr7	975539	51036	10.10287	Promoter (<=1kb)		10	*ADAP1*
77	chr21	31227154	36088	10.07419	Intron	intron 7 of 28	-188678	*KRTAP19-8*
78	chr21	10720104	35017	10.03897	Distal Intergenic		237366	*TPTE*
79	chr6	149863476	50466	9.99403	Promoter (<=1kb)		-18	*RAET1E-AS1*
80	chr10	118046637	8230	9.99403	Promoter (<=1kb)		-34	*RAB11FIP2*
81	chr6	79078432	49424	9.99403	Promoter (<=1kb)		-196	*PHIP*

82	chr6	17862890	48588	9.99403	Intron	intron 4 of 37	124733	*KIF13A*
83	chr11	14484615	9147	9.99403	Intron	intron 7 of 21	15412	*COPB1*
84	chr6	26593920	48812	9.99403	Distal Intergenic		-3032	*ABT1*
85	chr21	30834384	36045	9.99403	Distal Intergenic		-4625	*KRTAP7-1*
86	chr21	31088651	36068	9.99403	Distal Intergenic		-50175	*KRTAP19-8*
87	chr1	28737036	1324	9.95334	Promoter (<=1kb)		415	*YTHDF2*
88	chr1	52552968	1965	9.93451	Promoter (<=1kb)		519	*ZCCHC11*
89	chr22	11617078	38003	9.93451	Distal Intergenic		-3911080	*OR11H1*
90	chr21	37310290	36728	9.9151	Distal Intergenic		-42371	*DSCR3*
91	chr3	58334990	41360	9.91335	Promoter (2-3kb)		2110	*PXK*
92	chr8	37697765	54601	9.91335	Promoter (2-3kb)		2014	*ZNF703*
93	chr7	155457377	54119	9.91335	Promoter (<=1kb)		-752	*EN2*
94	chr10	22436997	6529	9.91335	Intron	intron 2 of 3	91552	*SPAG6*
95	chr6	15248706	48513	9.86546	Promoter (<=1kb)		-72	*JARID2-AS1*
96	chr10	133121768	8819	9.84493	Intron	intron 14 of 15	-35071	*ADGRA1-AS1*
97	chr15	79432183	18321	9.83739	Promoter (<=1kb)		-333	*KIAA1024*
98	chr17	49301573	23305	9.83739	Intron	intron 6 of 7	-13119	*MIR6129*
99	chr21	41150765	37388	9.77697	Promoter (2-3kb)		-2632	*LINC00323*
100	chr19	18280672	26432	9.77697	Promoter (<=1kb)		950	*JUND*
101	chr10	63521808	7025	9.76365	Promoter (<=1kb)		42	*JMJD1C*
102	chr1	115642069	3382	9.76365	Promoter (<=1kb)		116	*VANGL1*
103	chr1	231477926	5607	9.76365	Distal Intergenic		-50727	*TSNAX-DISC1*
104	chr21	30715984	36031	9.72735	Promoter (2-3kb)		2793	*KRTAP21-3*
105	chr4	98928887	44345	9.72735	Promoter (1-2kb)		1750	*EIF4E*
106	chr1	220046462	5262	9.72735	Promoter (<=1kb)		196	*EPRS*
107	chr15	95326862	18806	9.72735	Promoter (<=1kb)		267	*LINC01197*
108	chr5	98774086	46692	9.72735	Promoter (<=1kb)		-617	*RGMB-AS1*
109	chr4	118352462	44516	9.72735	Promoter (<=1kb)		541	*PRSS12*
110	chr5	172283735	47773	9.72735	Promoter (<=1kb)		336	*UBTD2*
111	chr21	9007831	34904	9.72735	Intron	intron 1 of 1	799358	*MIR3648-1*
112	chr6	107181262	49716	9.72735	Intron	intron 7 of 7	-65993	*BEND3*
113	chr21	40828781	37330	9.72735	Intron	intron 1 of 32	18358	*DSCAM*
114	chr12	31542279	12013	9.72735	Intron	intron 1 of 20	-47644	*DENND5B-AS1*
115	chr19	31733261	27033	9.72735	Distal Intergenic		145065	*THEG5*
116	chr5	166632365	47626	9.72735	Distal Intergenic		294005	*LINC01947*
117	chr2	126344270	30890	9.72735	Distal Intergenic		-311663	*GYPC*
118	chr10	46058922	6835	9.72735	Distal Intergenic		-12653	*MSMB*
119	chr9	69021148	56881	9.72735	Distal Intergenic		-7035	*PRKACG*
120	chr16	31073287	20303	9.68937	Promoter (1-2kb)		1033	*ZNF668*
121	chr1	143254234	3582	9.68937	Distal Intergenic		445385	*RNVU1-17*
122	chr21	10704684	35010	9.68937	Distal Intergenic		221946	*TPTE*
123	chr17	28903132	22854	9.66084	Promoter (<=1kb)		-61	*DHRS13*
124	chr17	19428155	22653	9.66084	Intron	intron 1 of 3	17030	*RNF112*
125	chr4	7131798	43483	9.66084	Distal Intergenic		-28418	*FLJ36777*
126	chr7	26864088	51486	9.61114	Intron	intron 1 of 12	131151	*SKAP2*
127	chr21	39313762	37036	9.60647	Promoter (<=1kb)		-173	*BRWD1-IT2*
128	chr15	20341532	17537	9.60647	Distal Intergenic		200268	*GOLGA6L6*
129	chr6	1611776	48157	9.60134	Promoter (1-2kb)		1804	*FOXC1*
130	chr21	26844709	35803	9.57817	Promoter (<=1kb)		700	*ADAMTS1*
131	chr1	3408742	534	9.57817	Intron	intron 8 of 15	-45684	*ARHGEF16*
132	chr8	142519702	56263	9.57817	Intron	intron 19 of 30	70272	*ADGRB1*
133	chr21	7950364	34889	9.57817	Distal Intergenic		205402	*SMIM11B*
134	chr17	82065876	24473	9.55711	Promoter (<=1kb)		11	*DUS1L*
135	chr21	42840617	37666	9.55711	Downstream (2-3kb)		38951	*WDR4*
136	chr13	62670557	14751	9.55711	Downstream (1-2kb)		61802	*LINC00448*
137	chr1	109984966	3189	9.51353	Promoter (<=1kb)		280	*AHCYL1*
138	chr7	140381885	53671	9.51353	Intron	intron 2 of 15	-22158	*RAB19*
139	chr1	231400939	5603	9.46576	Intron	intron 1 of 4	24105	*EGLN1*
140	chr16	34581995	20387	9.43561	Distal Intergenic		-421959	*LINC00273*
141	chr3	37911851	40742	9.41164	Intron	intron 1 of 6	49891	*CTDSPL*
142	chr20	57266055	34482	9.39462	Promoter (<=1kb)		574	*BMP7*
143	chr13	50046421	14646	9.31814	Promoter (2-3kb)		2640	*MIR16-1*
144	chr7	100015614	53062	9.31814	Promoter (<=1kb)		42	*ZKSCAN1*
145	chrX	41084724	59489	9.31814	Promoter (<=1kb)		-911	*USP9X*
146	chr14	31700628	15768	9.31814	Intron	intron 3 of 5	210672	*NUBPL*
147	chr5	137944641	47147	9.31814	Intron	intron 19 of 22	56673	*PKD2L2*
148	chr21	39749508	37113	9.31814	Intron	intron 2 of 8	4101	*IGSF5*
149	chr21	37642297	36763	9.31814	Intron	intron 3 of 3	274149	*KCNJ6*
150	chr4	49578713	43854	9.31814	Distal Intergenic		592466	*CWH43*
151	chr22	11619856	38004	9.31814	Distal Intergenic		-3908302	*OR11H1*
152	chr6	21595497	48698	9.25532	Promoter (2-3kb)		2728	*SOX4*

153	chr9	71910635	56934	9.25532	Promoter (<=1kb)		296	*ABHD17B*
154	chr3	112561740	42227	9.25532	Promoter (<=1kb)		31	*SLC35A5*
155	chr22	31944726	38678	9.25532	Promoter (<=1kb)		265	*YWHAH*
156	chr8	23290265	54479	9.25532	Exon	exon 5 of 9	20145	*R3HCC1*
157	chr19	4610958	25650	9.25247	Distal Intergenic		-28560	*TNFAIP8L1*
158	chr21	41606412	37482	9.20181	Distal Intergenic		-72769	*LINC00111*
159	chr14	90396992	17021	9.1871	Promoter (<=1kb)		490	*CALM1*
160	chr6	149718138	50458	9.1871	Promoter (<=1kb)		118	*LATS1*
161	chr1	233502941	5657	9.1871	Distal Intergenic		-111063	*KCNK1*
162	chr5	178726088	47960	9.17485	3' UTR		4614	*ZNF354A*
163	chr1	28553111	1317	9.17237	Promoter (<=1kb)		26	*TRNAU1AP*
164	chr21	33643106	36306	9.17237	Promoter (<=1kb)		706	*ITSN1*
165	chr9	37800811	56788	9.17237	Promoter (<=1kb)		309	*DCAF10*
166	chr17	80101428	24390	9.15994	Promoter (<=1kb)		-128	*GAA*
167	chr7	155458216	54120	9.14165	Promoter (<=1kb)		87	*EN2*
168	chrX	40168141	59458	9.09848	Intron	intron 1 of 13	9188	*BCOR*
169	chr16	68526793	21098	9.09848	Distal Intergenic		-3297	*ZFP90*
170	chr1	228487523	5508	9.05042	Promoter (<=1kb)		462	*RNF187*
171	chr7	140307918	53669	9.02338	Intron	intron 13 of 15	-96125	*RAB19*
172	chr19	5456075	25681	8.96525	Promoter (<=1kb)		644	*ZNRF4*
173	chr10	101694792	7859	8.96525	Promoter (<=1kb)		503	*FBXW4*
174	chr21	41456811	37447	8.96525	Intron	intron 18 of 18	36507	*MX1*
175	chr21	26246547	35757	8.96525	Distal Intergenic		-75419	*APP*
176	chr13	99969418	15376	8.94938	Promoter (2-3kb)		2491	*ZIC5*
177	chr9	111483389	57936	8.94938	Promoter (1-2kb)		1356	*KIAA0368*
178	chr5	56815082	46119	8.94938	Promoter (<=1kb)		-492	*MAP3K1*
179	chr3	5122173	40019	8.94938	Promoter (<=1kb)		-47	*ARL8B*
180	chr21	39797408	37129	8.94938	Intron	intron 8 of 8	52001	*IGSF5*
181	chr14	91051118	17037	8.94938	Intron	intron 1 of 16	-9215	*C14orf159*
182	chr22	34832985	38980	8.94938	Distal Intergenic		-233151	*ISX*
183	chr1	160084910	4042	8.94769	Intron	intron 2 of 2	3340	*KCNJ9*
184	chr12	27333531	11947	8.92328	Promoter (<=1kb)		677	*ARNTL2*
185	chr15	89020107	18606	8.87657	Distal Intergenic		-67352	*ABHD2*
186	chr10	75399355	7247	8.86322	Promoter (2-3kb)		-2164	*ZNF503-AS2*
187	chr13	40771199	14486	8.86322	Promoter (<=1kb)		-26	*MRPS31*
188	chr1	89524934	2779	8.86322	Promoter (<=1kb)		98	*LRRC8B*
189	chr21	31559084	36127	8.86322	Promoter (<=1kb)		893	*TIAM1*
190	chr1	163321811	4153	8.86322	Promoter (<=1kb)		83	*RGS5*
191	chr1	62688173	2257	8.86322	Promoter (<=1kb)		195	*DOCK7*
192	chr12	113300034	13350	8.86322	Intron	intron 16 of 16	8511	*MIR6762*
193	chr10	41910621	6711	8.86322	Distal Intergenic		-564922	*LINC00839*
194	chr1	58782617	2149	8.79059	Promoter (1-2kb)		1710	*JUN*
195	chr21	44386272	37764	8.73432	Intron	intron 10 of 32	36109	*TRPM2*
196	chr2	197516012	31756	8.73429	Promoter (<=1kb)		441	*MOB4*
197	chr12	121467031	13628	8.73429	Intron	intron 12 of 22	-22679	*MIR7107*
198	chr7	877104	51028	8.73429	Intron	intron 9 of 16	25154	*GET4*
199	chr16	30875336	20288	8.69914	Promoter (<=1kb)		-13	*MIR4519*
200	chr17	50866010	23362	8.69914	Promoter (<=1kb)		-669	*TOB1-AS1*
201	chr15	101489067	18972	8.69914	Intron	intron 1 of 21	36135	*PCSK6*
202	chr22	32867732	38765	8.66475	Intron	intron 5 of 13	66031	*TIMP3*
203	chr9	134599234	58585	8.66475	Distal Intergenic		-42540	*COL5A1*
204	chr13	85720997	15155	8.66475	Distal Intergenic		78491	*SLITRK6*
205	chr21	10710192	35013	8.65712	Distal Intergenic		227454	*TPTE*
206	chr5	60524247	46188	8.64961	Promoter (2-3kb)		-2127	*PDE4D*
207	chr14	45250852	16011	8.64961	Promoter (2-3kb)		2688	*MIS18BP1*
208	chr22	48545920	39690	8.64961	Promoter (1-2kb)		1467	*LOC284933*
209	chr21	39612473	37088	8.64961	Promoter (<=1kb)		348	*B3GALT5-AS1*
210	chr5	113488303	46832	8.64961	Promoter (<=1kb)		527	*MCC*
211	chr5	68216416	46315	8.64961	Promoter (<=1kb)		696	*PIK3R1*
212	chr14	91510595	17051	8.64961	Promoter (<=1kb)		-41	*PPP4R3A*
213	chr10	25174934	6573	8.64961	Promoter (<=1kb)		-128	*GPR158*
214	chr3	177197918	43097	8.64961	Promoter (<=1kb)		-522	*TBL1XR1*
215	chr18	500744	24518	8.64961	Promoter (<=1kb)		-22	*COLEC12*
216	chr5	127792717	46994	8.64961	Intron	intron 3 of 6	89327	*CCDC192*
217	chr21	35513072	36492	8.64961	Intron	intron 8 of 12	67688	*LOC100506403*
218	chr19	56034325	28316	8.64961	Intron	intron 8 of 13	-32359	*LINC01864*
219	chr18	37080582	25076	8.64961	Intron	intron 6 of 10	-156768	*LOC105372069*
220	chr16	7457179	19286	8.64961	Intron	intron 4 of 11	1133014	*TMEM114*
221	chr2	235669523	32309	8.64961	Intron	intron 1 of 9	163772	*AGAP1-IT1*
222	chr17	42905168	23145	8.64961	Intron	intron 2 of 4	4371	*G6PC*
223	chr4	86576602	44223	8.64961	Intron	intron 1 of 6	17529	*MAPK10*

224	chr10	73181132	7204	8.64961	Intron	intron 3 of 13	-12077	*ECD*
225	chr9	124634685	58281	8.64961	Intron	intron 2 of 9	-23782	*MIR181A2HG*
226	chr4	86797333	44228	8.64961	Intron	intron 40 of 46	51930	*SLC10A6*
227	chr3	149535794	42846	8.64961	Intron	intron 5 of 7	61820	*TM4SF4*
228	chr22	47656190	39584	8.64961	Intron	intron 2 of 8	24516	*LOC284930*
229	chr22	44986291	39400	8.64961	Intron	intron 2 of 13	23708	*PHF21B*
230	chr21	40344052	37226	8.64961	Intron	intron 6 of 32	-39031	*DSCAM-AS1*
231	chr21	15812594	35206	8.64961	Intron	intron 9 of 23	82569	*USP25*
232	chr16	82997014	21604	8.64961	Intron	intron 2 of 13	-6716	*LOC101928417*
233	chr16	27991652	20199	8.64961	Intron	intron 1 of 6	71857	*GSG1L*
234	chr4	143558293	44726	8.64961	Distal Intergenic		-43658	*SMARCA5-AS1*
235	chr5	1684134	45082	8.64961	Distal Intergenic		24734	*MIR4277*
236	chr5	3031543	45164	8.64961	Distal Intergenic		-146292	*LINC01377*
237	chr11	119771648	11031	8.64961	Distal Intergenic		42065	*LOC102724301*
238	chr10	42938517	6740	8.64961	Distal Intergenic		42990	*LINC01264*
239	chrX	81000306	60440	8.64961	Distal Intergenic		-190618	*BRWD3*
240	chr18	21126965	24681	8.64961	Distal Intergenic		-15114	*ROCK1*
241	chr21	41540027	37469	8.64961	Distal Intergenic		-8911	*TMPRSS2*
242	chr21	10617330	34975	8.64961	Distal Intergenic		134592	*TPTE*
243	chr17	56770607	23595	8.64961	Distal Intergenic		-63492	*DGKE*
244	chr22	34364183	38932	8.64961	Distal Intergenic		-441342	*LARGE1*
245	chr15	66293820	17994	8.6123	Promoter (<=1kb)		603	*DIS3L*
246	chr11	62727203	9948	8.6123	Promoter (<=1kb)		146	*HNRNPUL2*
247	chr16	4949472	19189	8.6123	Intron	intron 1 of 9	-8845	*SEC14L5*
248	chr22	20839651	38245	8.6123	Intron	intron 1 of 54	-19332	*SNAP29*
249	chr21	31501743	36118	8.6123	Intron	intron 1 of 28	58234	*TIAM1*
250	chr21	41459470	37448	8.6123	Distal Intergenic		39166	*MX1*
251	chr21	10666213	34991	8.6123	Distal Intergenic		183475	*TPTE*
252	chr22	20411103	38228	8.6123	Distal Intergenic		16988	*ZNF74*
253	chr1	185317285	4710	8.59624	Promoter (<=1kb)		44	*IVNS1ABP*
254	chr15	72677331	18150	8.59624	Intron	intron 2 of 2	8818	*HIGD2B*
255	chr7	158518884	54273	8.59624	Intron	intron 1 of 21	13929	*MIR595*
256	chr3	19939012	40354	8.59624	Intron	intron 1 of 14	7997	*EFHB*
257	chr19	57975141	28473	8.58861	Promoter (<=1kb)		-607	*C19orf18*
258	chr19	35269534	27260	8.58861	Promoter (<=1kb)		556	*USF2*
259	chr5	146380648	47327	8.58861	Distal Intergenic		41624	*POU4F3*
260	chr13	92878188	15261	8.52873	Distal Intergenic		-156574	*GPC5-AS1*
261	chr17	16000063	22546	8.5111	Promoter (<=1kb)		-346	*ZSWIM7*
262	chr20	25390645	33332	8.5111	Promoter (<=1kb)		338	*ABHD12*
263	chr19	43154509	27646	8.5111	Intron	intron 2 of 2	32027	*PSG5*
264	chr17	66318742	23895	8.5111	Intron	intron 2 of 9	16106	*PRKCA*
265	chr16	4720672	19181	8.5111	Intron	intron 4 of 15	-13600	*C16orf71*
266	chr16	1972009	19086	8.49902	Promoter (<=1kb)		-28	*TBL3*
267	chr2	86440942	30101	8.4622	Promoter (<=1kb)		295	*KDM3A*
268	chr7	129502590	53422	8.43495	Promoter (<=1kb)		111	*SMKR1*
269	chr12	110742904	13251	8.43495	Promoter (<=1kb)		35	*PPP1CC*
270	chr19	30222666	26944	8.43495	Intron	intron 1 of 1	-5848	*ZNF536*
271	chr1	114733378	3353	8.43495	Intron	intron 8 of 18	-16484	*NRAS*
272	chr6	1313661	48144	8.41774	Promoter (1-2kb)		1188	*FOXQ1*
273	chr7	26202240	51478	8.41774	Promoter (<=1kb)		-711	*HNRNPA2B1*
274	chr13	41198549	14493	8.39662	Intron	intron 3 of 3	-3983	*KBTBD7*
275	chr6	2971085	48199	8.36011	Promoter (1-2kb)		1080	*SERPINB6*
276	chr5	96807824	46669	8.36011	Promoter (<=1kb)		276	*ERAP1*
277	chr21	44598166	37793	8.33197	Promoter (2-3kb)		-2431	*KRTAP10-7*
278	chr10	114044170	8150	8.33197	Promoter (<=1kb)		114	*ADRB1*
279	chr1	180502673	4594	8.33197	Promoter (<=1kb)		281	*ACBD6*
280	chr17	76487356	24265	8.33197	Intron	intron 2 of 18	14434	*RHBDF2*
281	chr17	49797200	23325	8.33197	Intron	intron 3 of 14	-8020	*FAM117A*
282	chr21	38660559	36915	8.32612	Promoter (1-2kb)		1221	*ERG*
283	chr9	97412009	57632	8.32612	Promoter (<=1kb)		59	*TDRD7*
284	chr20	31968604	33478	8.32612	Promoter (<=1kb)		602	*XKR7*
285	chr1	86704595	2724	8.32612	Promoter (<=1kb)		25	*SH3GLB1*
286	chr9	83921087	57195	8.32612	Promoter (<=1kb)		378	*KIF27*
287	chr20	50795000	34186	8.32612	Promoter (<=1kb)		106	*BCAS4*
288	chr16	81314965	21528	8.32612	Promoter (<=1kb)		13	*GAN*
289	chr21	10559715	34957	8.32612	Intron	intron 18 of 31	76977	*TPTE*
290	chr10	6461026	6170	8.32612	Intron	intron 13 of 16	119275	*PRKCQ*
291	chr21	38611366	36905	8.32612	Intron	intron 1 of 11	50414	*ERG*
292	chr3	52625817	41178	8.32612	Intron	intron 12 of 27	-55339	*GNL3*
293	chr21	26038541	35742	8.32612	Intron	intron 5 of 17	132587	*APP*
294	chr20	2957112	32599	8.32612	Intron	intron 7 of 27	-86510	*GNRH2*

