Identification and validation of putative pathogenic variants in a patient with intellectual disability, epilepsy, autism and macrocephaly

Negar Zohoorian

Master Thesis

Department of Biosciences

Faculty of Mathematics and Natural Sciences

UNIVERSITY OF OSLO

June 2019

2019

Identification and validation of putative pathogenic variants in a patient with intellectual disability, epilepsy, autism and macrocephaly

http://www.duo.uio.no

Print: Reprosentralen, University of Oslo

Acknowledgment

This thesis is conducted at the department of Medical Genetics, Oslo University Hospital (OUS), Norway, June 2018 to June 2019.

I would like to first thank my supervisor Eirik Frengen whose expertise was invaluable in the formulating the whole project, and for his supports and continuous feedbacks especially during the writing process.

I would like to thank my internal supervisor Kristian Prydz. I would also like to acknowledge Doriana Misceo for her guidance during the whole project. You provided me with the tools that I needed to choose the right directions. Asbjørn Holmgren deserves special thanks for teaching me practical lab works during the first months. I would also like to thank all members of the Frengen group; Samudita, Dulika, Thilini, Ingunn for your helps. I am also thankful to Lene Wierød for her assistance in using fluorescence microscopy, and Marit was very helpful with all technical and practical matters relating to the lab.

Finally, I must express my very profound gratitude to my family, friends specially Fatemeh and to my boyfriend Arash for providing me with unfailing support and continuous encouragement throughout my study. This accomplishment would not have been possible without them. Thank you.

Oslo, June 2019
University of Oslo, Negar Zohoorian

Abstract

Whole exome sequencing was performed to identify putative disease-causing variants in a girl with macrocephaly, epilepsy and severe mental retardation, who had an unknown molecular diagnosis. Compound heterozygous *SZT2* variants were detected in the girl: The maternally inherited *SZT2* Chr1: g.43902997 A>T splice site variant was shown to cause skipping of the exon 42, and the missense variant Chr1: g.43905659 G>A NP_056099.3p: Asp2327Asn was paternally inherited. Previously 13 patients have been described with mutations in the *SZT2* gene with phenotypes overlapping with the patient in this study. The SZT2 protein is a component of the KICSTOR complex which has an important role in the inactivation of MTORC1. Under absence of nutrients, MTORC1 will be activated and translocated to the lysosomal membrane. Deficiency in the SZT2 protein may result in the constitutive localization of an active MTORC1 on the lysosomal membrane. In order to establish tools to study the impact of the SZT2 variants on the localization of MTOR in patient cells, specific siRNAs and antibodies were assessed for functionality. The MTOR and SZT2 mRNA expression level was shown to be reduced in HEK293T cells transfected with the MTOR or SZT2 siRNAs, and Western blotting detected reduction in the MTOR protein level, while with the SZT2 antibody gave no specific bands. Therefore, patient fibroblasts were studied only with the MTOR antibody using immunofluorescence staining to assess the cellular localization of the MTOR protein in the presence and absence of amino acids. Transcriptome analyses were also performed on mRNA extracted from fibroblast patient and her parents to identify differentially expressed genes, and pathway enrichments investigation was carried out to retrieve enriched pathways in the data from patient cells. These analyses indicated down regulation of *EREG, FGF7* and *SHC3* genes which play a role in the transduction of cell signaling leading to the RAS/RAF/MEK/ERK pathway activation, which is up stream of MTORC1.

We identified compound heterozygous *SZT2* variants Chr1: g.43902997 A>T, and Chr1: g.43905659 G>A. Further transcriptome analyses with samples from additional patients are required to reveal details on the molecular disease mechanisms. Furthermore, the MTOR antibody has been validated for future study on the localization of MTOR in nutrient deprived cells.