295	chr1	176764710	4484	8.32612	Intron	intron 14 of 22	264742	*MIR488*
296	chr8	47222181	54828	8.32612	Distal Intergenic		31409	*LOC100287846*
297	chr7	67409780	52425	8.32612	Distal Intergenic		73804	*LINC01372*
298	chr3	95267484	41973	8.32612	Distal Intergenic		329312	*LINC00879*
299	chr6	159598216	50689	8.32612	Distal Intergenic		-127369	*WTAP*
300	chr1	230534712	5575	8.32009	Distal Intergenic		-107777	*COG2*
301	chr21	38957841	36973	8.31503	Promoter (1-2kb)		-1374	*LOC400867*
302	chr3	42600817	40885	8.3113	Promoter (<=1kb)		203	*NKTR*
303	chr22	34103113	38909	8.28654	Intron	intron 3 of 7	-180272	*LARGE1*
304	chrX	154706197	62028	8.28654	Intron	intron 3 of 8	45386	*GAB3*
305	chr3	142785613	42787	8.28654	Intron	intron 7 of 12	61539	*TRPC1*
306	chrX	130327874	61559	8.28654	Distal Intergenic		-12014	*SLC25A14*
307	chr3	123067075	42365	8.26822	Promoter (<=1kb)		13	*PDIA5*
308	chr5	6447944	45292	8.26822	Promoter (<=1kb)		-679	*UBE2QL1*
309	chr21	44666611	37801	8.26822	Promoter (<=1kb)		316	*KRTAP12-2*
310	chr3	128680553	42489	8.26822	Promoter (<=1kb)		522	*RPN1*
311	chr22	17053320	38054	8.26822	Intron	intron 3 of 3	16750	*CECR7*
312	chr21	40530946	37267	8.26822	Intron	intron 3 of 32	99821	*DSCAM-IT1*
313	chr21	23833895	35631	8.26822	Distal Intergenic		-449034	*D21S2088E*
314	chrX	71013937	60110	8.26822	Distal Intergenic		59489	*SNX12*
315	chr15	85160522	18497	8.26822	Distal Intergenic		180082	*PDE8A*
316	chr9	133013348	58533	8.26822	Distal Intergenic		-17327	*GTF3C5*
317	chr19	58519802	28532	8.26822	5' UTR		-7737	*SLC27A5*
318	chr1	16616592	1010	8.25726	Distal Intergenic		-3030	*NBPF1*
319	chr5	93582095	46632	8.2142	Promoter (1-2kb)		-1242	*NR2F1*
320	chr5	10601556	45404	8.2142	Intron	intron 1 of 4	37226	*ANKRD33B*
321	chr6	43106966	49080	8.2142	Intron	intron 1 of 19	30698	*PTK7*
322	chr6	22124324	48716	8.2142	Intron	intron 9 of 11	22869	*NBAT1*
323	chr5	169730098	47706	8.2142	Intron	intron 22 of 51	92851	*DOCK2*
324	chr18	21111545	24679	8.20536	Promoter (<=1kb)		306	*ROCK1*
325	chr21	40139788	37194	8.20536	Intron	intron 18 of 32	72643	*MIR4760*
326	chr22	17171757	38068	8.15297	Distal Intergenic		-6470	*HDHD5*
327	chr19	30133823	26934	8.14335	Distal Intergenic		-94691	*ZNF536*
328	chr21	42839552	37665	8.11576	Distal Intergenic		40016	*WDR4*
329	chr20	17975402	33004	8.08219	Intron	intron 2 of 4	-6422	*SNX5*
330	chr13	43879659	14542	8.07308	Promoter (<=1kb)		68	*CCDC122*
331	chr12	112418849	13317	8.07308	Promoter (<=1kb)		-11	*RPL6*
332	chr8	119892091	55842	8.07308	Intron	intron 1 of 6	18374	*DEPTOR*
333	chr5	1271136	45076	8.07308	Exon	exon 8 of 16	23933	*TERT*
334	chr3	51538867	41150	8.04474	Promoter (2-3kb)		-2713	*RAD54L2*
335	chr1	117366983	3452	8.04474	Promoter (<=1kb)		-466	*MAN1A2*
336	chr1	32651007	1468	8.04474	Promoter (<=1kb)		-104	*ZBTB8OS*
337	chr2	33710769	29245	8.04474	Intron	intron 1 of 5	3883	*LINC01317*
338	chr1	61035695	2196	8.04474	Intron	intron 1 of 10	-64919	*NFIA-AS2*
339	chr22	36275886	39087	8.04474	Distal Intergenic		11021	*MIR6819*
340	chr22	35565210	39069	8.04474	Distal Intergenic		24342	*RASD2*
341	chr1	238415629	5795	8.04474	Distal Intergenic		70394	*LINC01139*
342	chr21	10696586	35006	8.04474	Distal Intergenic		213848	*TPTE*
343	chr7	5421053	51224	8.02184	Intron	intron 1 of 1	4361	*TNRC18*
344	chr12	118136210	13533	8.00955	Promoter (<=1kb)		-262	*VSIG10*
345	chr19	1668270	25499	8.00423	Distal Intergenic		-15665	*TCF3*
346	chr16	69565906	21143	7.98868	Promoter (<=1kb)		-38	*MIR1538*
347	chr15	84716829	18478	7.98868	Promoter (<=1kb)		-113	*SEC11A*
348	chr14	31207715	15758	7.98868	Promoter (<=1kb)		89	*HECTD1*
349	chr17	76736752	24277	7.98868	Promoter (<=1kb)		-204	*MIR636*
350	chr21	36966334	36674	7.98868	Intron	intron 1 of 11	23902	*HLCS*
351	chr21	43629433	37707	7.98868	Intron	intron 6 of 8	-19444	*MIR6070*
352	chr10	21516639	6501	7.98868	Exon	exon 4 of 4	-8036	*MLLT10*
353	chr13	80241654	15047	7.98868	Distal Intergenic		99297	*SPRY2*
354	chr1	236956607	5751	7.98868	Distal Intergenic		47811	*MT1HL1*
355	chr3	54382538	41224	7.96229	Intron	intron 4 of 38	257319	*ESRG*
356	chr19	6717807	25748	7.96229	Intron	intron 4 of 40	12755	*C3*
357	chr7	156498743	54166	7.93894	Intron	intron 1 of 1	-41748	*LINC00244*
358	chr5	140158966	47204	7.93647	Promoter (1-2kb)		1647	*LOC101929719*
359	chr12	42589788	12168	7.93647	Promoter (<=1kb)		567	*PRICKLE1*
360	chr16	87491927	21755	7.93647	Promoter (<=1kb)		118	*ZCCHC14*
361	chr6	149318112	50452	7.93647	Intron	intron 1 of 1	-82115	*SUMO4*
362	chr21	38907310	36957	7.93647	Intron	intron 2 of 3	49157	*LOC400867*
363	chr4	140276383	44700	7.93647	Intron	intron 1 of 4	19097	*SCOC*
364	chr22	35974611	39075	7.93647	Intron	intron 1 of 13	191566	*APOL3*
365	chr2	202224014	31815	7.93647	Intron	intron 1 of 4	14594	*SUMO1*

366	chr11	57603529	9803	7.93647	Intron	intron 4 of 7	6142	*SERPING1*
367	chr2	130372337	31018	7.93647	Exon	exon 13 of 15	16330	*PTPN18*
368	chr3	150731795	42875	7.93647	Distal Intergenic		-27824	*ERICH6*
369	chr10	130002731	8650	7.93647	Distal Intergenic		-38890	*EBF3*
370	chr21	9759190	34916	7.93647	Distal Intergenic		54819	*LINC01667*
371	chr1	228310499	5495	7.93213	Intron	intron 49 of 80	86610	*MIR6742*
372	chr4	49511652	43849	7.93213	Distal Intergenic		525405	*CWH43*
373	chr22	19432492	38185	7.93188	Promoter (<=1kb)		590	*MRPL40*
374	chr21	34886841	36424	7.93188	Intron	intron 1 of 5	141084	*LINC01426*
375	chr21	39445741	37055	7.92424	Promoter (<=1kb)		64	*LCA5L*
376	chr17	81871488	24461	7.92424	Promoter (<=1kb)		-82	*ARHGDIA*
377	chr21	42893533	37670	7.92424	Intron	intron 3 of 5	13889	*NDUFV3*
378	chr1	225706465	5395	7.92424	Intron	intron 1 of 2	-53323	*ENAH*
379	chr7	150653328	53905	7.92424	Distal Intergenic		-20680	*GIMAP6*
380	chr13	94737348	15294	7.90352	Distal Intergenic		24632	*SOX21-AS1*
381	chr21	39015663	36982	7.90352	Distal Intergenic		-37889	*LINC01700*
382	chr7	151642248	53954	7.86992	Intron	intron 4 of 15	-122128	*RHEB*
383	chr13	65037639	14785	7.86992	Distal Intergenic		-828535	*LINC01052*
384	chr1	175279870	4450	7.86977	Distal Intergenic		-86871	*KIAA0040*
385	chr22	50337805	39917	7.86977	Distal Intergenic		-5499	*PPP6R2*
386	chr21	37073276	36699	7.86093	Promoter (<=1kb)		50	*TTC3*
387	chr2	231710289	32215	7.86093	Exon	exon 1 of 5	3394	*PTMA*
388	chr6	3068624	48201	7.86093	5' UTR		4633	*RIPK1*
389	chrX	119872396	61296	7.84552	Promoter (<=1kb)		-569	*RNF113A*
390	chr1	110674653	3205	7.80411	Promoter (<=1kb)		380	*KCNA3*
391	chr17	6069832	22065	7.80411	Intron	intron 4 of 4	296906	*WSCD1*
392	chr21	42075113	37561	7.80411	Intron	intron 1 of 21	12154	*UMODL1*
393	chr19	46860638	27793	7.80411	Distal Intergenic		-9646	*AP2S1*
394	chr21	39081464	37002	7.78826	Distal Intergenic		102387	*PSMG1*
395	chr6	145735443	50391	7.73945	Promoter (<=1kb)		581	*EPM2A*
396	chr1	204918743	4970	7.73945	Intron	intron 1 of 25	90092	*NFASC*
397	chr21	41806934	37522	7.73425	Intron	intron 1 of 6	-39828	*RIPK4*
398	chr7	44796285	51861	7.73172	Promoter (<=1kb)		-395	*PPIA*
399	chr22	31519591	38657	7.73172	Intron	intron 3 of 3	-23483	*EIF4ENIF1*
400	chr21	41933783	37542	7.73172	Intron	intron 2 of 13	20107	*C2CD2*
401	chr19	5799464	25698	7.73172	Distal Intergenic		-8226	*DUS3L*
402	chr18	28175840	24884	7.6759	Promoter (1-2kb)		1606	*CDH2*
403	chr3	111071908	42197	7.6759	Promoter (<=1kb)		-355	*PVRL3-AS1*
404	chr22	21736099	38264	7.6759	Promoter (<=1kb)		-265	*YPEL1*
405	chr7	55365809	52130	7.6759	Promoter (<=1kb)		361	*LANCL2*
406	chr2	74958739	29931	7.6759	Promoter (<=1kb)		247	*POLE4*
407	chr13	39655875	14457	7.6759	Promoter (<=1kb)		248	*COG6*
408	chr3	24495385	40447	7.6759	Promoter (<=1kb)		-103	*THRB*
409	chr2	3703558	28630	7.6759	Promoter (<=1kb)		-34	*DCDC2C*
410	chr21	34982740	36436	7.6759	Intron	intron 2 of 8	66604	*RUNX1*
411	chr22	49595882	39831	7.6759	Intron	intron 1 of 4	-52416	*MIR3667*
412	chr1	71612119	2468	7.6759	Intron	intron 4 of 6	224893	*NEGR1-IT1*
413	chr21	41712351	37507	7.6759	Exon	exon 6 of 6	3424	*LINC00479*
414	chr4	112560842	44477	7.6759	5' UTR		-44670	*NEUROG2*
415	chr5	168291861	47665	7.67577	Promoter (<=1kb)		210	*WWC1*
416	chr1	143213129	3579	7.67577	Distal Intergenic		486490	*RNVU1-17*
417	chrX	96670769	60693	7.67577	Distal Intergenic		-13894	*DIAPH2*
418	chr9	113633618	58009	7.67577	Distal Intergenic		188887	*RGS3*
419	chr7	141551434	53719	7.63519	5' UTR		-212663	*TAS2R3*
420	chr22	36507037	39099	7.62078	Promoter (<=1kb)		64	*FOXRED2*
421	chr6	36409957	48922	7.62078	Intron	intron 3 of 4	-22157	*ETV7*
422	chr5	172492817	47781	7.62078	Distal Intergenic		-38294	*SH3PXD2B*
423	chr21	40405858	37237	7.61304	Intron	intron 3 of 32	22775	*DSCAM-AS1*
424	chr20	57709101	34504	7.60615	Promoter (1-2kb)		-1082	*NKILA*
425	chr10	44384628	6796	7.60615	Promoter (1-2kb)		1865	*CXCL12*
426	chr11	72436304	10303	7.60615	Promoter (1-2kb)		-1624	*CLPB*
427	chr2	134919207	31104	7.60615	Promoter (<=1kb)		-497	*CCNT2-AS1*
428	chr21	30717966	36032	7.60615	Promoter (<=1kb)		811	*KRTAP21-3*
429	chr12	103930465	13077	7.60615	Promoter (<=1kb)		40	*MIR3652*
430	chr20	58889158	34575	7.60615	Promoter (<=1kb)		-349	*LOC101927932*
431	chr10	114994767	8172	7.60615	Promoter (<=1kb)		110	*LOC102724589*
432	chr12	76559583	12735	7.60615	Promoter (<=1kb)		226	*OSBPL8*
433	chr6	88166290	49530	7.60615	Promoter (<=1kb)		69	*CNR1*
434	chr2	131093332	31038	7.60615	Promoter (<=1kb)		128	*FAM168B*
435	chr4	134201674	44643	7.60615	Promoter (<=1kb)		74	*PABPC4L*
436	chrX	55089301	59875	7.60615	Promoter (<=1kb)		293	*PAGE2*

437	chr13	51222565	14665	7.60615	Promoter (<=1kb)		231	*FAM124A*
438	chr18	70205757	25339	7.60615	Promoter (<=1kb)		188	*RTTN*
439	chr18	11689831	24620	7.60615	Promoter (<=1kb)		875	*GNAL*
440	chr13	49997282	14644	7.60615	Promoter (<=1kb)		469	*KCNRG*
441	chr20	25081544	33307	7.60615	Promoter (<=1kb)		821	*VSX1*
442	chr10	21527255	6503	7.60615	Promoter (<=1kb)		-887	*SKIDA1*
443	chr14	89417487	16983	7.60615	Promoter (<=1kb)		133	*FOXN3-AS1*
444	chr2	209423847	31930	7.60615	Promoter (<=1kb)		-211	*MAP2*
445	chr1	220528420	5275	7.60615	Promoter (<=1kb)		237	*MARK1*
446	chr4	40056914	43784	7.60615	Promoter (<=1kb)		88	*N4BP2*
447	chr2	108719534	30523	7.60615	Promoter (<=1kb)		53	*RANBP2*
448	chr15	56918642	17901	7.60615	Promoter (<=1kb)		19	*TCF12*
449	chrX	44873281	59590	7.60615	Promoter (<=1kb)		104	*KDM6A*
450	chr19	7395325	25778	7.60615	Intron	intron 2 of 21	12491	*ARHGEF18*
451	chrX	37814447	59404	7.60615	Intron	intron 3 of 8	33190	*DYNLT3*
452	chr9	76669499	57027	7.60615	Intron	intron 1 of 10	-94937	*PCA3*
453	chr7	56046585	52173	7.60615	Intron	intron 1 of 6	5019	*PSPH*
454	chr19	36396470	27337	7.60615	Intron	intron 1 of 1	-17271	*ZFP14*
455	chr19	32421436	27070	7.60615	Intron	intron 3 of 3	-15876	*LOC400684*
456	chr19	20227275	26523	7.60615	Intron	intron 3 of 4	60047	*ZNF486*
457	chr17	51700529	23417	7.60615	Intron	intron 4 of 8	363814	*LINC02072*
458	chr16	71479953	21190	7.60615	Intron	intron 3 of 11	9358	*ZNF23*
459	chr15	76225546	18245	7.60615	Intron	intron 10 of 10	-44060	*LOC101929439*
460	chr15	69021446	18054	7.60615	Intron	intron 2 of 3	-51480	*EWSAT1*
461	chr14	34345838	15824	7.60615	Intron	intron 1 of 4	116518	*SPTSSA*
462	chr13	42245833	14512	7.60615	Intron	intron 29 of 30	-26320	*AKAP11*
463	chr12	78370907	12756	7.60615	Intron	intron 1 of 2	-11161	*LINC02424*
464	chr12	28410783	11965	7.60615	Intron	intron 3 of 5	277534	*CCDC91*
465	chr12	5481822	11586	7.60615	Intron	intron 1 of 4	49710	*NTF3*
466	chr10	77553034	7281	7.60615	Intron	intron 1 of 28	85561	*KCNMA1*
467	chr10	43467327	6764	7.60615	Intron	intron 1 of 3	30486	*ZNF487*
468	chr7	100745177	53089	7.60615	Intron	intron 7 of 47	11551	*ZAN*
469	chr10	59306688	6956	7.60615	Intron	intron 4 of 13	56493	*FAM13C*
470	chr5	167438932	47646	7.60615	Intron	intron 2 of 28	154133	*TENM2*
471	chr3	62426333	41464	7.60615	Intron	intron 26 of 27	-52009	*FEZF2*
472	chr21	38572138	36895	7.60615	Intron	intron 2 of 8	89642	*ERG*
473	chr1	45442916	1807	7.60615	Intron	intron 1 of 8	48250	*TESK2*
474	chrX	53987577	59827	7.60615	Intron	intron 15 of 21	61381	*PHF8*
475	chr9	33890647	56686	7.60615	Intron	intron 2 of 4	43729	*SNORD121B*
476	chr7	218320	50966	7.60615	Intron	intron 3 of 9	-161039	*LOC442497*
477	chr6	136264342	50197	7.60615	Intron	intron 11 of 12	-14007	*MTFR2*
478	chr5	95572740	46653	7.60615	Intron	intron 3 of 7	17666	*ARSK*
479	chr5	40718527	45887	7.60615	Intron	intron 4 of 4	37448	*TTC33*
480	chr4	10571585	43587	7.60615	Intron	intron 5 of 18	113280	*CLNK*
481	chr3	60927661	41425	7.60615	Intron	intron 3 of 9	-309791	*MIR548BB*
482	chr3	37874617	40740	7.60615	Intron	intron 1 of 6	12657	*CTDSPL*
483	chr22	49881229	39883	7.60615	Intron	intron 1 of 1	27387	*ZBED4*
484	chr22	49580670	39826	7.60615	Intron	intron 1 of 4	-37204	*MIR3667*
485	chr22	33602520	38854	7.60615	Intron	intron 6 of 15	-122494	*LARGE-AS1*
486	chr21	40083224	37183	7.60615	Intron	intron 24 of 32	129207	*MIR4760*
487	chr17	10183791	22284	7.60615	Intron	intron 1 of 13	14760	*GAS7*
488	chr16	10471354	19445	7.60615	Intron	intron 7 of 11	109344	*EMP2*
489	chr15	27088019	17637	7.60615	Intron	intron 3 of 9	73300	*GABRG3-AS1*
490	chr14	73443975	16661	7.60615	Intron	intron 1 of 10	14642	*NUMB*
491	chr12	118180337	13534	7.60615	Intron	intron 15 of 20	-44389	*VSIG10*
492	chr12	98545857	12977	7.60615	Intron	intron 7 of 8	-29435	*TMPO-AS1*
493	chr12	86447195	12814	7.60615	Intron	intron 2 of 8	391803	*MGAT4C*
494	chr12	41063959	12150	7.60615	Intron	intron 23 of 23	-124489	*PDZRN4*
495	chr11	118216175	10970	7.60615	Intron	intron 1 of 9	8919	*JAML*
496	chr1	215668867	5195	7.60615	Intron	intron 64 of 71	101475	*KCTD3*
497	chr4	76689572	44091	7.60615	Exon	exon 1 of 4	114031	*MIR548AH*
498	chr2	100105152	30312	7.60615	Exon	exon 1 of 1	37587	*AFF3*
499	chr15	22222299	17553	7.60615	Downstream (2-3kb)		3030	*MIR1268A*
500	chr17	21815193	22741	7.60615	Distal Intergenic		122670	*KCNJ18*

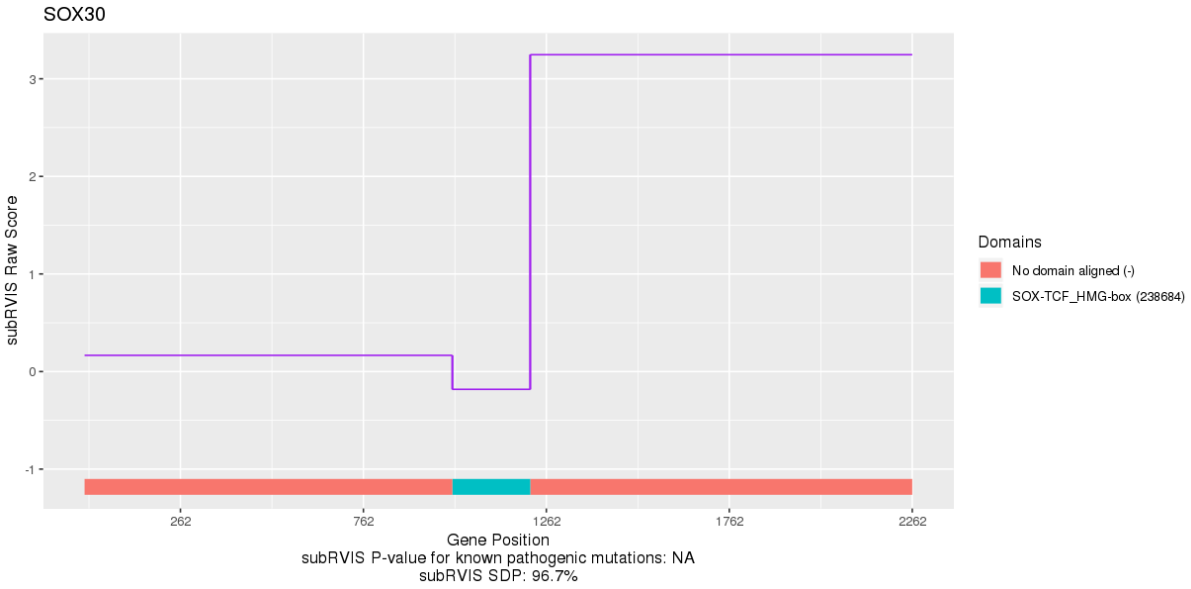


Figure S3.4. SubRVIS percentiles values across SOX30 protein. The subRVIS percentiles are a scaled version of the SubRVIS raw score, where the raw scores were converted to percentiles across the genome. Lower subRVIS percentiles correspond to more intolerant regions with threshold of 35 percentiles

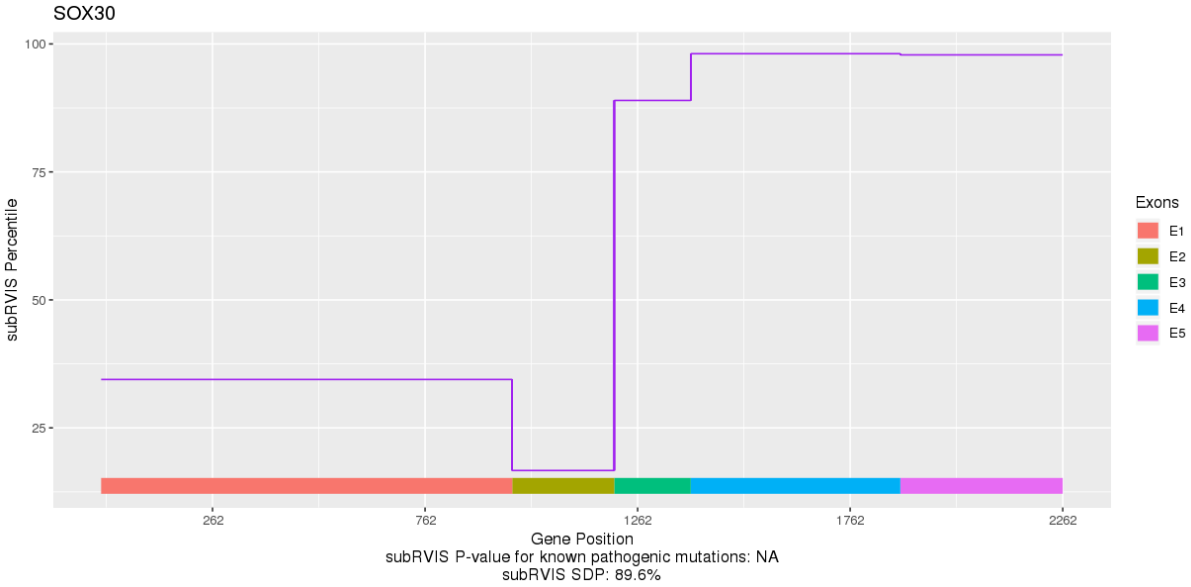


Figure S3.5. SubRVIS percentiles values across *SOX30* exons. The subRVIS percentiles are a scaled version of the SubRVIS raw scores, where the raw scores were converted to percentiles across the genome. Lower subRVIS percentiles correspond to more intolerant regions with threshold of 35 percentiles.

Appendix IV for Chapter 4

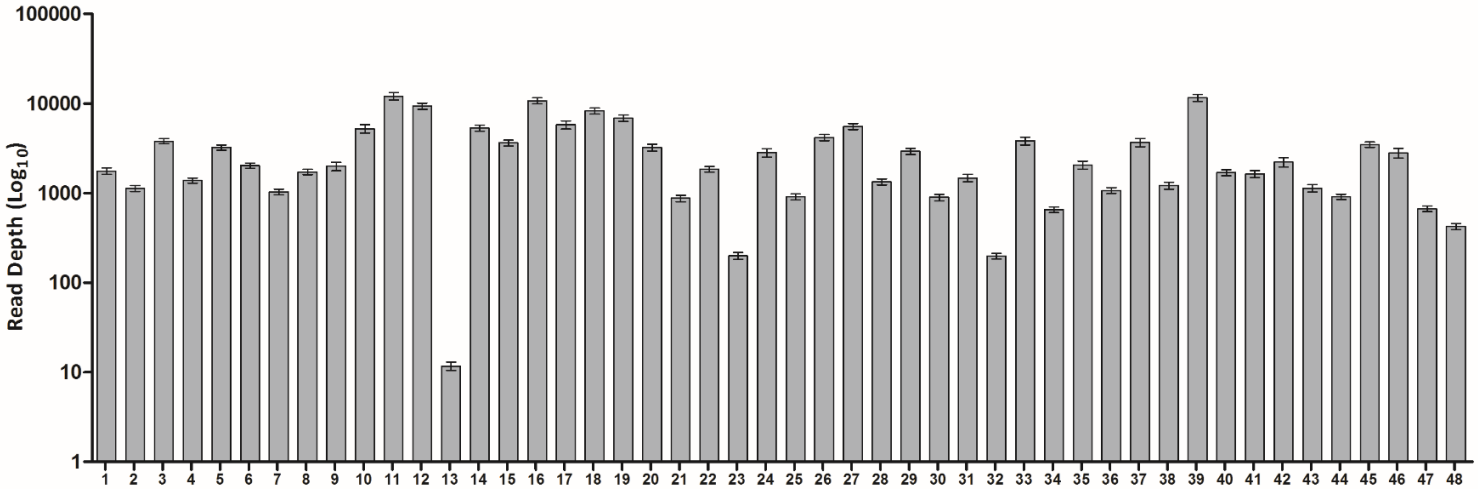


Figure A4.1.1. Average read depth plot across all exons per individual for samples 1-48 for *CHD2*

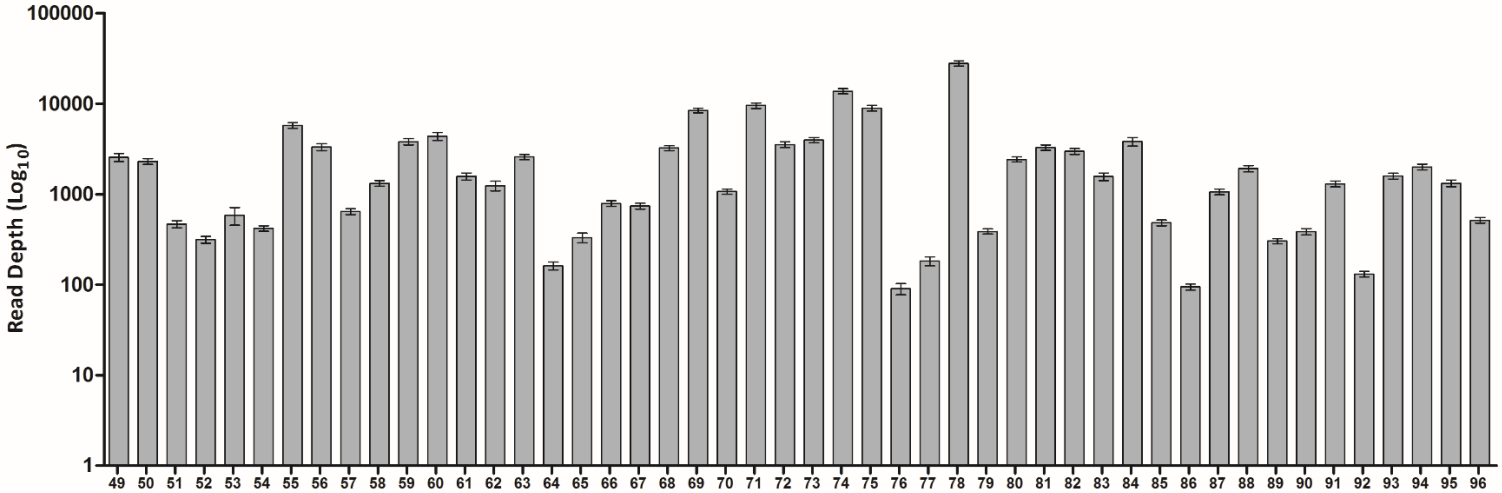


Figure A4.1.2. Average read depth plot across all exons per individual for samples 49-96 for *CHD2*

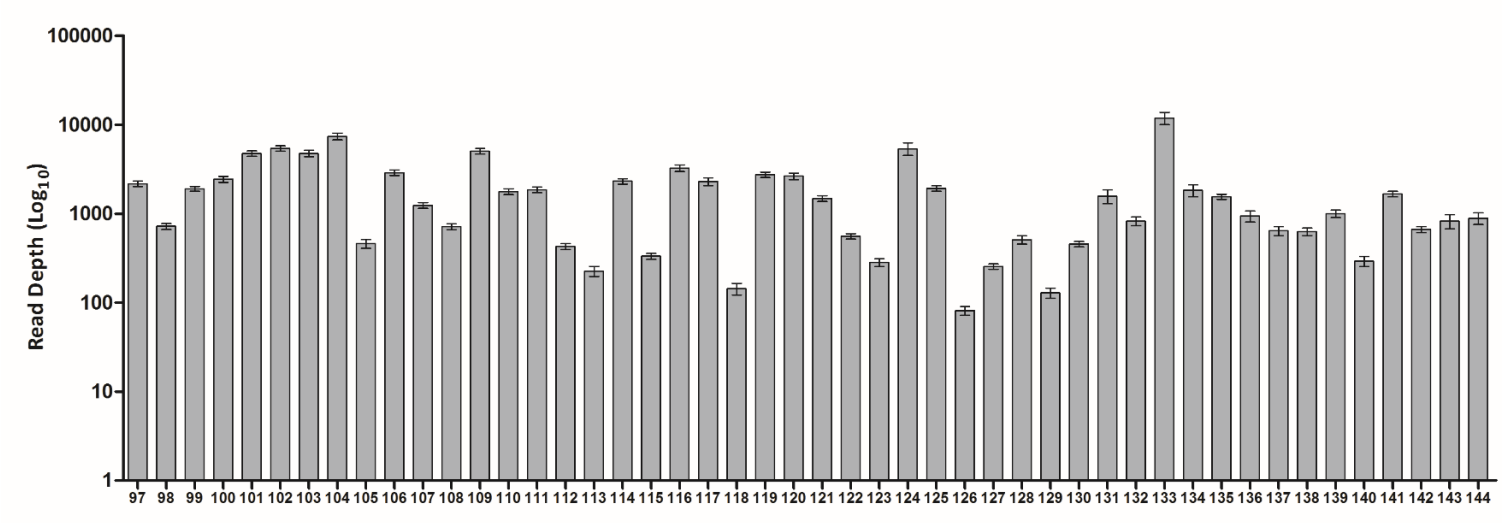


Figure A4.1.3. Average read depth across all exons per individual for samples 97-144 for *CHD2*

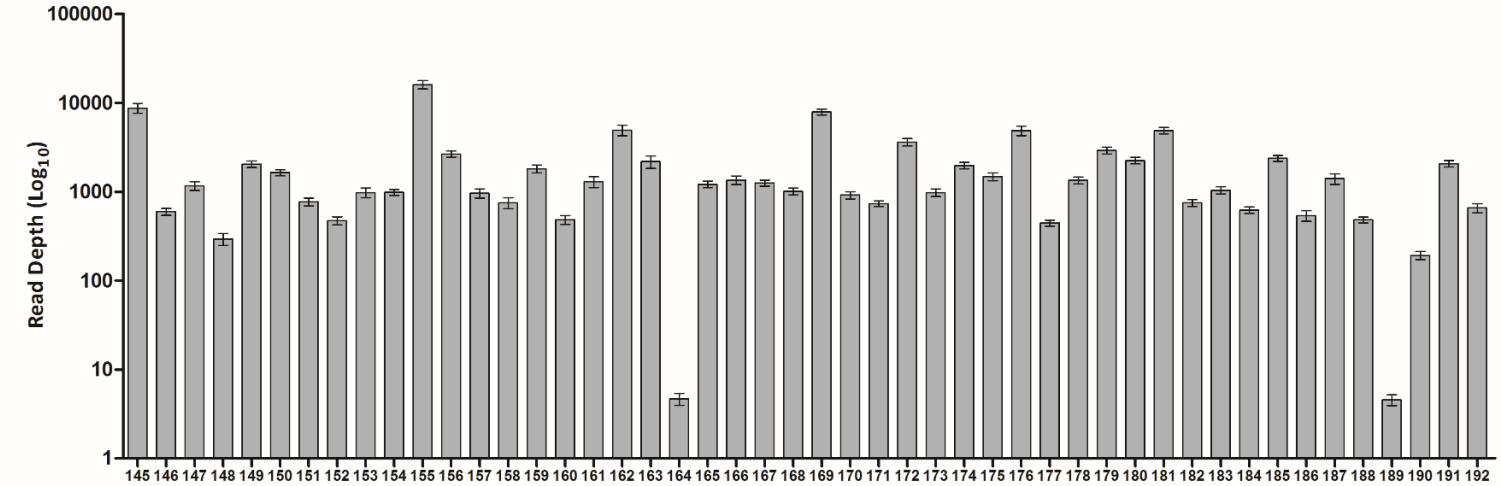


Figure A4.1.4. Average read depth plot across all exons for per individual samples 145-192 for *CHD2*

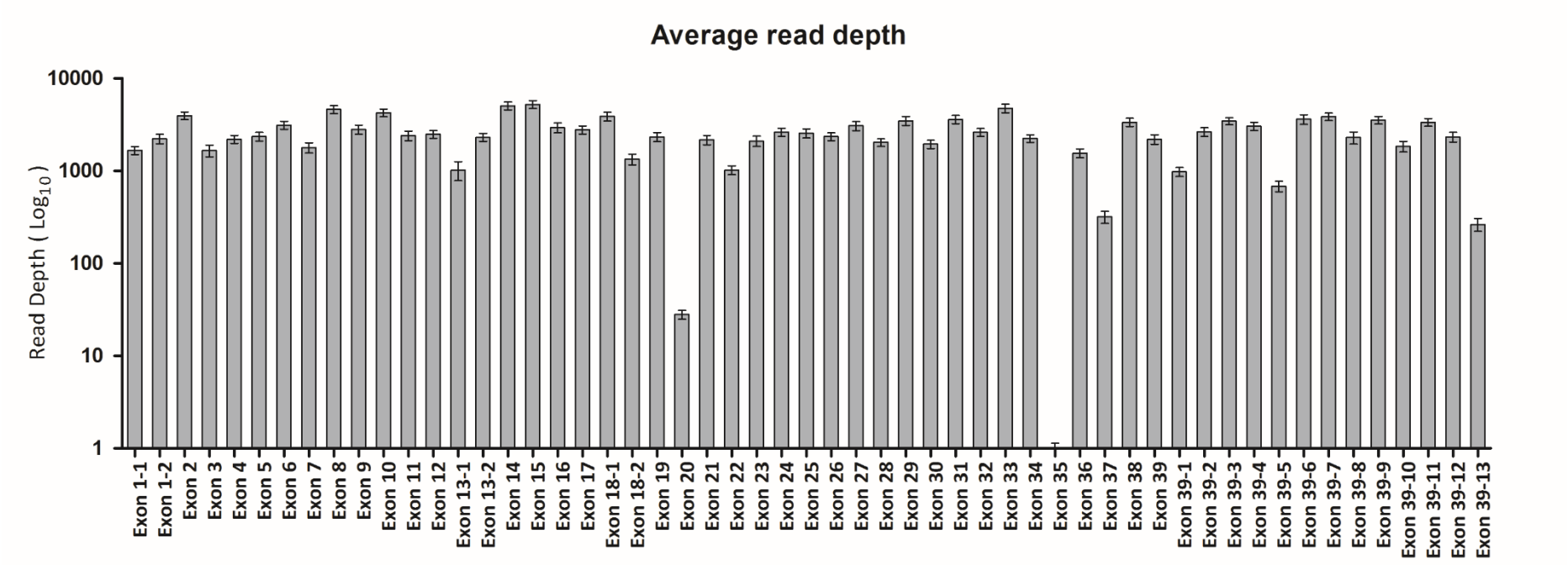


Figure A4.2. Average read depth across all samples for 55 amplicons covering 39 exons of *CHD2*