Table of Contents

1 Introduction

1.1 The human genome

The human genome is defined as a complete DNA content in the 46 chromosomes within a cell nucleus and DNA of mitochondria (1). Protein-coding sequences in the human genome represent approximately 1.5% of the human genome(2), and almost 50% of the total DNA consists of **unique** or **single-copy DNA**, which are specific sequences of nucleotides present in the genome only once. The other half of the human genome is **repetitive DNA,** which is composed of various groups of repeated nucleotide sequences (Figure1.1). Approximately 20,000 protein-coding genes are identified in the human genome(3), and a large portion of the protein-coding genes are single-copy DNA sequences. Repetitive DNA contributes to sustaining chromosome structure, and it is a main source of individuals' variations (4).

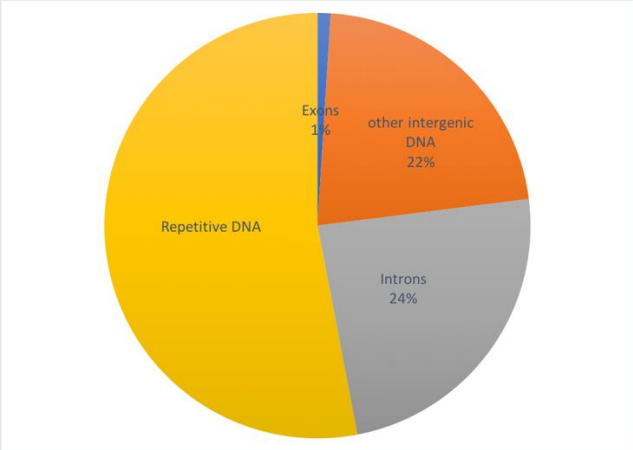


Figure 1.1 The human genome distribution of coding and non-coding elements in percentage (data from(5))

The repetitive DNA in the human genome is generally divided into two categories: **tandem repeats** and **interspersed repeats** (4). Tandem repeats consist of DNA sequence copies, which are repeated adjacently, that can be **direct tandem repeats** (in the same direction) or **inverted tandem repeats** (in the opposite orientation). They consist of : i) **Satellites,** the high copy-number tandem repeats (6), which mainly are present within heterochromatic regions and play a role in centromere structure (7), ii) **Minisatellites** that are variable-number tandem repeats (VNTRs) (8) usually containing motifs of 6-100 nucleotides(9), and iii) **Microsatellites** that are simple sequence repeats (SSRs) (8) in which a short segment of DNA (1-5 nucleotides) is being repeated(9).The interspersed repeats are

copies of DNA sequences that repeated non-adjacently. These are specific sequence elements that can be subdivided into the two main groups according to their size (10): i) Short interspersed elements (SINEs), which are 100-400(bp) in size(11), e.g. the **Alu** elements(12), and ii) Long interspersed nuclear elements (LINEs). The most abundant LINE family in the human genome is **LINE-1,** which can be up to 6kb(13).

1.1.1 The human reference genome

The human reference genome is the sequence of human DNA, which is a more complete version of the draft sequence published by the Human Genome Project in 2001(3). The Human Genome Project was an international collaboration aiming at sequencing the whole human genome with high coverage and accuracy, which has led to the sequence of almost the entire euchromatic genome (Build 35) with 2.85 billion nucleotides, which interrupted by just 341 gaps. The error rate is approximately 1 per 100,000 bp. Moreover, the human genome was estimated to be 3.08Gb (3). The reference genome is a haploid sequence compiled from DNA obtained from several individuals.

1.1.2 Variations in a genome

The genome sequence is approximately 99.5% identical between two randomly selected individuals(4). Thus, the remaining 0.5% contributes in the differences between each individual. Two genome sequences can differ in many ways.

1.1.2.1 Chromosomal variations

Variations in the human chromosomes can be broadly divided into two groups: numerical and structural. **Numerical variations**: Alterations in the numbers of chromosomes, which may result from mis-segregation of chromosomes in meiosis, lead to aneuploidy (Figure1.2)(14). Having one or two extra copies of a chromosomes is called trisomy or tetrasomy, respectively, while missing a chromosome from a pair is called monosomy. Numerical abnormalities can for instance cause Down syndrome, which is usually caused by trisomy 21.

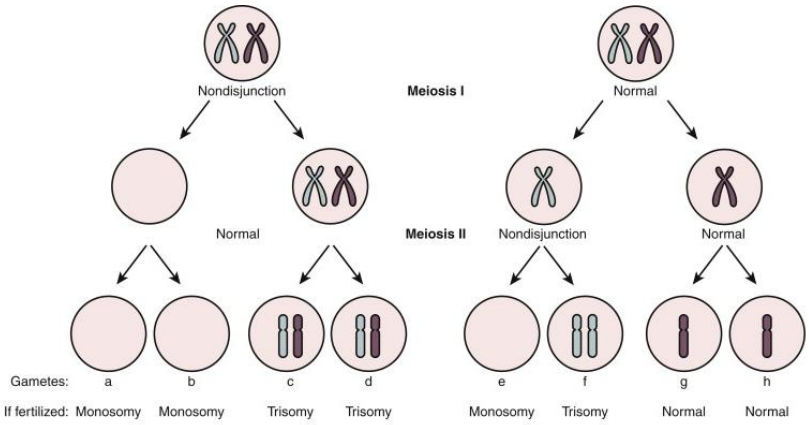


Figure 1.2 nondisjunction during meiosis leads to an aneuploidy. Nondisjunction in meiosis I results in two gametes with an extra chromosome, and the two with no copies of the mis segregated chromosome. While in meiosis II, nondisjunction leads to two normal gametes, one with an extra chromosome and one with one chromosome missing (figure from (https://www.sciencedirect.com/topics/medicine-and-dentistry/nondisjunction)).

Structural variations (SV), Structural variants are defined as alterations in the structure of chromosomes. The length of SV can vary between 50 (bp) and many Mega base pairs (Mb). SVs can be categorized in three main groups consist of copy number variants (CNVs), inversions and translocations. **Copy number variants (CNVs)** are DNA segments which are at least 50 bp and their copy numbers are variable between individuals in the population (15), and contain insertions, deletions and duplications. An **inversion** is a DNA segment which has a reversed orientation in one chromosome compared to its homolog. The inverted region can include the centromere, which is then called a pericentric inversion, or without a centromere called paracentric inversion. In **translocations** a chromosomal segment is moved to another position in the genome in which the DNA content may remain intact(16) (Figure1.3).

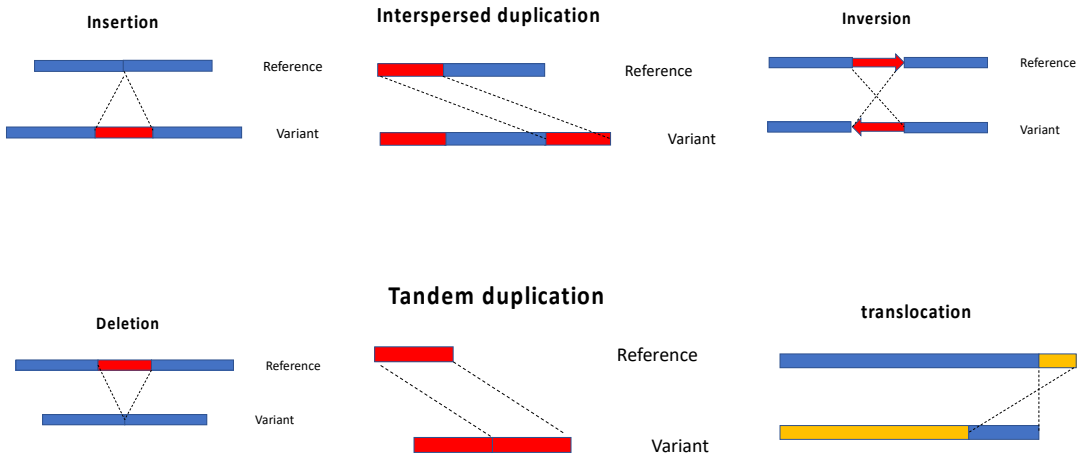


Figure 1.3 Schematic illustration of classes of structural variants.