Table S4.1. Variants identified in *CHD2* by targeted sequencing in 189 JME individuals

Genomic Location (GRCh38/hg38)	Ref	Alt	Quality	Consequence	mRNA change	Amino acid change	rsID	MAF	No. of patients with variant
92900632	G	GT	8751.43	5'UTR	c.-256dupT	-	rs560673016	0.00006	1
92900722	C	T	4403.47	5'UTR	c.-174C>T	-	rs540465851	0.0004	1
92901330	CT	C	148.85	intron	c.62+40delT	-	rs769211092	0.01288	2
92928960	G	T	13014.42	intron	c.382-70G>T	-	-	-	1
92929162	C	T	39398.72	intron	c.443+71C>T	-	rs915633303	0.000127	3
92937531	C	G	49020.37	missense	c.457C>G	p.Gln153Glu	rs755510106	0.000025	1
92942923	G	A	201380	missense	c.907G>A	p.Gly303Ser	rs200687736	0.0028	9
92949144	AT	A	132334.5	intron	c.1502+75_150+76	-	rs1002924182	0	170
92955420	T	C	92124.35	splice region	c.1720-3T>C	-	rs759231634	0.00004	1
92967289	A	G	14287.36	intron	c.2001-36A>G	-	rs375992486	0.000008	1
92967373	A	G	9672450	synonymous	c.2049A>G	p.Glu683=	rs4777755	0.22836	34
92971719	T	G	32930.42	intron	c.2190-46T>C	-	rs745804520	0.000032	1
92974811	T	C	11.97	intron	c.2506-68T>C	-	-	-	1
92974833	A	G	277.46	intron	c.2506-46A>G	-	rs549581041	0.000267	1
92978374	A	G	9459064	synonymous	c.2718A>G	p.Gln906=	rs11074121	0.229063	36
92980968	A	G	444350	intron	c.2973+57A>G	-	rs8041241	0.074024	21
92984389	C	T	24224.69	synonymous	c.3126C>T	p.Asp1042=	rs150268140	0.002819	2
92985486	G	A	13256.37	intron	c.3238-12G>A	-	-	-	1
92992967	C	T	1169626	synonymous	c.3564C>T	p.Tyr1188=	rs2272457	0.224846	51
92997085	GA	G	579101.9	frameshift	c.3735delA	p.Lys1245fs	rs752940775	0	183
93002203	GA	G	33.63	frameshift	c.4173delA	p.Lys1391fs	rs749969667	0	2
93002258	T	A	24844.69	missense	c.4219T>A	p.Ser1407Thr	rs61756301	0.015300	2
93012311	A	G	20267.44	intron	c.4593-34A>G	-	rs117430127	0.008228	1
93012445	G	C	4773.44	splice donor	c.4692+1G>C	-	-	-	1
93012487	A	G	207014	intron	c.4692+43A>G	-	rs2119010	0.185788	21
93014724	G	C	9001.97	missense	c.4721G>C	p.Gly1574Ala	rs56227200	0.0239	3
93020200	C	T	21370.35	missense	c.5095C>T	p.Pro1699Ser	rs777481677	0.00003	1
93024158	T	TGTAC	3139005	intron	c.5154-214_5154-213IinsGTAC	-	rs559995137	-	95
93024192	C	G	6211.62	intron	c.5154-180C>G	-	-	-	2
93024296	G	A	4177348	intron	c.5154-76G>A	-	rs28558199	0.39081	93
93024297	C	T	513762.8	intron	c.5154-75C>T	-	rs12905032	0.177251	23
93024298	C	T	506193.2	intron	c.5154-74C>T	-	rs12905033	0.177259	23
93024582	C	T	1778.63	synonymous	c.5364C>T	p.Arg1788=	rs145582926	0.0006	2
93024634	A	C	732416	synonymous	c.5416A>C	p.Arg1806=	rs12906163	0.294623	81
93024778	G	A	56534.14	3'UTR	c.*73G>A	-	rs4778070	0.013228	5
93024953	C	T	19973.4	3'UTR	c.*248C>T	-	rs1051478439	0.000032	1
93025190	G	C	116158	3'UTR	c.*485G>C	-	rs553437821	0.0006	9
93025236	A	AGGG	11224359	3'UTR	c.*532_*535insGGG	-	rs3840036	0.3978	92
93025411	A	G	21.54	3'UTR	c.*706A>G	-	-	-	1
93025500	AC	A	15.42	3'UTR	c.*796delC	-	-	-	1
93025707	T	C	6509.41	3'UTR	c.*1002T>C	-	rs1449923910	0.00003	1
93025767	A	G	2208.4	3'UTR	c.*1062A>G	-	rs1440077664	0.000008	1
93025804	A	G	1444.65	3'UTR	c.*1099A>G	-	-		2
93026624	G	A	441540.6	3'UTR	c.*1919>A	-	rs72765626	0.04798	33
93026790	A	G	15154.69	3'UTR	c.*2085A>G	-	-	-	9
93026820	A	G	6690.35	3'UTR	c.*2115A>G	-	-	-	1
93027493	T	C	7331929	3'UTR	c.*2788T>C	-	rs13759	0.4595	99
93027767	T	C	135928.1	3'UTR	c.*3062T>C	-	rs58542701	0.015	5
93027791	G	A	8320.39	3'UTR	c.*3086>A	-	rs184443272	0.00198	1

Table S4.2. Variants identified in exon 35 of *CHD2* by Sanger sequencing.

Genomic Location (GRCh38/ hg38)	mRNA change	Amino acid change	Consequence	SNP rsiD	Occurrence in patients (n = 189)	Controls Chromosome	MAF
93009100	c.4414-45C>T		intron	rs12915582	Het-30, Hom -159	362/384	0.24118
93009119	c.4414-26G>A		intron	rs72647789	Het-5	2/384	0.025927
93009231	c.4500A>T	p.Glu1500Asp	missense	rs776113114	Het-1	0/384	0.000074
93009249	c.4518G>A	p.Leu1506=	synonymous	rs557766333	Het-1	0/384	0.000173
93009258	c.4527C>T	p.Ile1509=	synonymous	rs34315566	Het-2	3/384	0.05613
93009417	c.4592+94G>A		intron		Het-1	0/384	
93009426	c.4592+103C>T		intron	rs12900461	Het-30, Hom -159	353/384	0.23884

Table S4.3. Primer sequences for uncovered exons and Sanger confirmations of variants novel/rare (MAF<0.005) in *CHD2*

Amplicon	Forward primer	Reverse primer
CHD2-EX1-1	CTATGGTACATCACCTGAGTTG	ACAACGACCTGGGCTCTCAG
CHD2 EX1-2	TTCTTTTCGTGCCTTGTTGG	CACCAATGATCGTGAACTTAC
CHD2 EX6	TGTGCATGTTCTGCCCTTTGG	CATTTCAGCACGACATTATCC
CHD2 EX9	GATAATTGCATTCTTGGGAAGC	CTTCTTTCAGGCTGCAAATAC
CHD2 EX15	GCTTAGGTCATTTTAGAATGTGAG	CTAACATGCTCCACAGAACAG
CHD2-EX35	CTCTAGCAATTATATGGCATAGG	TTGCCAATAGATGAGTTGTGC
CHD2 EX36	CTATGCTTAGGGAGTATACC	GCTATTTCACCCATCTGGTC
CHD2 EX38	GCAGTCATCAGATCATTCTTTC	AGAAACAGCTCCTGGCAAAGG
CHD2-EX39-1	CTGATTCGGGTCACCACTCC	TGTACCTGCTCCCCTTCAAG
CHD2-EX39-2	AGCTACAGGACCTGGAGAAG	GAAACTTGTTCACCACAGGTAC
CHD2-EX39-3	TAGGCCACACAGTGATGAGG	CTGATGACTGTGGAAACCCTG

Table S4.4. Cloning and site-directed mutagenesis primer sequences for amplification of exons and flanking introns for minigene splicing assay

Name	Forward primer	Reverse primer
CHD2 EX15-SacI-BamHI	CTGAGAGCTCCTGATTCTGGCAATAAAGC	CTGAGGATCCCAACTGGACTTGGTGACC
CHD2 EX15-SDM	ATGTTTTTTTCTTACAGATACGGGAATATG	CATATTCCCGTATCTGTAAGAAAAAAACAT
CHD2 EX36-SacI-BamHI	CTGAGAGCTCTAAGTCCTCACATAACG	CTGAGGATCCAAACCTGGTGAACAAATG
CHD2 EX36-SDM	CTCAAGAAGAAGAGCTAAAGTACAAATTCA	TGAATTTGTACTTTAGCTCTTCTTCTTGAG

Table S4.5. Clinical features of patients having *CHD2* variations.

Sample no.	Mutation	Zygosity	Sex	Age of onset	Types of seizure	Precipitating factors	Current AEDs	EEG	PPR
191	c.-264_263insT	Het	M	11	MYO, GTCS	Sleep deprivation	Carbamazepine	Gen S&W	UA
122	c.-174C>T	Het	F	13	MYO	None	None	Gen S&W	NO
14	c.457C>G	Het	F	17	MYO	Within 2 hours of waking up, Emotional stress	Valproic acid	Gen S&W	NO
71	c.1720-3T>C	Het	F	17	MYO, GTCS	Emotional stress	Valproic acid	Gen S&W	YES
155	c.4500A>T	Het	F	10	MYO, GTCS	Sleep deprivation, Menses	Clobazam	UA	UA
168	c.4692+1G>C	Het	F	-	-	-	Valproic acid	-	-
135	c.5095C>T	Het	F	10	MYO, GTCS	Photic stimulation, Emotional stress	Valproic acid	Gen S&W	UA
174	c.*248C>T	Hom	M				Valproic acid	UA	UA
45	c.*1002T>C	Het	M	16	MYO, GTCS	Within 2 hours of waking up	Phenobarbital	Gen S&W	YES
180	c.*1062A>G	Het	F	12	MYO, GTCS	Within 2 hours of waking up, Sleep deprivation, Emotional stress	Valproic acid	Gen S&W	DF
171	c.*1099A>G	Het	M	-	-	-	Valproic acid	-	-
157	c.*2115A>G	Het	F	14	ABS, MYO, GTCS	Within 2 hours of waking up, Sleep deprivation	Valproic acid	UA	UA

MYO-Myoclonic jerks, GTCS – Generalised tonic clonic seizure, ABS – Absence seizures, M - Male, F – Female, UA- Unavailable, DF- Doubtful, Het – Heterozygous, Hom – Homozygous, Gen S &W – Generalised spikes and waves

References

Alkuraya FS, Cai X, Emery C, et al. Human mutations in NDE1 cause extreme microcephaly with lissencephaly. Am J Hum Genet. 2011;88(5):536-547.

Amemiya HM, Kundaje A, Boyle AP. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. Sci Rep. 2019;9(1):9354.

Amessou M, Fradagrada A, Falguières T, et al. Syntaxin 16 and syntaxin 5 are required for efficient retrograde transport of several exogenous and endogenous cargo proteins. J Cell Sci. 2007;120(Pt 8):1457-1468.

Angelozzi M, Lefebvre V. SOXopathies: Growing Family of Developmental Disorders Due to SOX Mutations. Trends Genet. 2019;35(9):658-671.

Anitha A, Senthilkumaran B. Role of sox30 in regulating testicular steroidogenesis of common carp. J Steroid Biochem Mol Biol. 2020;204:105769.

Annesi F, Gambardella A, Michelucci R, et al. Mutational analysis of EFHC1 gene in Italian families with juvenile myoclonic epilepsy. Epilepsia. 2007;48(9):1686-1690.

Arain F, Zhou C, Ding L, Zaidi S, Gallagher MJ. The developmental evolution of the seizure phenotype and cortical inhibition in mouse models of juvenile myoclonic epilepsy. Neurobiol Dis. 2015;82:164-175.

Arain FM, Boyd KL, Gallagher MJ. Decreased viability and absence-like epilepsy in mice lacking or deficient in the GABAA receptor α1 subunit. Epilepsia. 2012;53(8):e161-e165.

Aslam A, Mittal S, Koch F, Andrau JC, Winkler GS. The Ccr4-NOT deadenylase subunits CNOT7 and CNOT8 have overlapping roles and modulate cell proliferation. Mol Biol Cell. 2009;20(17):3840-3850.

Bai D, Alonso ME, Medina MT, et al. Juvenile myoclonic epilepsy: linkage to chromosome 6p12 in Mexico families. Am J Med Genet. 2002;113(3):268-274.

Bai D, Bailey JN, Durón RM, et al. DNA variants in coding region of EFHC1: SNPs do not associate with juvenile myoclonic epilepsy. Epilepsia. 2009;50(5):1184-1190.

Bai M. Structure-function relationship of the extracellular calcium-sensing receptor. Cell Calcium. 2004;35(3):197-207.

Bai S, Fu K, Yin H, et al. Sox30 initiates transcription of haploid genes during late meiosis and spermiogenesis in mouse testes. Development. 2018;145(13):dev164855.

Bai YF, Chiu M, Chan ES, et al. Pathophysiology of and therapeutic options for a GABRA1 variant linked to epileptic encephalopathy. Mol Brain. 2019;12(1):92.

Bailey JN, de Nijs L, Bai D, et al. Variant Intestinal-Cell Kinase in Juvenile Myoclonic Epilepsy. N Engl J Med. 2018;378(11):1018-1028.

Bailey ME, Matthews DA, Riley BP, et al. Genomic mapping and evolution of human GABA(A) receptor subunit gene clusters. Mamm Genome. 1999;10(8):839-843.

Bailey TL, Boden M, Buske FA, et al. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 2009;37(Web Server issue):W202-W208.

Banerjee A, Northrup D, Boukarabila H, Jacobsen SE, Allman D. Transcriptional repression of Gata3 is essential for early B cell commitment. Immunity. 2013;38(5):930-942.

Bashamboo A, Ferraz-de-Souza B, Lourenço D, et al. Human male infertility associated with mutations in NR5A1 encoding steroidogenic factor 1 . Am J Hum Genet. 2010;87(4):505-512.

Baulac S, Gourfinkel-An I, Picard F, et al. A second locus for familial generalized epilepsy with febrile seizures plus maps to chromosome 2q21-q33. Am J Hum Genet. 1999;65(4):1078-1085.

Baulac S, Huberfeld G, Gourfinkel-An I, Mitropoulou G, Beranger A, Prud'homme JF, Baulac M, Brice A, Bruzzone R, LeGuern E. First genetic evidence of GABA(A) receptor dysfunction in epilepsy: a mutation in the gamma2-subunit gene. Nat Genet. 2001 May;28(1):46-8.

Baykan B, Wolf P. Juvenile myoclonic epilepsy as a spectrum disorder: A focused review. Seizure. 2017;49:36-41.

Beghi E. The Epidemiology of Epilepsy. Neuroepidemiology. 2020;54(2):185-191.

Belkadi A, Bolze A, Itan Y, et al. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. Proc Natl Acad Sci U S A. 2015;112(17):5473-5478.

Belluardo N, Trovato-Salinaro A, Mudò G, Hurd YL, Condorelli DF. Structure, chromosomal localization, and brain expression of human Cx36 gene. J Neurosci Res. 1999;57(5):740-752.

Berger I, Dor T, Halvardson J, et al. Intractable epilepsy of infancy due to homozygous mutation in the EFHC1 gene. Epilepsia. 2012;53(8):1436-1440.

Bianchin C, Mauxion F, Sentis S, Séraphin B, Corbo L. Conservation of the deadenylase activity of proteins of the Caf1 family in human. RNA. 2005;11(4):487-494.

Bowles J, Schepers G, Koopman P. Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. Dev Biol. 2000;227(2):239-255.

Brinciotti M, Fioriello F, Mittica A, Bernardini L, Goldoni M, Matricardi M. Epilepsy phenotype in patients with Xp22.31 microduplication. Epilepsy Behav Case Rep. 2018;11:31-34.

Brown EM, Gamba G, Riccardi D, et al. Cloning and characterization of an extracellular Ca(2+)-sensing receptor from bovine parathyroid. Nature. 1993;366(6455):575-580.

Brown EM. Role of the calcium-sensing receptor in extracellular calcium homeostasis. Best Pract Res Clin Endocrinol Metab. 2013;27(3):333-343.

Buono RJ, Lohoff FW, Sander T, et al. Association between variation in the human KCNJ10 potassium ion channel gene and seizure susceptibility. Epilepsy Res. 2004;58(2-3):175-183.

Burgess DL, Jones JM, Meisler MH, Noebels JL. Mutation of the Ca2+ channel beta subunit gene Cchb4 is associated with ataxia and seizures in the lethargic (lh) mouse. Cell. 1997;88(3):385-392.

Camfield CS, Striano P, Camfield PR. Epidemiology of juvenile myoclonic epilepsy. Epilepsy Behav. 2013;28 Suppl 1:S15-S17.

Capelli LP, Krepischi AC, Gurgel-Giannetti J, et al. Deletion of the RMGA and CHD2 genes in a child with epilepsy and mental deficiency. Eur J Med Genet. 2012;55(2):132-134.

Caputo D, Trivisano M, Vigevano F, Fusco L. CHD2-epilepsy: Polygraphic documentation of self-induced seizures due to fixation-off sensitivity. Seizure. 2018;57:8-10.

Carranza Rojo D, Hamiwka L, McMahon JM, et al. De novo SCN1A mutations in migrating partial seizures of infancy. Neurology. 2011;77(4):380-383.

Carvill GL, Heavin SB, Yendle SC, et al. Targeted resequencing in epileptic encephalopathies identifies de novo mutations in CHD2 and SYNGAP1. Nat Genet. 2013;45(7):825-830.

Carvill GL, Mefford HC. CHD2-Related Neurodevelopmental Disorders. In: Adam MP, Ardinger HH, Pagon RA, et al., eds. GeneReviews®. Seattle (WA): University of Washington, Seattle; December 10, 2015.

Carvill GL, Weckhuysen S, McMahon JM, et al. GABRA1 and STXBP1: novel genetic causes of Dravet syndrome. Neurology. 2014;82(14):1245-1253.

Cavalleri GL, Walley NM, Soranzo N, et al. A multicenter study of BRD2 as a risk factor for juvenile myoclonic epilepsy. Epilepsia. 2007;48(4):706-712.

Chachua T, Goletiani C, Maglakelidze G, et al. Sex-specific behavioral traits in the Brd2 mouse model of juvenile myoclonic epilepsy. Genes Brain Behav. 2014;13(7):702-712.

Chan CK, Low JS, Lim KS, Low SK, Tan CT, Ng CC. Whole exome sequencing identifies a novel SCN1A mutation in genetic (idiopathic) generalized epilepsy and juvenile myoclonic epilepsy subtypes. Neurol Sci. 2020;41(3):591-598.

Chao HT, Davids M, Burke E, et al. A Syndromic Neurodevelopmental Disorder Caused by De Novo Variants in EBF3. Am J Hum Genet. 2017;100(1):128-137.

Chatterjee S, Pal JK. Role of 5'- and 3'-untranslated regions of mRNAs in human diseases. Biol Cell. 2009;101(5):251-262.

Chaves J, Martins-Ferreira R, Ferreira AM, et al. Immunogenetic protective factors in Genetic Generalized Epilepsy. Epilepsy Res. 2020;166:106396.

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

Chen C, Ito K, Takahashi A, et al. Distinct expression patterns of the subunits of the CCR4-NOT deadenylase complex during neural development. Biochem Biophys Res Commun. 2011;411(2):360-364.