1.1.2.2 Small insertions and deletions

There are several micro deletions and insertions (indel) less than 50 bp, which have been recognized in the human populations (17, 18). Indels which are located in the regulatory regions or protein-coding regions, may have functional consequences. For example, an indel which is not multiple of 3 bp located within a protein-coding region causes frameshift. Consequently, it may lead to either the formation of premature stop-codon resulting in mRNA degradation or translation of mRNA to a truncated protein which may not be functional. On the other hand, indels in coding regions which are multiple of 3 bp, lead to in-frame changes, nevertheless it may result in alteration of the resulting protein through adding or removal of one or more amino acids. As a result, the produced protein may be functional, or show altered functionality.

1.1.2.3 Single nucleotide variations (SNV)

SNV is a single nucleotide substitution in the particular genome position (19). SNVs are located both in coding and non-coding regions. In the protein coding regions SNVs have two types of affects; i) **Synonymous substitution**, due to the degeneracy of codons some substitutions in exons do not result in amino acid changes. ii) **non-synonymous substitution**, SNVs that will result in nonsense or missense codon. Missense means that the alteration results in an amino acid change, while nonsense substitutions lead to a stop codon. SNVs in intronic regions may have functional effects such as causing defects in splicing sites which can further lead to aberrant splicing and potentially a truncated protein.

1.1.2.4 Repeat variations

A large portion of the genome consists of repeated sequences and the number of repeats can vary in the population. Some repeat expansions are known to cause particular diseases, e.g. the CAG-repeat expansion causing Huntington disease (OMIM 143100) (19).

1.1.3 Genetic polymorphism

Cavalli-Sforza & Bodmer defined genetic polymorphism as “the occurrence in the same population of two or more alleles at one locus, each with appreciable frequency" (20), and the frequency is defined as 1% or above. Two major groups of genetic polymorphisms comprise of single nucleotide polymorphisms (SNVs) and copy number polymorphisms (CNPs).

Single nucleotide polymorphism (SNP) is a single nucleotide variation, which occurs at particular position in the genome with the frequency of >1% in a population. SNPs are responsible for the largest

number of variants in the genome, and they can be located in coding or non-coding regions of the genome.

In **Copy number polymorphisms (CNPs),** the CNV alleles have a frequency of at least 1% in a population (16).

1.2 Types of genetic disorders and mode of inheritance

According to the number of genes involved, genetic disorders can be sorted into two classifications; monogenic disorders and multigenic disorders. i) **Monogenic disorders** in which dysfunction of a single-gene causes the disorder. Dysfunction can be as a result of alteration within the gene itself or in the gene's regulatory elements. According to the position of the gene on sex chromosomes or autosomal chromosomes, potential transmission patterns includes autosomal dominant, autosomal recessive, X-linked dominant, X-linked recessive, and Y-linked (21). Disorders caused by genetic variants in mitochondrial genes have a different inheritance pattern. Since a child only obtains mitochondria from a mother, mitochondrial disease-causing variants are maternally inherited(19). ii) In **Multigenic disorders**, more than one gene is involved and each abnormal gene have a contribution in the disease manifestation(22).

Genetic variants can have various consequences for the resulting protein. i) **Loss of function variants (LoF)** which result in no protein, a protein with no or reduced function. LoF variants are often the cause in recessive disorders. In addition, **Haploinsufficient** genes in which both alleles are vital for a normal phenotype. LoF in one allele in such genes, for example by a deletion, may cause a disorder because of a reduction in the production of a protein. On the other hand, ii) **Gain of function variants (GoF)** usually result in production of proteins with an increased function or expression. GoF variants usually cause dominant disorders.

1.3 Multifactorial disorders

There are some disorders that tend to cluster in families, but without a distinct inheritance pattern. These disorders may be caused by several genetic and environmental factors. Some examples are congenital malformations such as congenital heart defects, schizophrenia and autism(5).