Chen J, Zhang J, Liu A, et al. CHD2-related epilepsy: novel mutations and new phenotypes. Dev Med Child Neurol. 2020;62(5):647-653.

Chen M, Zhang L, Cui X, et al. Wt1 directs the lineage specification of sertoli and granulosa cells by repressing Sf1 expression. Development. 2017;144(1):44-53.

Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884-i890.

Chen W, Bergsman JB, Wang X, et al. Presynaptic external calcium signaling involves the calcium-sensing receptor in neocortical nerve terminals. PLoS One. 2010;5(1):e8563.

Chen Y, Lu J, Pan H, et al. Association between genetic variation of CACNA1H and childhood absence epilepsy. Ann Neurol. 2003;54(2):239-243.

Chénier S, Yoon G, Argiropoulos B, et al. CHD2 haploinsufficiency is associated with developmental delay, intellectual disability, epilepsy and neurobehavioural problems. J Neurodev Disord. 2014;6(1):9.

Chipuk JE, Bouchier-Hayes L, Kuwana T, Newmeyer DD, Green DR. PUMA couples the nuclear and cytoplasmic proapoptotic function of p53. Science. 2005;309(5741):1732-1735.

Cho A, Lima de Carvalho JR Jr, Tanaka AJ, et al. Fundoscopy-directed genetic testing to re-evaluate negative whole exome sequencing results. Orphanet J Rare Dis. 2020;15(1):32.

Christensen J, Kjeldsen MJ, Andersen H, Friis ML, Sidenius P. Gender differences in epilepsy. Epilepsia. 2005;46(6):956-960.

Chua CE, Tang BL. Syntaxin 16 is enriched in neuronal dendrites and may have a role in neurite outgrowth. Mol Membr Biol. 2008;25(1):35-45.

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

Coll M, Allegue C, Partemi S, et al. Genetic investigation of sudden unexpected death in epilepsy cohort by panel target resequencing. Int J Legal Med. 2016;130(2):331-339.

Collart MA. The Ccr4-Not complex is a key regulator of eukaryotic gene expression. Wiley Interdiscip Rev RNA. 2016;7(4):438-454.

Cossette P, Liu L, Brisebois K, et al. Mutation of GABRA1 in an autosomal dominant form of juvenile myoclonic epilepsy. Nat Genet. 2002;31(2):184-189.

Coste de Bagneaux P, von Elsner L, Bierhals T, et al. A homozygous missense variant in CACNB4 encoding the auxiliary calcium channel beta4 subunit causes a severe neurodevelopmental disorder and impairs channel and non-channel functions. PLoS Genet. 2020;16(3):e1008625. Published 2020 Mar 16.

Courage C, Houge G, Gallati S, Schjelderup J, Rieubland C. 15q26.1 microdeletion encompassing only CHD2 and RGMA in two adults with moderate intellectual disability, epilepsy and truncal obesity. Eur J Med Genet. 2014;57(9):520-523.

Crespel A, Gelisse P, Reed RC, et al. Management of juvenile myoclonic epilepsy. Epilepsy Behav. 2013;28 Suppl 1:S81-S86.

da Silva Sousa P, Lin K, Garzon E, Sakamoto AC, Yacubian EM. Self-perception of factors that precipitate or inhibit seizures in juvenile myoclonic epilepsy. Seizure. 2005;14(5):340-346.

Daghsni M, Lahbib S, Fradj M, et al. TOP3B: A Novel Candidate Gene in Juvenile Myoclonic Epilepsy?. Cytogenet Genome Res. 2018;154(1):1-5.

Das DK, Tapias V, D'Aiuto L, et al. Genetic and morphological features of human iPSC-derived neurons with chromosome 15q11.2 (BP1-BP2) deletions. Mol Neuropsychiatry. 2015;1(2):116-123.

Day NC, Volsen SG, McCormack AL, et al. The expression of voltage-dependent calcium channel beta subunits in human hippocampus. Brain Res Mol Brain Res. 1998;60(2):259-269.

de Araújo Filho GM, Pascalicchio TF, Sousa Pda S, Lin K, Ferreira Guilhoto LM, Yacubian EM. Psychiatric disorders in juvenile myoclonic epilepsy: a controlled study of 100 patients. Epilepsy Behav. 2007;10(3):437-441.

de Kovel CG, Trucks H, Helbig I, et al. Recurrent microdeletions at 15q11.2 and 16p13.11 predispose to idiopathic generalized epilepsies. Brain. 2010;133(Pt 1):23-32.

de Lange IM, Verrijn Stuart AA, van der Luijt RB, Ploos van Amstel HK, van Haelst MM. Macrosomia, obesity, and macrocephaly as first clinical presentation of PHP1b caused by STX16 deletion. Am J Med Genet A. 2016;170(9):2431-2435.

de Nijs L, Lakaye B, Coumans B, et al. EFHC1, a protein mutated in juvenile myoclonic epilepsy, associates with the mitotic spindle through its N-terminus. Exp Cell Res. 2006;312(15):2872-2879.

de Nijs L, Léon C, Nguyen L, et al. EFHC1 interacts with microtubules to regulate cell division and cortical development. Nat Neurosci. 2009;12(10):1266-1274.

de Nijs L, Wolkoff N, Coumans B, Delgado-Escueta AV, Grisar T, Lakaye B. Mutations of EFHC1, linked to juvenile myoclonic epilepsy, disrupt radial and tangential migrations during brain development. Hum Mol Genet. 2012;21(23):5106-5117.

Deans MR, Gibson JR, Sellitto C, Connors BW, Paul DL. Synchronous activity of inhibitory networks in neocortex requires electrical synapses containing connexin36. Neuron. 2001;31(3):477-485.

Dejanovic B, Lal D, Catarino CB, et al. Exonic microdeletions of the gephyrin gene impair GABAergic synaptic inhibition in patients with idiopathic generalized epilepsy. Neurobiol Dis. 2014;67:88-96.

Delgado-Escueta AV, Koeleman BP, Bailey JN, Medina MT, Durón RM. The quest for juvenile myoclonic epilepsy genes. Epilepsy Behav. 2013;28 Suppl 1:S52-S57.

Delgado-Escueta AV. Advances in genetics of juvenile myoclonic epilepsies. Epilepsy Curr. 2007;7(3):61-67.

Dhamija R, Breningstall G, Wong-Kisiel L, Dolan M, Hirsch B, Wirrell E. Microdeletion of chromosome 15q26.1 in a child with intractable generalized epilepsy. Pediatr Neurol. 2011;45(1):60-62.

Dibbens LM, Feng HJ, Richards MC, et al. GABRD encoding a protein for extra- or peri-synaptic GABAA receptors is a susceptibility locus for generalized epilepsies. Hum Mol Genet. 2004;13(13):1315-1319.

Dibbens LM, Mullen S, Helbig I, et al. Familial and sporadic 15q13.3 microdeletions in idiopathic generalized epilepsy: precedent for disorders with complex inheritance. Hum Mol Genet. 2009;18(19):3626-3631.

Ding L, Feng HJ, Macdonald RL, Botzolakis EJ, Hu N, Gallagher MJ. GABA(A) receptor alpha1 subunit mutation A322D associated with autosomal dominant juvenile myoclonic epilepsy reduces the expression and alters the composition of wild type GABA(A) receptors. J Biol Chem. 2010;285(34):26390-26405.

Doidge R, Mittal S, Aslam A, Winkler GS. The anti-proliferative activity of BTG/TOB proteins is mediated via the Caf1a (CNOT7) and Caf1b (CNOT8) deadenylase subunits of the Ccr4-not complex. PLoS One. 2012;7(12):e51331.

Doobin DJ, Kemal S, Dantas TJ, Vallee RB. Severe NDE1-mediated microcephaly results from neural progenitor cell cycle arrests at multiple specific stages. Nat Commun. 2016;7:12551.

Durner M, Keddache MA, Tomasini L, et al. Genome scan of idiopathic generalized epilepsy: evidence for major susceptibility gene and modifying genes influencing the seizure type. Ann Neurol. 2001;49(3):328-335.

Ellis CA, Churilov L, Epstein MP, et al. Epilepsy in families: Age at onset is a familial trait, independent of syndrome. Ann Neurol. 2019;86(1):91-98.

Ellis CA, Ottman R, Epstein MP, Berkovic SF; Epi4K Consortium. Generalized, focal, and combined epilepsies in families: New evidence for distinct genetic factors. Epilepsia. 2020;61(12):2667-2674.

Elmslie FV, Rees M, Williamson MP, et al. Genetic mapping of a major susceptibility locus for juvenile myoclonic epilepsy on chromosome 15q. Hum Mol Genet. 1997;6(8):1329-1334.

Epi25 Collaborative. Sub-genic intolerance, ClinVar, and the epilepsies: A whole-exome sequencing study of 29,165 individuals [published online ahead of print, 2021 Apr 28]. Am J Hum Genet. 2021;S0002-9297(21)00140-3.

Epi25 Collaborative. Ultra-Rare Genetic Variation in the Epilepsies: A Whole-Exome Sequencing Study of 17,606 Individuals. Am J Hum Genet. 2019;105(2):267-282.

Epi4K Consortium. De Novo Mutations in SLC1A2 and CACNA1A Are Important Causes of Epileptic Encephalopathies. Am J Hum Genet. 2016;99(2):287-298.

Epi4K Consortium. Epi4K: gene discovery in 4,000 genomes. Epilepsia. 2012;53(8):1457-1467.

Epi4K Consortium. Phenotypic analysis of 303 multiplex families with common epilepsies. Brain. 2017b;140(8):2144-2156.

Epi4K Consortium; Epilepsy Phenome/Genome Project, Allen AS, et al. De novo mutations in epileptic encephalopathies. Nature. 2013;501(7466):217-221.

Epi4K consortium; Epilepsy Phenome/Genome Project. Ultra-rare genetic variation in common epilepsies: a case-control sequencing study. Lancet Neurol. 2017a;16(2):135-143.

Escayg A, De Waard M, Lee DD, et al. Coding and noncoding variation of the human calcium-channel beta4-subunit gene CACNB4 in patients with idiopathic generalized epilepsy and episodic ataxia. Am J Hum Genet. 2000b;66(5):1531-1539.

Escayg A, Heils A, MacDonald BT, Haug K, Sander T, Meisler MH. A novel SCN1A mutation associated with generalized epilepsy with febrile seizures plus--and prevalence of variants in patients with epilepsy. Am J Hum Genet. 2001;68(4):866-873.

Escayg A, Jones JM, Kearney JA, Hitchcock PF, Meisler MH. Calcium channel beta 4 (CACNB4): human ortholog of the mouse epilepsy gene lethargic. Genomics. 1998;50(1):14-22.

Escayg A, MacDonald BT, Meisler MH, et al. Mutations of SCN1A, encoding a neuronal sodium channel, in two families with GEFS+2. Nat Genet. 2000a;24(4):343-345.

Etemad S, Campiglio M, Obermair GJ, Flucher BE. The juvenile myoclonic epilepsy mutant of the calcium channel β(4) subunit displays normal nuclear targeting in nerve and muscle cells. Channels (Austin). 2014;8(4):334-343.

Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32(19):3047-3048.

Farnaes L, Nahas SA, Chowdhury S, et al. Rapid whole-genome sequencing identifies a novel GABRA1 variant associated with West syndrome. Cold Spring Harb Mol Case Stud. 2017;3(5):a001776.

Fejgin K, Nielsen J, Birknow MR, et al. A mouse model that recapitulates cardinal features of the 15q13.3 microdeletion syndrome including schizophrenia- and epilepsy-related alterations. Biol Psychiatry. 2014;76(2):128-137.

Feng CA, Spiller C, Merriner DJ, O'Bryan MK, Bowles J, Koopman P. SOX30 is required for male fertility in mice. Sci Rep. 2017;7(1):17619.

Feng Y, Kapornai K, Kiss E, et al. Association of the GABRD gene and childhood-onset mood disorders. Genes Brain Behav. 2010;9(6):668-672.

Feng Y, Walsh CA. Mitotic spindle regulation by Nde1 controls cerebral cortical size. Neuron. 2004;44(2):279-293.

Ferlin A, Rocca MS, Vinanzi C, Ghezzi M, Di Nisio A, Foresta C. Mutational screening of NR5A1 gene encoding steroidogenic factor 1 in cryptorchidism and male factor infertility and functional analysis of seven undescribed mutations. FertilSteril. 2015;104(1):163-9.e1.

Fisher JL. A mutation in the GABAA receptor alpha 1 subunit linked to human epilepsy affects channel gating properties. Neuropharmacology. 2004;46(5):629-637.

Fisher RS, Acevedo C, Arzimanoglou A, et al. ILAE official report: a practical clinical definition of epilepsy. Epilepsia. 2014;55(4):475-482.

Fisher RS, Cross JH, French JA, et al. Operational classification of seizure types by the International League Against Epilepsy: Position Paper of the ILAE Commission for Classification and Terminology. Epilepsia. 2017;58(4):522-530.

Fogarty LC, Song B, Suppiah Y, et al. Bcl-xL dependency coincides with the onset of neurogenesis in the developing mammalian spinal cord. Mol Cell Neurosci. 2016;77:34-46.

Forsingdal A, Fejgin K, Nielsen V, Werge T, Nielsen J. 15q13.3 homozygous knockout mouse model display epilepsy-, autism- and schizophrenia-related phenotypes. Transl Psychiatry. 2016;6(7):e860.

Frésard L, Montgomery SB. Diagnosing rare diseases after the exome. Cold Spring Harb Mol Case Stud. 2018;4(6):a003392.

Fu Z, Gailey CD, Wang EJ, Brautigan DL. Ciliogenesis associated kinase 1: targets and functions in various organ systems. FEBS Lett. 2019;593(21):2990-3002.

Galizia EC, Myers CT, Leu C, et al. CHD2 variants are a risk factor for photosensitivity in epilepsy. Brain. 2015;138(Pt 5):1198-1207.

Gallagher MJ, Song L, Arain F, Macdonald RL. The juvenile myoclonic epilepsy GABA(A) receptor alpha1 subunit mutation A322D produces asymmetrical, subunit position-dependent reduction of heterozygous receptor currents and alpha1 subunit protein expression. J Neurosci. 2004;24(24):5570-5578.

Gama L, Breitwieser GE. A carboxyl-terminal domain controls the cooperativity for extracellular Ca2+ activation of the human calcium sensing receptor. A study with receptor-green fluorescent protein fusions. J Biol Chem. 1998;273(45):29712-29718.

Gao F, Maiti S, Alam N, et al. The Wilms tumor gene, Wt1, is required for Sox9 expression and maintenance of tubular architecture in the developing testis. Proc Natl Acad Sci U S A. 2006;103(32):11987-11992

Gao H, Mejhert N, Fretz JA, et al. Early B cell factor 1 regulates adipocyte morphology and lipolysis in white adipose tissue. Cell Metab. 2014;19(6):981-992.

Garcia-Gutierrez P, Juarez-Vicente F, Wolgemuth DJ, Garcia-Dominguez M. Pleiotrophin antagonizes Brd2 during neuronal differentiation. J Cell Sci. 2014;127(Pt 11):2554-2564.

Garel S, Marín F, Grosschedl R, Charnay P. Ebf1 controls early cell differentiation in the embryonic striatum. Development. 1999;126(23):5285-5294.

GBD 2016 Epilepsy Collaborators. Global, regional, and national burden of epilepsy, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016 Lancet Neurol. 2019;18(4):357-375.

Gee HY, Tang BL, Kim KH, Lee MG. Syntaxin 16 binds to cystic fibrosis transmembrane conductance regulator and regulates its membrane trafficking in epithelial cells. J Biol Chem. 2010;285(46):35519-35527.

Gelisse P, Genton P, Thomas P, Rey M, Samuelian JC, Dravet C. Clinical factors of drug resistance in juvenile myoclonic epilepsy. J Neurol Neurosurg Psychiatry. 2001;70(2):240-243.

Gelžinienė G, Endzinienė M, Jurkevičienė G. EEG activation by neuropsychological tasks in idiopathic generalized epilepsy of adolescence. Brain Dev. 2015;37(4):409-417.

Genton P, Gelisse P. The history of juvenile myoclonic epilepsy. Epilepsy Behav. 2013;28 Suppl 1:S2-S7.

Gilsoul M, Grisar T, Delgado-Escueta AV, de Nijs L, Lakaye B. Subtle Brain Developmental Abnormalities in the Pathogenesis of Juvenile Myoclonic Epilepsy. Front Cell Neurosci. 2019;13:433. Published 2019 Sep 27.

Gonsales MC, Ribeiro PAO, Betting LE, et al. Revisiting the clinical impact of variants in EFHC1 in patients with different phenotypes of genetic generalized epilepsy. Epilepsy Behav. 2020;112:107469.

González-García M, Pérez-Ballestero R, Ding L, et al. bcl-XL is the major bcl-x mRNA form expressed during murine development and its product localizes to mitochondria. Development. 1994;120(10):3033-3042.

Greenberg DA, Cayanis E, Strug L, et al. Malic enzyme 2 may underlie susceptibility to adolescent-onset idiopathic generalized epilepsy. Am J Hum Genet. 2005;76(1):139-146.

Greenberg DA, Durner M, Shinnar S, et al. Association of HLA class II alleles in patients with juvenile myoclonic epilepsy compared with patients with other forms of adolescent-onset generalized epilepsy. Neurology. 1996;47(3):750-755.

Grisar T, Lakaye B, de Nijs L, LoTurco J, Daga A, Delgado-Escueta AV. Myoclonin1/EFHC1 in cell division, neuroblast migration, synapse/dendrite formation in juvenile myoclonic epilepsy. In: Noebels JL, Avoli M, Rogawski MA, Olsen RW, Delgado-Escueta AV, eds. Jasper's Basic Mechanisms of the Epilepsies. 4th ed. Bethesda (MD): National Center for Biotechnology Information (US); 2012.

Gu W, Sander T, Heils A, Lenzen KP, Steinlein OK. A new EF-hand containing gene EFHC2 on Xp11.4: tentative evidence for association with juvenile myoclonic epilepsy. Epilepsy Res. 2005;66(1-3):91-98.

Guaranha MS, da Silva Sousa P, de Araújo-Filho GM, et al. Provocative and inhibitory effects of a video-EEG neuropsychologic protocol in juvenile myoclonic epilepsy. Epilepsia. 2009;50(11):2446-2455.

Guaranha MS, Filho GM, Lin K, Guilhoto LM, Caboclo LO, Yacubian EM. Prognosis of juvenile myoclonic epilepsy is related to endophenotypes. Seizure. 2011;20(1):42-48.

Gubbay J, Collignon J, Koopman P, et al. A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. Nature. 1990;346(6281):245-250.

Guerrini R, Marini C, Barba C. Generalized epilepsies. Handb Clin Neurol. 2019;161:3-15.

Gussow AB, Petrovski S, Wang Q, Allen AS, Goldstein DB. The intolerance to functional genetic variation of protein domains predicts the localization of pathogenic mutations within genes. Genome Biol. 2016;17:9.

Gyuris A, Donovan DJ, Seymour KA, et al. The chromatin-targeting protein Brd2 is required for neural tube closure and embryogenesis. Biochim Biophys Acta. 2009;1789(5):413-421.

Haenisch S, Zhao Y, Chhibber A, et al. SOX11 identified by target gene evaluation of miRNAs differentially expressed in focal and non-focal brain tissue of therapy-resistant epilepsy patients. Neurobiol Dis. 2015;77:127-140.

Hagman J, Belanger C, Travis A, Turck CW, Grosschedl R. Cloning and functional characterization of early B-cell factor, a regulator of lymphocyte-specific gene expression. Genes Dev. 1993;7(5):760-773.

Hagman J, Travis A, Grosschedl R. A novel lineage-specific nuclear factor regulates mb-1 gene transcription at the early stages of B cell differentiation. EMBO J. 1991;10(11):3409-3417.