1.4 Identification of genetic variations

Two different approaches used for identification of genetic variants are hypothesis-free and hypothesis-driven methods. For example, when the patient's phenotype points to a particular diagnosis, a hypothesis-driven method such as Sanger sequencing may be used to investigate one specific locus. While, in hypothesis-free approaches the whole genome of an individual may be explored for the identification of genetic variation. High throughput sequencing (HTS) is a powerful hypothesis-free approach. Although, in some cases HTS is performed using gene panels, which is a hypothesis-driven method often used in diagnostics. In such cases, HTS are restricted to analysis of genes in a gene panel, which is a list of genes associated with a specific phenotype(23).

High throughput sequencing (HTS) allows sequencing the whole genome or whole exome with high accuracy by massively parallel sequencing through fragmenting the DNA and allowing the machine to read the sequence of a large number of fragments at the same time. Therefore, the whole genome may be sequenced but the number of reads for each position can vary. Two main applications of HTS commonly used in medical genetics include **whole genome sequencing (WGS)** and **whole exome sequencing (WES)**(24). WES is a powerful technique to identify the disease-causing variants in a patient with an unknown diagnosis, having complex phenotype, or in multigenic diseases(25-27). Moreover, it has been a successful method for identifying disease-causing indels and point mutations in novel disease-causing genes(28). WES may be performed on family trios, most often through sequencing of the exomes from one patient and both parents, or by analyzing two affected siblings and the mother (inverted trio). This method has been successful for the identification of disease causing variants(29-31), particularly for detection of *de novo* variants(32, 33). WGS has been shown to be more efficient in detection of protein-coding variants compared to WES(34, 35), and it can also identify structural variants such as copy number variants and non-coding variants such as those, which are located in regulatory elements (36, 37).

1.5 Functional validation

By high throughput sequencing technologies, one may identify novel potential disease-causing variants contributing to a human disease. However, in order to document causality of the variant in relation to a patient's phenotype, further investigations are required such as functional validation assays which are used to demonstrate an association between a genetic variant and measurable effects on a cellular phenotype. Examples of functional validation assays in a molecular level can be,

i) studying the intracellular localization of a protein using antibodies tagged by immunofluorescence colors to analyze the effect of a variant on the localization of a specific protein, or ii) exploring a genetic variation's impact on the mRNA expression using genome wide (total mRNA sequencing) or targeted (quantitative real-time PCR) mRNA expression studies on a transcript level.

1.6 Neurological disorders

Neurological disorders can affect the central nervous system (CNS), peripheral nervous system, or even both. These can have various causes including environmental factors such as exposure to radiations, toxins, or drugs, un-normal pregnancy or delivery, as well as genetic defects. Disorders which have a childhood onset may manifest from birth to late childhood(38). Two main categories of pediatric neurological disorders according to the patients' phenotypes are: neurodevelopmental and neurodegenerative disorders. **Neurodevelopmental disorders** perturb brain development, which typically have a childhood onset(39) such as intellectual disability (ID), autism spectrum disorder, attention deficit hyperactivity disorder (ADHD), and epilepsy. The prevalence of ID, ASD, ADHD and epilepsy is 1%, 1%, 3.4% and 0.5-1% , respectively(40-43). **Neurodegenerative disorders,** in these disorders the progressive decrease of neurons typically leads to disturbing in a nervous system through loss of motor or cognitive skills, or increased sensory deficiencies(44). Pediatric movement disorders usually are categorized in neurodegenerative disorders which are defined by abnormalities in spontaneous movements(45).