Hamdan FF, Srour M, Capo-Chichi JM, et al. De novo mutations in moderate or severe intellectual disability. PLoS Genet. 2014;10(10):e1004772.

Han F, Dong Y, Liu W, et al. Epigenetic regulation of sox30 is associated with testis development in mice. PLoS One. 2014;9(5):e97203.

Han F, Jiang X, Li ZM, et al. Epigenetic Inactivation of SOX30 Is Associated with Male Infertility and Offers a Therapy Target for Non-obstructive Azoospermia. Mol Ther Nucleic Acids. 2020;19:72-83.

Han F, Liu W, Jiang X, et al. SOX30, a novel epigenetic silenced tumor suppressor, promotes tumor cell apoptosis by transcriptional activating p53 in lung cancer. Oncogene. 2015a;34(33):4391-4402.

Han F, Liu W, Xiao H, et al. High expression of SOX30 is associated with favorable survival in human lung adenocarcinoma. Sci Rep. 2015b;5:13630.

Han F, Liu WB, Li JJ, et al. SOX30 is a prognostic biomarker and chemotherapeutic indicator for advanced-stage ovarian cancer. Endocr Relat Cancer. 2019;26(3):303-319.

Han F, Liu WB, Shi XY, et al. SOX30 Inhibits Tumor Metastasis through Attenuating Wnt-Signaling via Transcriptional and Posttranslational Regulation of β-Catenin in Lung Cancer. EBioMedicine. 2018a;31:253-266.

Han F, Zhang MQ, Liu WB, et al. SOX30 specially prevents Wnt-signaling to suppress metastasis and improve prognosis of lung adenocarcinoma patients. Respir Res. 2018b;19(1):241.

Hao X, Han F, Ma B, et al. SOX30 is a key regulator of desmosomal gene suppressing tumor growth and metastasis in lung adenocarcinoma. J Exp Clin Cancer Res. 2018;37(1):111.

Harkin LA, Bowser DN, Dibbens LM, et al. Truncation of the GABA(A)-receptor gamma2 subunit in a family with generalized epilepsy with febrile seizures plus. Am J Hum Genet. 2002;70(2):530-536.

Harms FL, Girisha KM, Hardigan AA, et al. Mutations in EBF3 Disturb Transcriptional Profiles and Cause Intellectual Disability, Ataxia, and Facial Dysmorphism. Am J Hum Genet. 2017;100(1):117-127.

Hayward A, Ghazal A, Andersson G, Andersson L, Jern P. ZBED evolution: repeated utilization of DNA transposons as regulators of diverse host functions. PLoS One. 2013;8(3):e59940.

Heenan P, Zondag L, Wilson MJ. Evolution of the Sox gene family within the chordate phylum. Gene. 2016;575(2 Pt 2):385-392.

Heinzen EL, Depondt C, Cavalleri GL, et al. Exome sequencing followed by large-scale genotyping fails to identify single rare variants of large effect in idiopathic generalized epilepsy. Am J Hum Genet. 2012;91(2):293-302.

Helbig I, Mefford HC, Sharp AJ, et al. 15q13.3 microdeletions increase risk of idiopathic generalized epilepsy. Nat Genet. 2009;41(2):160-162.

Hempelmann A, Heils A, Sander T. Confirmatory evidence for an association of the connexin-36 gene with juvenile myoclonic epilepsy. Epilepsy Res. 2006;71(2-3):223-228.

Hendy GN, Guarnieri V, Canaff L. Calcium-sensing receptor and associated diseases. Prog Mol Biol Transl Sci. 2009;89:31-95.

Heron SE, Khosravani H, Varela D, et al. Extended spectrum of idiopathic generalized epilepsies associated with CACNA1H functional variants. Ann Neurol. 2007;62(6):560-568.

Heyne HO, Artomov M, Battke F, et al. Targeted gene sequencing in 6994 individuals with neurodevelopmental disorder with epilepsy. Genet Med. 2019;21(11):2496-2503.

Ho C, Conner DA, Pollak MR, et al. A mouse model of human familial hypocalciuric hypercalcemia and neonatal severe hyperparathyroidism. Nat Genet. 1995;11(4):389-394.

Hormuzdi SG, Pais I, LeBeau FE, et al. Impaired electrical signaling disrupts gamma frequency oscillations in connexin 36-deficient mice. Neuron. 2001;31(3):487-495.

Hornbeck PV, Kornhauser JM, Tkachev S, et al. PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. Nucleic Acids Res. 2012;40(Database issue):D261-D270.

Hosoyamada S, Sasaki M, Kobayashi T. The CCR4-NOT Complex Maintains Stability and Transcription of rRNA Genes by Repressing Antisense Transcripts. Mol Cell Biol. 2019;40(1):e00320-19.

Hough TA, Bogani D, Cheeseman MT, et al. Activating calcium-sensing receptor mutation in the mouse is associated with cataracts and ectopic calcification. Proc Natl Acad Sci U S A. 2004;101(37):13566-13571.

Huang C, Miller RT. The calcium-sensing receptor and its interacting proteins. J Cell Mol Med. 2007;11(5):923-934.

Huang H, Rambaldi I, Daniels E, Featherstone M. Expression of the Wdr9 gene and protein products during mouse development. Dev Dyn. 2003;227(4):608-614.

Huang Y, Zhou Y, Castiblanco A, Yang W, Brown EM, Yang JJ. Multiple Ca(2+)-binding sites in the extracellular domain of the Ca(2+)-sensing receptor corresponding to cooperative Ca(2+) response. Biochemistry. 2009;48(2):388-398.

Hung MC, Link W. Protein localization in disease and therapy. J Cell Sci. 2011;124(Pt 20):3381-3392.

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

ILAE International League Against Epilepsy Consortium on Complex Epilepsies. Genome-wide mega-analysis identifies 16 loci and highlights diverse biological mechanisms in the common epilepsies. Nat Commun. 2018;9(1):5269.

Jacob TC, Moss SJ, Jurd R. GABA(A) receptor trafficking and its role in the dynamic modulation of neuronal inhibition. Nat Rev Neurosci. 2008;9(5):331-343.

Jähn JA, von Spiczak S, Muhle H, et al. Iterative phenotyping of 15q11.2, 15q13.3 and 16p13.11 microdeletion carriers in pediatric epilepsies. Epilepsy Res. 2014;108(1):109-116.

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

Jara-Prado A, Martínez-Juárez IE, Ochoa A, et al. Novel Myoclonin1/EFHC1 mutations in Mexican patients with juvenile myoclonic epilepsy. Seizure. 2012;21(7):550-554.

Jayalakshmi SS, Mohandas S, Sailaja S, Borgohain R. Clinical and electroencephalographic study of first-degree relatives and probands with juvenile myoclonic epilepsy. Seizure. 2006;15(3):177-183.

Jiang T, Hou CC, She ZY, Yang WX. The SOX gene family: function and regulation in testis determination and male fertility maintenance. Mol Biol Rep. 2013;40(3):2187-2194.

Jimenez MA, Akerblad P, Sigvardsson M, Rosen ED. Critical role for Ebf1 and Ebf2 in the adipogenic transcriptional cascade. Mol Cell Biol. 2007;27(2):743-757.

Johannesen K, Marini C, Pfeffer S, et al. Phenotypic spectrum of GABRA1: From generalized epilepsies to severe epileptic encephalopathies. Neurology. 2016;87(11):1140-1151.

Jones BL, Smith SM. Calcium-Sensing Receptor: A Key Target for Extracellular Calcium Signaling in Neurons. Front Physiol. 2016;7:116.

Kamachi Y, Kondoh H. Sox proteins: regulators of cell fate specification and differentiation. Development. 2013;140(20):4129-4144.

Kamachi Y, Uchikawa M, Kondoh H. Pairing SOX off: with partners in the regulation of embryonic development. Trends Genet. 2000;16(4):182-187.

Kananura C, Haug K, Sander T, et al. A splice-site mutation in GABRG2 associated with childhood absence epilepsy and febrile convulsions. Arch Neurol. 2002;59(7):1137-1141.

Kapoor A, Satishchandra P, Ratnapriya R, et al. An idiopathic epilepsy syndrome linked to 3q13.3-q21 and missense mutations in the extracellular calcium sensing receptor gene. Ann Neurol. 2008;64(2):158-167.

Kapoor A, Vijai J, Ravishankar HM, Satishchandra P, Radhakrishnan K, Anand A. Absence of GABRA1 Ala322Asp mutation in juvenile myoclonic epilepsy families from India. J Genet. 2003;82(1-2):17-21.

Kasak L, Laan M. Monogenic causes of non-obstructive azoospermia: challenges, established knowledge, limitations and perspectives. Hum Genet. 2021;140(1):135-154.

Kasper, D. L., A. S. Fauci, S. L. Hauser, D. L. Longo, J. L. Jameson & J. Loscalzo. 2015. Harrison's principles of internal medicine.: McGraw Hill Education Medical

Katano M, Numata T, Aguan K, et al. The juvenile myoclonic epilepsy-related protein EFHC1 interacts with the redox-sensitive TRPM2 channel linked to cell death. Cell Calcium. 2012;51(2):179-185.

Kay B, Szaflarski JP. EEG/fMRI contributions to our understanding of genetic generalized epilepsies. Epilepsy Behav. 2014;34:129-135.

Kim YJ, Khoshkhoo S, Frankowski JC, et al. Chd2 Is Necessary for Neural Circuit Development and Long-Term Memory. Neuron. 2018;100(5):1180-1193.e6.

Kimura H, Tsuboi D, Wang C, et al. Identification of Rare, Single-Nucleotide Mutations in NDE1 and Their Contributions to Schizophrenia Susceptibility. Schizophr Bull. 2015;41(3):744-753.

Kjeldsen MJ, Corey LA, Solaas MH, et al. Genetic factors in seizures: a population-based study of 47,626 US, Norwegian and Danish twin pairs. Twin Res Hum Genet. 2005;8(2):138-147.

Klassen T, Davis C, Goldman A, et al. Exome sequencing of ion channel genes reveals complex profiles confounding personal risk assessment in epilepsy. Cell. 2011;145(7):1036-1048

Koch P, Löhr HB, Driever W. A mutation in cnot8, component of the Ccr4-not complex regulating transcript stability, affects expression levels of developmental regulators and reveals a role of Fgf3 in development of caudal hypothalamic dopaminergic neurons. PLoS One. 2014;9(12):e113829.

Kodera H, Ohba C, Kato M, et al. De novo GABRA1 mutations in Ohtahara and West syndromes. Epilepsia. 2016;57(4):566-573.

Kormish JD, Sinner D, Zorn AM. Interactions between SOX factors and Wnt/beta-catenin signaling in development and disease. Dev Dyn. 2010;239(1):56-68.

Kos CH, Karaplis AC, Peng JB, et al. The calcium-sensing receptor is required for normal calcium homeostasis independent of parathyroid hormone. J Clin Invest. 2003;111(7):1021-1028.

Kralic JE, Criswell HE, Osterman JL, et al. Genetic essential tremor in gamma-aminobutyric acidA receptor alpha1 subunit knockout mice. J Clin Invest. 2005;115(3):774-779.

Krenn M, Ernst M, Tomschik M, et al. Phenotypic variability of GABRA1-related epilepsy in monozygotic twins. Ann Clin Transl Neurol. 2019;6(11):2317-2322.

Kudrycki K, Stein-Izsak C, Behn C, Grillo M, Akeson R, Margolis FL. Olf-1-binding site: characterization of an olfactory neuron-specific promoter motif. Mol Cell Biol. 1993;13(5):3002-3014.

Kulkarni S, Nagarajan P, Wall J, et al. Disruption of chromodomain helicase DNA binding protein 2 (CHD2) causes scoliosis. Am J Med Genet A. 2008;146A(9):1117-1127.

Lachance-Touchette P, Brown P, Meloche C, et al. Novel α1 and γ2 GABAA receptor subunit mutations in families with idiopathic generalized epilepsy. Eur J Neurosci. 2011;34(2):237-249.

Lachance-Touchette P, Choudhury M, Stoica A, Di Cristo G, Cossette P. Single-cell genetic expression of mutant GABAA receptors causing Human genetic epilepsy alters dendritic spine and GABAergic bouton formation in a mutation-specific manner. Front Cell Neurosci. 2014;8:317.

Lahiry P, Wang J, Robinson JF, et al. A multiplex human syndrome implicates a key role for intestinal cell kinase in development of central nervous, skeletal, and endocrine systems [published correction appears in Am J Hum Genet. 2009 Jun;84(6):822]. Am J Hum Genet. 2009;84(2):134-147.

Lal D, Reinthaler EM, Dejanovic B, et al. Evaluation of Presumably Disease Causing SCN1A Variants in a Cohort of Common Epilepsy Syndromes. PLoS One. 2016;11(3):e0150426.

Lal D, Ruppert AK, Trucks H, et al. Burden analysis of rare microdeletions suggests a strong impact of neurodevelopmental genes in genetic generalised epilepsies. PLoS Genet. 2015;11(5):e1005226.

Lal D, Trucks H, Møller RS, et al. Rare exonic deletions of the RBFOX1 gene increase risk of idiopathic generalized epilepsy. Epilepsia. 2013;54(2):265-271.

Lamar KJ, Carvill GL. Chromatin Remodeling Proteins in Epilepsy: Lessons From CHD2-Associated Epilepsy. Front Mol Neurosci. 2018;11:208.

Landmark C J, Fløgstad I, Baftiu A, et al. Long-term follow-up with therapeutic drug monitoring of antiepileptic drugs in patients with juvenile myoclonic epilepsy. Epilepsy Res. 2019a;155:106148.

Landmark C J, Fløgstad I, Syvertsen M, et al. Treatment and challenges with antiepileptic drugs in patients with juvenile myoclonic epilepsy. Epilepsy Behav. 2019b;98(Pt A):110-116.

Landoulsi Z, Laatar F, Noé E, et al. Clinical and genetic study of Tunisian families with genetic generalized epilepsy: contribution of CACNA1H and MAST4 genes. Neurogenetics. 2018;19(3):165-178.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357-359.

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

Layouni S, Chouchane L, Malafosse A, Dogui M. Dimorphism of TAP-1 gene in Caucasian with juvenile myoclonic epilepsy and in Tunisian with idiopathic generalized epilepsies. Int J Immunogenet. 2010b;37(2):117-123.

Lee BD, Walss-Bass C, Thompson PM, et al. Malic enzyme 2 and susceptibility to psychosis and mania. Psychiatry Res. 2007;150(1):1-11.

Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536(7616):285-291.

Lenzen KP, Heils A, Lorenz S, et al. Supportive evidence for an allelic association of the human KCNJ10 potassium channel gene with idiopathic generalized epilepsy. Epilepsy Res. 2005a;63(2-3):113-118.

Lenzen KP, Heils A, Lorenz S, Hempelmann A, Sander T. Association analysis of malic enzyme 2 gene polymorphisms with idiopathic generalized epilepsy. Epilepsia. 2005b;46(10):1637-1641.

Lenzen KP, Heils A, Lorenz S, Hempelmann A, Sander T. Association analysis of the Arg220His variation of the human gene encoding the GABA delta subunit with idiopathic generalized epilepsy. Epilepsy Res. 2005c;65(1-2):53-57.

Léon C, de Nijs L, Chanas G, Delgado-Escueta AV, Grisar T, Lakaye B. Distribution of EFHC1 or Myoclonin 1 in mouse neural structures. Epilepsy Res. 2010;88(2-3):196-207.

Lerche H, Berkovic SF, Lowenstein DH; EuroEPINOMICS-CoGIE Consortium; EpiPGX Consortium; Epi4K Consortium/Epilepsy Phenome/Genome Project. Intestinal-Cell Kinase and Juvenile Myoclonic Epilepsy. N Engl J Med. 2019;380(16):e24.

Leu C, Stevelink R, Smith AW, et al. Polygenic burden in focal and generalized epilepsies. Brain. 2019;142(11):3473-3481.

Lin H, Grosschedl R. Failure of B-cell differentiation in mice lacking the transcription factor EBF. Nature. 1995;376(6537):263-267.

Lindenboim L, Borner C, Stein R. Bcl-x(S) can form homodimers and heterodimers and its apoptotic activity requires localization of Bcl-x(S) to the mitochondria and its BH3 and loop domains. Cell Death Differ. 2001;8(9):933-942.

Linglart A, Gensure RC, Olney RC, Jüppner H, Bastepe M. A novel STX16 deletion in autosomal dominant pseudohypoparathyroidism type Ib redefines the boundaries of a cis-acting imprinting control element of GNAS. Am J Hum Genet. 2005;76(5):804-814.

Lioubinski O, Müller M, Wegner M, Sander M. Expression of Sox transcription factors in the developing mouse pancreas. Dev Dyn. 2003;227(3):402-408.

Liu AW, Delgado-Escueta AV, Gee MN, et al. Juvenile myoclonic epilepsy in chromosome 6p12-p11: locus heterogeneity and recombinations. Am J Med Genet. 1996;63(3):438-446.

Liu AW, Delgado-Escueta AV, Serratosa JM, et al. Juvenile myoclonic epilepsy locus in chromosome 6p21.2-p11: linkage to convulsions and electroencephalography trait. Am J Hum Genet. 1995;57(2):368-381.

Liu Y, Wang W, Li Y, Huang Y. SOX30 confers a tumor suppressive effect in acute myeloid leukemia through inactivation of Wnt/β-catenin signaling. Mol Cell Probes. 2020;52:101578.

Lobo MK, Karsten SL, Gray M, Geschwind DH, Yang XW. FACS-array profiling of striatal projection neuron subtypes in juvenile and adult mouse brains. Nat Neurosci. 2006;9(3):443-452.

Loucks CM, Park K, Walker DS, et al. EFHC1, implicated in juvenile myoclonic epilepsy, functions at the cilium and synapse to modulate dopamine signaling. Elife. 2019;8:e37271.

Lund C, Brodtkorb E, Øye AM, Røsby O, Selmer KK. CHD2 mutations in Lennox-Gastaut syndrome. Epilepsy Behav. 2014;33:18-21.

Ma S, Abou-Khalil B, Blair MA, Sutcliffe JS, Haines JL, Hedera P. Mutations in GABRA1, GABRA5, GABRG2 and GABRD receptor genes are not a major factor in the pathogenesis of familial focal epilepsy preceded by febrile seizures. Neurosci Lett. 2006a;394(1):74-78.

Ma S, Blair MA, Abou-Khalil B, Lagrange AH, Gurnett CA, Hedera P. Mutations in the GABRA1 and EFHC1 genes are rare in familial juvenile myoclonic epilepsy. Epilepsy Res. 2006b;71(2-3):129-134.

Maguire J, Mody I. GABA(A)R plasticity during pregnancy: relevance to postpartum depression. Neuron. 2008;59(2):207-213.

Malas S, Postlethwaite M, Ekonomou A, et al. Sox1-deficient mice suffer from epilepsy associated with abnormal ventral forebrain development and olfactory cortex hyperexcitability. Neuroscience. 2003;119(2):421-432.

Malki S, Boizet-Bonhoure B, Poulat F. Shuttling of SOX proteins. Int J Biochem Cell Biol. 2010;42(3):411-416.

Mandal M, Maienschein-Cline M, Maffucci P, et al. BRWD1 orchestrates epigenetic landscape of late B lymphopoiesis. Nat Commun. 2018;9(1):3888.

Mantoan L, Walker M. Treatment options in juvenile myoclonic epilepsy. Curr Treat Options Neurol. 2011;13(4):355-370.

Marfella CG, Henninger N, LeBlanc SE, et al. A mutation in the mouse Chd2 chromatin remodeling enzyme results in a complex renal phenotype. Kidney Blood Press Res. 2008;31(6):421-432.

Marfella CG, Ohkawa Y, Coles AH, Garlick DS, Jones SN, Imbalzano AN. Mutation of the SNF2 family member Chd2 affects mouse development and survival . J Cell Physiol. 2006;209(1):162-171.