1.7 Background

This thesis focuses on genetic analysis of a girl with macrocephaly, epilepsy and severe mental retardation, who had an unknown molecular diagnosis (46). She was born after a normal pregnancy with a birth weight of 3600 g and occipitofrontal circumference (OFC) 36 cm (75^{th} centile). She had an increase in her head size during her first six months and her OFC was 2-3 cm above the 97.5^{th} percentile after this. She developed psychomotor delay, epilepsy, and autistic features. Developmental delay was observed as delayed milestones such as walking at 28 months and she was not capable to speak

consistently at 14 years of age. Generalized complex seizures were manifested with nocturnal apnoea, myoclonus, gurgling sounds and in increased salivation. At age of 14 she had a flat and monotypic mood with an autistic-like behavior, lack of contact and her IQ was estimated between 20 and 35.

The younger sister of the patient seemed to have similar clinical features as her sister (Figure 1.4). She was born after an uneventful pregnancy with apparently normal birth parameters. During her first six months she developed macrocephaly, she walked at the age of 24 months and she showed the same autistic behavior as her sister. Nocturnal seizures were predominant which frequently occurred at the time of awaking or falling asleep. She had a large forehead and a short philtrum. In a family pedigree in Figure 1.4 one of the filled circles belongs to her. She died at 5 years old, possibly by a nocturnal epileptic attack. A post-mortem examination showed a histologically and structurally normal brain and spinal cord.

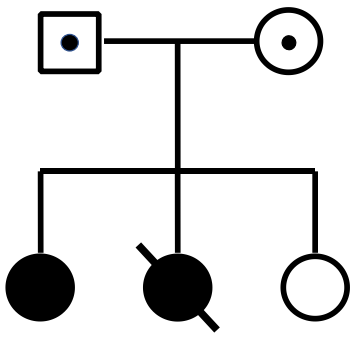


Figure 1.4 Family pedigree, affected individuals are displayed with filled symbols. The "•" sign shows the assumed carrier status of the parents. The "\" sign indicates the deceased sibling.

The unrelated parents had an additional younger healthy daughter. Because the parents were healthy, the main hypothesis is that the disorder in two older sisters were recessively inherited. However, it cannot be excluded that the disorder was caused by a dominant allele if one of the parents are mosaic. This thesis aims to identify potential disease-causing variants in the patient, and further to verify and assess their functional impact.

1.8 Aim of the thesis

The aim of this project was to identify the causative variant in the patient with intellectual disability (ID), epilepsy, autism, macrocephaly, regression in a girl with one affected sister, one un-affected sister and healthy un-consanguineous parents using WES. Moreover, we aimed to study the functional effect of the putative pathogenic variants in patient cells.

2 Materials and Methods

2.1 Cell culture

Two cell types were used in the present project: 1) The Human embryonic kidney 293 (HEK293T) cell line which is efficiently transfected with conventional techniques. 2) Fibroblast cells from the patient, her mother and father and an additional healthy control.

2.2 Primer design

Primers were designed using the online tool Primer3 (http://bioinfo.ut.ee/primer3-0.4.0/) and ordered from Eurofins Genomics (Ebersberg, Germany). Default settings were used except for; product size range (bp): 301-600, primer Length (bp): 20-24, and primer melting temperature (Tm): 62-68. In order to avoid potential hybridization on secondary loci in the genome, the designed primers were analyzed using BLAT (http://genome.ucsc.edu/cgi-bin/hgBlat) and In-Silico PCR (http://genome.ucsc.edu/cgi-bin/hgPcr) using GRCh37/hg19 reference genome (Table2.1). M13 sequences were incorporated on 5' end of some of the primers as it is described in section 2.5, to facilitate efficient sequencing of all PCR products.

Table 2.1 Primers used in the present study. M13 tails are indicated in small letters.

Name	**Sequence (5'→ 3')**	**Position in the human reference genome (GRCh37/hg19)**
SZT2-ex40-41-F	tgtaaaacgacggccagtGTCAACCAGCGACTGCTTCTTC	Chr1:43,900,148-43,900,883
SZT2-ex43-R	caggaaacagctatgaccGTACACAGCCTTTGTTGCTCGC	Chr1:43,903,150-43,903,171
GAPDH-cDNA-1F	AGGTCGGAGTCAACGGATTTG	Chr12:6,644,012-6,645,664
GAPDH-cDNA-1R	TGTAAACCATGTAGTTGAGGTCA	Chr12:6,645,744-6,645,856
PPIB-1F	CCAGGGCGGAGACTTCACCAGG	Chr15:64,452,340-64,452,319
PPIB-1R	CTCACCCAGCCAGGCCCGTA	Chr15:64,449,055-64,449,036
MTOR-cDNA-exon30-31F	GAGGCCTTGGGGGAATGG	Chr1:11,217,225-11,210,283
MTOR-cDNA-exon31-32R	GCTGTCCCACTGACCTAAACC	Chr1:11,210,189-11,206,835
MTOR-cDNA-exon52F	GCTATGACCCGAGAGAAGTTTCC	Chr1:11,174,941-11,174,919

MTOR-cDNA-exon52-53R	CAGGCCTGTAACCTCCATAGC	Chr1:11,174,878-11,174,499
MTOR-cDNA-exon16F	TGGCCTGGAAATGAGGAAATGG	Chr1:11,292,580-11,292,559
MTOR-cDNA-exon16-17R	CAGAGCCACCTGCCTTTTGG	Chr1:11,292,503-11,291,483
SZT2-F	tgtaaaacgacggccagtTGGGATCTAGGGACACCCTCAA	Chr1: 43,905,509-43,905,530
SZT2-R	caggaaacagctatgaccCACTGCTGCCAGTGTAGCACCT	Chr1: 43,905,886-43,905,907
SZT2-cDNA-ex41-42-F	CACTGCCCAGTGATGATTATGC	Chr1:43,900,986-43,902,848
SZT2-cDNA-ex42-R	CCCATTTTGAGGCGTGAATGG	Chr1:43,902,952-43,902,971
SZT2-cDNA-ex47-F	CACTCTCCCAAGTACACAGATAGC	Chr1:43,904,720-43,904,743
SZT2-cDNA-ex48-49-R	TGCAGGCAACCCCTTTGC	Chr1:43,905,108-43,905,287
SZT2-cDNA-ex55-56-F	ACACAGCAGGCTGCCAAAGC	Chr1:43,907,792-43,907,948
SZT2-cDNA-ex56-R	CTCTGCCGGGGAGCATTTCG	Chr1:43,907,989-43-908-008

2.3 Polymerase Chain Reaction (PCR)

PCR was performed to amplify the DNA segment containing a region within the *SZT2* gene, using AmpliTaq Gold DNA Polymerase (Applied Biosystems, California, USA) on a Veriti 96-Well Thermal Cycler PCR machine (Applied Biosystems). Components of the PCR reaction and PCR cycle parameters are summarized in Table2.2. 360 GC-enhancer, dNTP, $MgCl_2$, and Taq Gold 360 polymerase were supplied with AmpliTaq Gold® DNA Polymerase (Applied Biosystems). The Mixture of PCR reagents was prepared for several reactions to minimize pipetting errors.

Table 2.2 PCR reagents and cycle parameters

Reagent	Concentration	Volume per reaction (µL)
Nuclease free H_2O		10.38
360 GC-enhancer		5
10xBuffer Gold 360	1x	2.5
dNTP ABI	0.2 mM	2.0
MgCl2 360	1.5 mM	1.5
Primer mix	5 µM	2.5
DNA	50ng/µL	1
Taq Gold 360 polymerase	0.5 U/µL	0.12
Total volume		25

Step	Temperature (° C)	Duration	Cycles
Initial denaturation	95	10’	1
Denaturation	95	30’’	30
Annealing	60	30’’	30
Elongation	72	1’	30
Final extension	72	7’	1

2.4 Purification of PCR products

In order to remove excess primers, dNTPs, enzyme, primer dimers, and salts from PCR reaction mixtures PCR products were purified in a Biomek® FXP Laboratory automated workstation (Beckman Coulter, Indianapolis, USA), using Agencourt Ampure magnetic beads (Beckman Coulter Genomics, Indianapolis, USA). A volume ratio of 1.8: 1 (bead: DNA) was used. DNA fragments were washed with 70% ethanol and eluted in nuclease free water.