Marson AG, Al-Kharusi AM, Alwaidh M, et al. The SANAD study of effectiveness of valproate, lamotrigine, or topiramate for generalised and unclassifiable epilepsy: an unblinded randomised controlled trial. Lancet. 2007;369(9566):1016-1026.

Martínez-Juárez IE, Alonso ME, Medina MT, et al. Juvenile myoclonic epilepsy subsyndromes: family studies and long-term follow-up. Brain. 2006;129(Pt 5):1269-1280.

Mas C, Taske N, Deutsch S, et al. Association of the connexin36 gene with juvenile myoclonic epilepsy. J Med Genet. 2004;41(7):e93.

Mattheisen GB, Tsintsadze T, Smith SM. Strong G-Protein-Mediated Inhibition of Sodium Channels. Cell Rep. 2018;23(9):2770-2781.

May P, Girard S, Harrer M, et al. Rare coding variants in genes encoding GABAA receptors in genetic generalised epilepsies: an exome-based case-control study. Lancet Neurol. 2018;17(8):699-708.

Mayer TA, Schroeder F, May TW, Wolf PT. Perioral reflex myoclonias: a controlled study in patients with JME and focal epilepsies. Epilepsia. 2006;47(6):1059-1067.

McKenna MC, Stevenson JH, Huang X, Tildon JT, Zielke CL, Hopkins IB. Mitochondrial malic enzyme activity is much higher in mitochondria from cortical synaptic terminals compared with mitochondria from primary cultures of cortical neurons or cerebellar granule cells. Neurochem Int. 2000;36(4-5):451-459.

McKernan RM, Quirk K, Prince R, et al. GABAA receptor subtypes immunopurified from rat brain with alpha subunit-specific antibodies have unique pharmacological properties. Neuron. 1991;7(4):667-676.

Medina MT, Suzuki T, Alonso ME, et al. Novel mutations in Myoclonin1/EFHC1 in sporadic and familial juvenile myoclonic epilepsy. Neurology. 2008;70(22 Pt 2):2137-2144.

Mefford HC, Muhle H, Ostertag P, et al. Genome-wide copy number variation in epilepsy: novel susceptibility loci in idiopathic generalized and focal epilepsies. PLoS Genet. 2010;6(5):e1000962.

Meganathan K, Lewis EMA, Gontarz P, et al. Regulatory networks specifying cortical interneurons from human embryonic stem cells reveal roles for CHD2 in interneuron development. Proc Natl Acad Sci U S A. 2017;114(52):E11180-E11189.

Mehndiratta, M., K. B. Rao, S. Singh, S. Ganesh & G. Khwaja. Genetic study of juvenile myoclonic epilepsy (JME) patients and their family members in a University Hospital in North India. Neurology asia, 2007;12, 105 – 106.

Meienberg J, Bruggmann R, Oexle K, Matyas G. Clinical sequencing: is WGS the better WES?. Hum Genet. 2016;135(3):359-362.

Michels J, Kepp O, Senovilla L, et al. Functions of BCL-X L at the Interface between Cell Death and Metabolism. Int J Cell Biol. 2013;2013:705294.

Miller JE, Reese JC. Ccr4-Not complex: the control freak of eukaryotic cells. Crit Rev Biochem Mol Biol. 2012;47(4):315-333.

Millichap JG, Millichap JJ. Neuropsychological-EEG Activation in Genetic Generalized Epilepsy. Pediatr Neurol Briefs. 2015;29(3):21.

Miyamoto, T., M. Iijima, T. Shin, M. Ikezawa, Y. Utsuno, Y. Saijo, H. Okada & K. Sengoku (2020) SOX30 might not be associated with Sertoli cell-only syndrome in azoospermic Japanese men. Integr Mol Med 7.

Møller RS, Weber YG, Klitten LL, et al. Exon-disrupting deletions of NRXN1 in idiopathic generalized epilepsy [published correction appears in Epilepsia. 2013 Dec;54(12):2232]. Epilepsia. 2013;54(2):256-264.

Møller RS, Wuttke TV, Helbig I, et al. Mutations in GABRB3: From febrile seizures to epileptic encephalopathies. Neurology. 2017;88(5):483-492.

Moon H, Song J, Shin JO, et al. Intestinal cell kinase, a protein associated with endocrine-cerebro-osteodysplasia syndrome, is a key regulator of cilia length and Hedgehog signaling. Proc Natl Acad Sci U S A. 2014;111(23):8541-8546.

Moruzzo D, Nobbio L, Sterlini B, et al. The Transcription Factors EBF1 and EBF2 Are Positive Regulators of Myelination in Schwann Cells. Mol Neurobiol. 2017;54(10):8117-8127.

Motoyama N, Wang F, Roth KA, et al. Massive cell death of immature hematopoietic cells and neurons in Bcl-x-deficient mice. Science. 1995;267(5203):1506-1510.

Moulard B, Guipponi M, Chaigne D, Mouthon D, Buresi C, Malafosse A. Identification of a new locus for generalized epilepsy with febrile seizures plus (GEFS+) on chromosome 2q24-q33. Am J Hum Genet. 1999;65(5):1396-1400.

Muhle H, von Spiczak S, Gaus V, et al. Role of GRM4 in idiopathic generalized epilepsies analysed by genetic association and sequence analysis. Epilepsy Res. 2010;89(2-3):319-326.

Mullen SA, Berkovic SF; ILAE Genetics Commission. Genetic generalized epilepsies. Epilepsia. 2018;59(6):1148-1153.

Mullen SA, Carvill GL, Bellows S, et al. Copy number variants are frequent in genetic generalized epilepsy with intellectual disability [published correction appears in Neurology. 2013 Dec 10;81(24):2148. Trucks, Holger [added]; Lal, Dennis [added]; Sander, Thoman [added]]. Neurology. 2013;81(17):1507-1514.

Murai MJ, Sassonia RC, Zamboni AH, et al. Characterization of the C-terminal half of human juvenile myoclonic epilepsy protein EFHC1: dimer formation blocks Ca2+ and Mg2+ binding to its functional EF-hand. Arch Biochem Biophys. 2008;477(1):131-138.

Murphy T, Melville H, Fradkin E, et al. Knockdown of epigenetic transcriptional co-regulator Brd2a disrupts apoptosis and proper formation of hindbrain and midbrain-hindbrain boundary (MHB) region in zebrafish. Mech Dev. 2017;146:10-30.

Murthy JM, Rao CM, Meena AK. Clinical observations of juvenile myoclonic epilepsy in 131 patients: a study in South India. Seizure. 1998;7(1):43-47.

Nagasaki K, Tsuchiya S, Saitoh A, Ogata T, Fukami M. Neuromuscular symptoms in a patient with familial pseudohypoparathyroidism type Ib diagnosed by methylation-specific multiplex ligation-dependent probe amplification. Endocr J. 2013;60(2):231-236.

Nagase T, Ishikawa K, Suyama M, et al. Prediction of the coding sequences of unidentified human genes. XIII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. DNA Res. 1999;6(1):63-70.

Naseer MI, Faheem M, Chaudhary AG, et al. Genome wide analysis of novel copy number variations duplications/deletions of different epileptic patients in Saudi Arabia. BMC Genomics. 2015;16 Suppl 1(Suppl 1):S10.

Neto H, Kaupisch A, Collins LL, Gould GW. Syntaxin 16 is a master recruitment factor for cytokinesis. Mol Biol Cell. 2013;24(23):3663-3674.

Nicoletti A, Reggio A, Bartoloni A, et al. Prevalence of epilepsy in rural Bolivia: a door-to-door survey. Neurology. 1999;53(9):2064-2069.

Niestroj LM, Perez-Palma E, Howrigan DP, et al. Epilepsy subtype-specific copy number burden observed in a genome-wide study of 17 458 subjects. Brain. 2020;143(7):210

Nilo A, Gelisse P, Crespel A. Genetic/idiopathic generalized epilepsies: Not so good as that!. Rev Neurol (Paris). 2020;176(6):427-438.

Nissen PH, Rejnmark L. Expanding the spectrum of genetic variants in the calcium-sensing receptor (CASR) gene in hypercalcemic individuals. Clin Endocrinol (Oxf). 2019;91(5):683-690.

Niturad CE, Lev D, Kalscheuer VM, et al. Rare GABRA3 variants are associated with epileptic seizures, encephalopathy and dysmorphic features. Brain. 2017;140(11):2879-2894.

Nuyts S, D'Souza W, Bowden SC, Vogrin SJ. Structural brain abnormalities in genetic generalized epilepsies: A systematic review and meta-analysis. Epilepsia. 2017;58(12):2025-2037.

Ohmori I, Ouchida M, Miki T, et al. A CACNB4 mutation shows that altered Ca(v)2.1 function may be a genetic modifier of severe myoclonic epilepsy in infancy. Neurobiol Dis. 2008;32(3):349-354.

Okamoto N, Miya F, Tsunoda T, et al. Targeted next-generation sequencing in the diagnosis of neurodevelopmental disorders. Clin Genet. 2015;88(3):288-292.

Ooi A, Wong A, Esau L, Lemtiri-Chlieh F, Gehring C. A Guide to Transient Expression of Membrane Proteins in HEK-293 Cells for Functional Characterization. Front Physiol. 2016;7:300.

Opferman JT, Kothari A. Anti-apoptotic BCL-2 family members in development. Cell Death Differ. 2018;25(1):37-45.

Orenstein N, Goldberg-Stern H, Straussberg R, et al. A de novo GABRA2 missense mutation in severe early-onset epileptic encephalopathy with a choreiform movement disorder. Eur J Paediatr Neurol. 2018;22(3):516-524.

O'Roak BJ, Stessman HA, Boyle EA, et al. Recurrent de novo mutations implicate novel genes underlying simplex autism risk. Nat Commun. 2014;5:5595.

Osaki E, Nishina Y, Inazawa J, et al. Identification of a novel Sry-related gene and its germ cell-specific expression. Nucleic Acids Res. 1999;27(12):2503-2510.

Oud MM, Bonnard C, Mans DA, et al. A novel ICK mutation causes ciliary disruption and lethal endocrine-cerebro-osteodysplasia syndrome. Cilia. 2016;5:8.

Pais I, Hormuzdi SG, Monyer H, et al. Sharp wave-like activity in the hippocampus in vitro in mice lacking the gap junction protein connexin 36 [published correction appears in J Neurophysiol. 2003 Aug;90(2):1362]. J Neurophysiol. 2003;89(4):2046-2054.

Pal DK, Durner M, Klotz I, et al. Complex inheritance and parent-of-origin effect in juvenile myoclonic epilepsy. Brain Dev. 2006;28(2):92-98.

Pal DK, Evgrafov OV, Tabares P, Zhang F, Durner M, Greenberg DA. BRD2 (RING3) is a probable major susceptibility gene for common juvenile myoclonic epilepsy. Am J Hum Genet. 2003;73(2):261-270.

Panayiotopoulos CP, Obeid T, Tahan AR. Juvenile myoclonic epilepsy: a 5-year prospective study. Epilepsia. 1994;35(2):285-296.

Panayiotopoulos CP, Obeid T, Waheed G. Absences in juvenile myoclonic epilepsy: a clinical and video-electroencephalographic study. Ann Neurol. 1989;25(4):391-397.

Parihar R, Mishra R, Singh SK, Jayalakshmi S, Mehndiratta MM, Ganesh S. Association of the GRM4 gene variants with juvenile myoclonic epilepsy in an Indian population. J Genet. 2014;93(1):193-197.

Pathak S, Miller J, Morris EC, Stewart WCL, Greenberg DA. DNA methylation of the BRD2 promoter is associated with juvenile myoclonic epilepsy in Caucasians. Epilepsia. 2018;59(5):1011-1019.

Peña VN, Kohn TP, Herati AS. Genetic mutations contributing to non-obstructive azoospermia. Best Pract Res Clin Endocrinol Metab. 2020;34(6):101479.

Petit FG, Kervarrec C, Jamin SP, et al. Combining RNA and protein profiling data with network interactions identifies genes associated with spermatogenesis in mouse and human. Biol Reprod. 2015;92(3):71.

Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. Genic intolerance to functional variation and the interpretation of personal genomes. PLoS Genet. 2013;9(8):e1003709.

Phochanukul N, Russell S. No backbone but lots of Sox: Invertebrate Sox genes. Int J Biochem Cell Biol. 2010;42(3):453-464.

Pinto AM, Bianciardi L, Mencarelli MA, et al. Exome sequencing analysis in a pair of monozygotic twins re-evaluates the genetics behind their intellectual disability and reveals a CHD2 mutation. Brain Dev. 2016;38(6):590-596.

Pinto D, Louwaars S, Westland B, et al. Heterogeneity at the JME 6p11-12 locus: absence of mutations in the EFHC1 gene in linked Dutch families. Epilepsia. 2006;47(10):1743-1746.

Poisson A, Chatron N, Labalme A, et al. Chromatin remodeling dysfunction extends the etiological spectrum of schizophrenia: a case report. BMC Med Genet. 2020;21(1):10.

Portnoi MF, Dumargne MC, Rojo S, et al. Mutations involving the SRY-related gene SOX8 are associated with a spectrum of human reproductive anomalies. Hum Mol Genet. 2018;27(7):1228-1240.

Proctor KM, Miller SC, Bryant NJ, Gould GW. Syntaxin 16 controls the intracellular sequestration of GLUT4 in 3T3-L1 adipocytes. Biochem Biophys Res Commun. 2006;347(2):433-438.

Raju PK, Satishchandra P, Nayak S, Iyer V, Sinha S, Anand A. Microtubule-associated defects caused by EFHC1 mutations in juvenile myoclonic epilepsy. Hum Mutat. 2017;38(7):816-826.

Ramos VC, Vidal-Taboada J, Bergoñon S, et al. Characterisation and expression analysis of the WDR9 gene, located in the Down critical region-2 of the human chromosome 21. Biochim Biophys Acta. 2002;1577(3):377-383.

Ratnapriya R. Identification of genetic loci for human epilepsies [PhD thesis]. Bengaluru, India: Jawaharlal Nehru Centre for Advanced Scientific Research; 2009.

Rauch A, Wieczorek D, Graf E, et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet. 2012;380(9854):1674-1682.

Ray K, Fan GF, Goldsmith PK, Spiegel AM. The carboxyl terminus of the human calcium receptor. Requirements for cell-surface expression and signal transduction. J Biol Chem. 1997;272(50):31355-31361.

Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. Nature. 2006;444(7118):444-454.

Regala J, Cavaco B, Domingues R, Limbert C, Lopes L. Novel Mutation in the CASR Gene (p.Leu123Ser) in a Case of Autosomal Dominant Hypocalcemia. J Pediatr Genet. 2015;4(1):29-33.

Renganathan R, Delanty N. Juvenile myoclonic epilepsy: under-appreciated and under-diagnosed. Postgrad Med J. 2003;79(928):78-80.

Reyes-Nava NG, Yu HC, Coughlin CR 2nd, Shaikh TH, Quintana AM. Abnormal expression of GABAA receptor subunits and hypomotility upon loss of gabra1 in zebrafish. Biol Open. 2020;9(4):bio051367.

Rhee K, Brunori M, Besset V, Trousdale R, Wolgemuth DJ. Expression and potential role of Fsrg1, a murine bromodomain-containing homologue of the Drosophila gene female sterile homeotic. J Cell Sci. 1998;111 (Pt 23):3541-3550.

Rodríguez D, Bretones G, Quesada V, et al. Mutations in CHD2 cause defective association with active chromatin in chronic lymphocytic leukemia. Blood. 2015;126(2):195-202.

Röpke A, Tewes AC, Gromoll J, Kliesch S, Wieacker P, Tüttelmann F. Comprehensive sequence analysis of the NR5A1 gene encoding steroidogenic factor 1 in a large group of infertile males. Eur J Hum Genet. 2013;21(9):1012-1015.

Rossetto MG, Zanarella E, Orso G, et al. Defhc1.1, a homologue of the juvenile myoclonic gene EFHC1, modulates architecture and basal activity of the neuromuscular junction in Drosophila. Hum Mol Genet. 2011;20(21):4248-4257.

Rossi GC, Patterson AL, McGregor AL, Wheless JW. Intractable Generalized Epilepsy and Autosomal Dominant Hypocalcemia: A Case Report. Child Neurol Open. 2019;6:2329048X19876199.

Roumaud P, Haché J, Martin LJ. Expression profiles of Sox transcription factors within the postnatal rodent testes. Mol Cell Biochem. 2018;447(1-2):175-187.

Routier L, Verny F, Barcia G, et al. Exome sequencing findings in 27 patients with myoclonic-atonic epilepsy: Is there a major genetic factor? Clin Genet. 2019;96(3):254-260.

Rozycka A, Steinborn B, Trzeciak WH. The 1674+11C>T polymorphism of CHRNA4 is associated with juvenile myoclonic epilepsy. Seizure. 2009;18(8):601-603.

Ruat M, Molliver ME, Snowman AM, Snyder SH. Calcium sensing receptor: molecular cloning in rat and localization to nerve terminals. Proc Natl Acad Sci U S A. 1995;92(8):3161-3165.

Sakai N, Terami H, Suzuki S, et al. Identification of NR5A1 (SF-1/AD4BP) gene expression modulators by large-scale gain and loss of function studies. J Endocrinol. 2008;198(3):489-497.

Samarut É, Swaminathan A, Riché R, et al. γ-Aminobutyric acid receptor alpha 1 subunit loss of function causes genetic generalized epilepsy by impairing inhibitory network neurodevelopment. Epilepsia. 2018;59(11):2061-2074.

Sander T, Kretz R, Williamson MP, et al. Linkage analysis between idiopathic generalized epilepsies and the GABA(A) receptor alpha5, beta3 and gamma3 subunit gene cluster on chromosome 15. Acta Neurol Scand. 1997;96(1):1-7.

Sano S, Iwata H, Matsubara K, Fukami M, Kagami M, Ogata T. Growth hormone deficiency in monozygotic twins with autosomal dominant pseudohypoparathyroidism type Ib. Endocr J. 2015;62(6):523-529.

Sapio MR, Vessaz M, Thomas P, Genton P, Fricker LD, Salzmann A. Novel carboxypeptidase A6 (CPA6) mutations identified in patients with juvenile myoclonic and generalized epilepsy. PLoS One. 2015;10(4):e0123180.

Sarkar A, Hochedlinger K. The sox family of transcription factors: versatile regulators of stem and progenitor cell fate. Cell Stem Cell. 2013;12(1):15-30.

Sazgar M, Young MG. Seizures and Epilepsy. In: Absolute Epilepsy and EEG rotation review, 2019;pp. 9–46. Springer, New York

Sbrocchi AM, Rauch F, Lawson ML, et al. Osteosclerosis in two brothers with autosomal dominant pseudohypoparathyroidism type 1b: bone histomorphometric analysis. Eur J Endocrinol. 2011;164(2):295-301

Scheffer IE, Berkovic S, Capovilla G, et al. ILAE classification of the epilepsies: Position paper of the ILAE Commission for Classification and Terminology. Epilepsia. 2017;58(4):512-521.

Schepers GE, Teasdale RD, Koopman P. Twenty pairs of sox: extent, homology, and nomenclature of the mouse and human sox transcription factor gene families. Dev Cell. 2002;3(2):167-170.

Schulz H, Ruppert AK, Zara F, et al. No evidence for a BRD2 promoter hypermethylation in blood leukocytes of Europeans with juvenile myoclonic epilepsy. Epilepsia. 2019;60(5):e31-e36.

Seabra CM, Quental S, Lima AC, et al. The mutational spectrum of WT1 in male infertility. J Urol. 2015;193(5):1709-1715.

Serafini A, Gerard E, Genton P, Crespel A, Gelisse P. Treatment of Juvenile Myoclonic Epilepsy in Patients of Child-Bearing Potential. CNS Drugs. 2019;33(3):195-208.

Sesarini CV, Costa L, Naymark M, et al. Evidence for interaction between markers in GABA(A) receptor subunit genes in an Argentinean autism spectrum disorder population. Autism Res. 2014;7(1):162-166.

Shaikh TH. Copy Number Variation Disorders. Curr Genet Med Rep. 2017;5(4):183-190.

Shang E, Cui Q, Wang X, Beseler C, Greenberg DA, Wolgemuth DJ. The bromodomain-containing gene BRD2 is regulated at transcription, splicing, and translation levels. J Cell Biochem. 2011;112(10):2784-2793.

Shang E, Wang X, Wen D, Greenberg DA, Wolgemuth DJ. Double bromodomain-containing gene Brd2 is essential for embryonic development in mouse. Dev Dyn. 2009;238(4):908-917.

Shashi V, Schoch K, Spillmann R, et al. A comprehensive iterative approach is highly effective in diagnosing individuals who are exome negative. Genet Med. 2019;21(1):161-172.

She ZY, Yang WX. SOX family transcription factors involved in diverse cellular events during development. Eur J Cell Biol. 2015;94(12):547-563.

Shen D, Hernandez CC, Shen W, et al. De novo GABRG2 mutations associated with epileptic encephalopathies. Brain. 2017;140(1):49-67.

Shen T, Ji F, Yuan Z, Jiao J. CHD2 is Required for Embryonic Neurogenesis in the Developing Cerebral Cortex. Stem Cells. 2015;33(6):1794-1806.

Shorvon, S. D.-.-S. D., F. Andermann & R. Guerrini. 2010. The causes of epilepsy. Cambridge University Press

Silvennoinen K, de Lange N, Zagaglia S, et al. Comparative effectiveness of antiepileptic drugs in juvenile myoclonic epilepsy. Epilepsia Open. 2019;4(3):420-430.

Simon J, Wakimoto H, Fujita N, Lalande M, Barnard EA. Analysis of the set of GABA(A) receptor genes in the human genome. J Biol Chem. 2004;279(40):41422-41435.

Simonsen A, Bremnes B, Rønning E, Aasland R, Stenmark H. Syntaxin-16, a putative Golgi t-SNARE. Eur J Cell Biol. 1998;75(3):223-231.

Sinclair AH, Berta P, Palmer MS, et al. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature. 1990;346(6281):240-244.

Singh B, Monteil A, Bidaud I, et al. Mutational analysis of CACNA1G in idiopathic generalized epilepsy. Mutation in brief #962. Online. Hum Mutat. 2007;28(5):524-525.

Singh N, Ritaccio A. Juvenile myoclonic epilepsy mimic associated with CHD2 gene mutation [published correction appears in Epilepsy Behav Rep. 2020 Dec 08;14:100415]. Epilepsy Behav Rep. 2020;13:100355.

Şirinocak, P. B., B. Salman, F. Y. Kesim, N. Bebek, B. Baykan & S. A. U. İşeri. Susceptibility to Juvenile Myoclonic Epilepsy Associated with the EFHC1 Gene: First Case Report in Turkey. Türk Nöroloji Dergisi, 2019;25, 233-236.

Sisodiya SM, Ragge NK, Cavalleri GL, et al. Role of SOX2 mutations in human hippocampal malformations and epilepsy. Epilepsia. 2006;47(3):534-542.

Sleven H, Welsh SJ, Yu J, et al. De Novo Mutations in EBF3 Cause a Neurodevelopmental Syndrome. Am J Hum Genet. 2017;100(1):138-150.

Srivastava S, Cohen J, Pevsner J, et al. A novel variant in GABRB2 associated with intellectual disability and epilepsy. Am J Med Genet A. 2014;164A(11):2914-2921.

Steffens M, et al. EPICURE Consortium; EMINet Consortium, Genome-wide association analysis of genetic generalized epilepsies implicates susceptibility loci at 1q43, 2p16.1, 2q22.3 and 17q21.32. Hum Mol Genet. 2012;21(24):5359-5372.

Stepanchick A, McKenna J, McGovern O, Huang Y, Breitwieser GE. Calcium sensing receptor mutations implicated in pancreatitis and idiopathic epilepsy syndrome disrupt an arginine-rich retention motif. Cell Physiol Biochem. 2010;26(3):363-374.

Steudle F, Rehman S, Bampali K, et al. A novel de novo variant of GABRA1 causes increased sensitivity for GABA in vitro. Sci Rep. 2020;10(1):2379.

Stogmann E, Lichtner P, Baumgartner C, et al. Idiopathic generalized epilepsy phenotypes associated with different EFHC1 mutations. Neurology. 2006;67(11):2029-2031.

Strehlow V, Swinkels ME, Thomas RH, et al. Generalized Epilepsy and Myoclonic Seizures in 22q11.2 Deletion Syndrome. Mol Syndromol. 2016;7(4):239-246.

Stros M, Launholt D, Grasser KD. The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins. Cell Mol Life Sci. 2007;64(19-20):2590-2606.

Subaran RL, Conte JM, Stewart WC, Greenberg DA. Pathogenic EFHC1 mutations are tolerated in healthy individuals dependent on reported ancestry. Epilepsia. 2015;56(2):188-194.

Sudhakar DVS, Nizamuddin S, Manisha G, et al. NR5A1 mutations are not associated with male infertility in Indian men. Andrologia. 2018;50(3):10.1111/and.12931.

Suebsoonthron J, Jaroonwitchawan T, Yamabhai M, Noisa P. Inhibition of WNT signaling reduces differentiation and induces sensitivity to doxorubicin in human malignant neuroblastoma SH-SY5Y cells. Anticancer Drugs. 2017;28(5):469-479.

Suls A, Jaehn JA, Kecskés A, et al. De novo loss-of-function mutations in CHD2 cause a fever-sensitive myoclonic epileptic encephalopathy sharing features with Dravet syndrome. Am J Hum Genet. 2013;93(5):967-975.

Suzuki T, Delgado-Escueta AV, Aguan K, et al. Mutations in EFHC1 cause juvenile myoclonic epilepsy. Nat Genet. 2004;36(8):842-849.

Suzuki T, Inoue I, Yamagata T, Morita N, Furuichi T, Yamakawa K. Sequential expression of Efhc1/myoclonin1 in choroid plexus and ependymal cell cilia. Biochem Biophys Res Commun. 2008;367(1):226-233.

Suzuki T, Miyamoto H, Nakahari T, et al. Efhc1 deficiency causes spontaneous myoclonus and increased seizure susceptibility. Hum Mol Genet. 2009;18(6):1099-1109.

Tadmouri A, Kiyonaka S, Barbado M, et al. Cacnb4 directly couples electrical activity to gene expression, a process defective in juvenile epilepsy. EMBO J. 2012;31(18):3730-3744.

Tanaka AJ, Cho MT, Willaert R, et al. De novo variants in EBF3 are associated with hypotonia, developmental delay, intellectual disability, and autism. Cold Spring Harb Mol Case Stud. 2017;3(6):a002097.

Tang BL, Low DY, Lee SS, Tan AE, Hong W. Molecular cloning and localization of human syntaxin 16, a member of the syntaxin family of SNARE proteins. Biochem Biophys Res Commun. 1998;242(3):673-679.

Tang Y, Li X, Xiao H, et al. Transcription of the Sox30 Gene Is Positively Regulated by Dmrt1 in Nile Tilapia. Int J Mol Sci. 2019;20(21):5487.

Taniguchi Y. The Bromodomain and Extra-Terminal Domain (BET) Family: Functional Anatomy of BET Paralogous Proteins. Int J Mol Sci. 2016;17(11):1849.

Thakker RV. Diseases associated with the extracellular calcium-sensing receptor. Cell Calcium. 2004;35(3):275-282.

Thijs RD, Surges R, O'Brien TJ, Sander JW. Epilepsy in adults. Lancet. 2019;393(10172):689-701.

Thomas RH, Zhang LM, Carvill GL, et al. CHD2 myoclonic encephalopathy is frequently associated with self-induced seizures. Neurology. 2015;84(9):951-958.

Thounaojam R, Langbang L, Itisham K, et al. EFHC1 mutation in Indian juvenile myoclonic epilepsy patient. Epilepsia Open. 2017;2(1):84-89.

Togawa K, Yan YX, Inomoto T, Slaugenhaupt S, Rustgi AK. Intestinal cell kinase (ICK) localizes to the crypt region and requires a dual phosphorylation site found in map kinases. J Cell Physiol. 2000;183(1):129-139.

Tolchin D, Yeager JP, Prasad P, et al. De Novo SOX6 Variants Cause a Neurodevelopmental Syndrome Associated with ADHD, Craniosynostosis, and Osteochondromas. Am J Hum Genet. 2020;106(6):830-845.

Trivisano M, Striano P, Sartorelli J, et al. CHD2 mutations are a rare cause of generalized epilepsy with myoclonic-atonic seizures. Epilepsy Behav. 2015;51:53-56.

Trousdale RK, Wolgemuth DJ. Bromodomain containing 2 (Brd2) is expressed in distinct patterns during ovarian folliculogenesis independent of FSH or GDF9 action. Mol Reprod Dev. 2004;68(3):261-268.

Tu Q, Pi M, Karsenty G, Simpson L, Liu S, Quarles LD. Rescue of the skeletal phenotype in CasR-deficient mice by transfer onto the Gcm2 null background. J Clin Invest. 2003;111(7):1029-1037.

Turan S, Ignatius J, Moilanen JS, et al. De novo STX16 deletions: an infrequent cause of pseudohypoparathyroidism type Ib that should be excluded in sporadic cases. J Clin Endocrinol Metab. 2012;97(12):E2314-E2319.

Usui N, Kotagal P, Matsumoto R, Kellinghaus C, Lüders HO. Focal semiologic and electroencephalographic features in patients with juvenile myoclonic epilepsy. Epilepsia. 2005;46(10):1668-1676.

Vadlamudi L, Milne RL, Lawrence K, et al. Genetics of epilepsy: The testimony of twins in the molecular era. Neurology. 2014;83(12):1042-1048.

Vahe C, Benomar K, Espiard S, et al. Diseases associated with calcium-sensing receptor. Orphanet J Rare Dis. 2017;12(1):19.

Velíšek L, Shang E, Velíšková J, et al. GABAergic neuron deficit as an idiopathic generalized epilepsy mechanism: the role of BRD2 haploinsufficiency in juvenile myoclonic epilepsy. PLoS One. 2011;6(8):e23656.

Veredice C, Bianco F, Contaldo I, et al. Early onset myoclonic epilepsy and 15q26 microdeletion: observation of the first case. Epilepsia. 2009;50(7):1810-1815.

Vijai J, Cherian PJ, Stlaja PN, Anand A, Radhakrishnan K. Clinical characteristics of a South Indian cohort of juvenile myoclonic epilepsy probands. Seizure. 2003a;12(7):490-496.

Vijai J, Kapoor A, Ravishankar HM, et al. Genetic association analysis of KCNQ3 and juvenile myoclonic epilepsy in a South Indian population. Hum Genet. 2003b;113(5):461-463.

Vizard TN, O'Keeffe GW, Gutierrez H, Kos CH, Riccardi D, Davies AM. Regulation of axonal and dendritic growth by the extracellular calcium-sensing receptor. Nat Neurosci. 2008;11(3):285-291.

Vysotskaya ZV, Moss CR 2nd, Gilbert CA, Gabriel SA, Gu Q. Modulation of BK channel activities by calcium-sensing receptor in rat bronchopulmonary sensory neurons. Respir Physiol Neurobiol. 2014;203:35-44.

Wallace RH, Marini C, Petrou S, et al. Mutant GABA(A) receptor gamma2-subunit in childhood absence epilepsy and febrile seizures. Nat Genet. 2001;28(1):49-52.

Wang EJ, Gailey CD, Brautigan DL, Fu Z. Functional Alterations in Ciliogenesis-Associated Kinase 1 (CILK1) that Result from Mutations Linked to Juvenile Myoclonic Epilepsy. Cells. 2020;9(3):694.

Wang M, Greenberg DA, Stewart WCL. Replication, reanalysis, and gene expression: ME2 and genetic generalized epilepsy. Epilepsia. 2019;60(3):539-546.

Wang MM, Reed RR. Molecular cloning of the olfactory neuronal transcription factor Olf-1 by genetic selection in yeast. Nature. 1993;364(6433):121-126.

Wang S, Wang X, Ma L, et al. Retinoic Acid Is Sufficient for the In Vitro Induction of Mouse Spermatocytes. Stem Cell Reports. 2016;7(1):80-94.

Wang XN, Li ZS, Ren Y, et al. The Wilms tumor gene, Wt1, is critical for mouse spermatogenesis via regulation of sertoli cell polarity and is associated with non-obstructive azoospermia in humans. PLoS Genet. 2013;9(8):e1003645.

Wang Y, Du X, Bin R, et al. Genetic Variants Identified from Epilepsy of Unknown Etiology in Chinese Children by Targeted Exome Sequencing [published correction appears in Sci Rep. 2017 May 04;7:46520]. Sci Rep. 2017;7:40319.

Ward DT. Calcium receptor-mediated intracellular signalling. Cell Calcium. 2004;35(3):217-228.

Wegner M. All purpose Sox: The many roles of Sox proteins in gene expression. Int J Biochem Cell Biol. 2010;42(3):381-390.

Whissell PD, Rosenzweig S, Lecker I, Wang DS, Wojtowicz JM, Orser BA. γ-aminobutyric acid type A receptors that contain the δ subunit promote memory and neurogenesis in the dentate gyrus. Ann Neurol. 2013;74(4):611-621.

Wiel L, Baakman C, Gilissen D, Veltman JA, Vriend G, Gilissen C. MetaDome: Pathogenicity analysis of genetic variants through aggregation of homologous human protein domains. Hum Mutat. 2019;40(8):1030-1038.

Williams AJ, Khachigian LM, Shows T, Collins T. Isolation and characterization of a novel zinc-finger protein with transcription repressor activity. J Biol Chem. 1995;270(38):22143-22152.

Windpassinger C, Kroisel PM, Wagner K, Petek E. The human gamma-aminobutyric acid A receptor delta (GABRD) gene: molecular characterisation and tissue-specific expression. Gene. 2002;292(1-2):25-31.

Wirrell EC, Camfield CS, Camfield PR, Gordon KE, Dooley JM. Long-term prognosis of typical childhood absence epilepsy: remission or progression to juvenile myoclonic epilepsy. Neurology. 1996;47(4):912-918.

Wolf P, Yacubian EM, Avanzini G, et al. Juvenile myoclonic epilepsy: A system disorder of the brain. Epilepsy Res. 2015;114:2-12.

Xicoy H, Wieringa B, Martens GJ. The SH-SY5Y cell line in Parkinson's disease research: a systematic review. Mol Neurodegener. 2017;12(1):10.

Xie HR, Hu LS, Li GY. SH-SY5Y human neuroblastoma cell line: in vitro cell model of dopaminergic neurons in Parkinson's disease. Chin Med J (Engl). 2010;123(8):1086-1092.

Xu J, Jiang L, Yu W, et al. A novel functional variant in Wilms' Tumor 1 (WT1) is associated with idiopathic non-obstructive azoospermia. Mol Reprod Dev. 2017;84(3):222-228.

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

Yacubian EM, Wolf P. Praxis induction. Definition, relation to epilepsy syndromes, nosological and prognostic significance. A focused review. Seizure. 2014;23(4):247-251.

Yacubian EM. Juvenile myoclonic epilepsy: Challenges on its 60th anniversary. Seizure. 2017;44:48-52.

Yang T, Jiang Y, Chen J. The identification and subcellular localization of human MRK. Biomol Eng. 2002;19(1):1-4.

Yavuz EN, Ozdemir O, Catal S, Bebek N, Ozbek U, Baykan B. Bromodomain-containing protein 2 gene in photosensitive epilepsy. Seizure. 2012;21(8):646-648.

Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics. 2015;31(14):2382-2383.

Zawerton A, Mignot C, Sigafoos A, et al. Widening of the genetic and clinical spectrum of Lamb-Shaffer syndrome, a neurodevelopmental disorder due to SOX5 haploinsufficiency. Genet Med. 2020;22(3):524-537.

Zeng T, Dong ZF, Liu SJ, et al. A novel variant in the 3' UTR of human SCN1A gene from a patient with Dravet syndrome decreases mRNA stability mediated by GAPDH's binding. Hum Genet. 2014;133(6):801-811.

Zhang D, Xie D, Lin X, et al. The transcription factor SOX30 is a key regulator of mouse spermiogenesis. Development. 2018;145(11):dev164723.

Zhang X, Ning Y, Liu W, et al. Variation analysis of SOX8 gene in Chinese men with non-obstructive azoospermia or oligozoospermia. Andrologia. 2020;52(4):e13531.

Zhang Y, Liu T, Meyer CA, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137.

Zhao Y, Shi J, Winey M, Klymkowsky MW. Identifying domains of EFHC1 involved in ciliary localization, ciliogenesis, and the regulation of Wnt signaling. Dev Biol. 2016;411(2):257-265.

Zheng QS, Wang XN, Wen Q, et al. Wt1 deficiency causes undifferentiated spermatogonia accumulation and meiotic progression disruption in neonatal mice. Reproduction. 2013;147(1):45-52.

Zhou C, Huang Z, Ding L, et al. Altered cortical GABAA receptor composition, physiology, and endocytosis in a mouse model of a human genetic absence epilepsy syndrome. J Biol Chem. 2013;288(29):21458-21472.

Zorrilla M, Yatsenko AN. The Genetics of Infertility: Current Status of the Field. Curr Genet Med Rep. 2013;1(4):10.1007/s40142-013-0027-1.

Zou F, McWalter K, Schmidt L, et al. Expanding the phenotypic spectrum of GABRG2 variants: a recurrent GABRG2 missense variant associated with a severe phenotype. J Neurogenet. 2017;31(1-2):30-36.

Web Resources

1000 Genomes Project: https://www.internationalgenome.org/home

Align-GVGD: http://agvgd.hci.utah.edu/agvgd_input.php

Allen brain atlas: https://portal.brain-map.org/

BioGRID[4.4]: https://thebiogrid.org/

BrainStars: http://brainstars.org/

BrainTx: https://www.cdtdb.neuroinf.jp/CDT/Top.jsp

CADD: https://cadd.gs.washington.edu/

Clustal Omega: https://www.ebi.ac.uk/Tools/msa/clustalo/

ConSurf Server: https://consurf.tau.ac.il/

dbSNP: https://www.ncbi.nlm.nih.gov/snp/

Envision: https://envision.gs.washington.edu/shiny/envision_new/

Epi25K: https://epi25.broadinstitute.org/

Exome Variant Server (EVS): https://evs.gs.washington.edu/EVS/

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

Genome Asia 100k: https://genomeasia100k.org/

GnomAD: https://gnomad.broadinstitute.org/

HPA Mouse brain RNA-Seq: https://www.proteinatlas.org/

Human Splicing Fimder: https://www.genomnis.com/access-hsf

Index-DB: http://indexdb.ncbs.res.in/

IndiGenomes: http://clingen.igib.res.in/indigen/index.php

IntAct: https://www.ebi.ac.uk/intact/home

MaxEntScan: http://hollywood.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html

MINT: https://mint.bio.uniroma2.it/

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

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

MutPred2: http://mutpred.mutdb.org/

NCBI Gene: https://www.ncbi.nlm.nih.gov/gene/

OligoCalc: http://biotools.nubic.northwestern.edu/OligoCalc.html

Online Mendelian Inheritance in Man: https://www.omim.org/

PANTHER-PSEP: http://pantherdb.org/tools/csnpScoreForm.jsp

PhD-SNP: https://snps.biofold.org/phd-snp/phd-snp.html

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

Primer3: https://bioinfo.ut.ee/primer3/

PROVEAN: http://provean.jcvi.org/index.php

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

SNAP2: https://rostlab.org/services/snap2web/

SNPs&GO: https://snps-and-go.biocomp.unibo.it/snps-and-go/

STRING: https://string-db.org/

TOPMED: https://bravo.sph.umich.edu/freeze8/hg38/

UCSC Genome Browser: https://genome.ucsc.edu/

UMD-Predictor: http://umd-predictor.eu/