2.5 Sanger sequencing

In this method DNA primers hybridizing to the DNA region of interest are extended by DNA polymerase. In each reaction, deoxynucleotides (dNTPs) allow chain reaction to be continued and Fluorescent-labelled dideoxynucleotides (ddNTPs), which lack the 3'-hydroxyl group, and each is labelled with one of the four colors, lead to termination of polymerization. Therefore, DNA fragments of different length, terminated with ddNTPs, are produced. The fragments with terminator nucleotides (ddNTPs) are then identified by each specific fluorescent color. Based on the primers used in PCR, Sanger sequencing could be performed with 2 different kind of primers; M13 universal forward and reverse primers (Figure2.1), and specific primers which have been designed for a specific genomic sequence next to the region of interest. The sequencing procedure was performed using BigDye Terminator v3.1 Cycle Sequencing Kit (Life Technologies, California, USA) on a Veriti™ 96-Well Fast Thermal Cycler (Applied Biosystems, California, USA). The reaction's components are shown in Table2.3 PCR master mix was prepared for several reactions in the same tube and divided into a number of tubes to minimize pipetting errors.

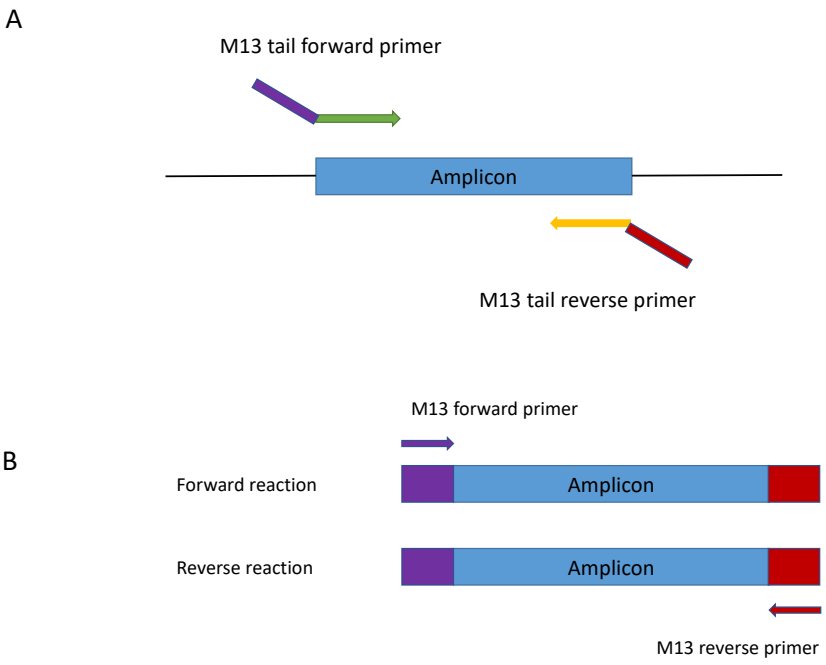


Figure 2.1 M13- tailed primers in PCR and Sanger sequencing. M13-tailed forward and reverse primers which contain complementary sequence to either side of the region of interest that flanked with M13 forward and reverse tail for using in PCR (A). The M13 universal forward and reverse primers for using in Sanger sequencing (B).

Table 2.3 Components and cycle parameters of Sanger sequencing reactions. 5X Sequencing Buffer, Ready Reaction Mix and M13 Primer supplied with the BigDye Terminator v3.1 Cycle Sequencing Kit. Step 1 contains denaturation, annealing and elongation which went for 25 cycles.

components	Volume µL (per reaction)
Nuclease free water	4.15
Purified PCR product	2
5x Sequencing buffer	2
Ready Reaction Mix	0.25
Primer	1.6

Step	Temperature	Time
Step1	96°C	10 sec
	50°C	5 sec
	60°C	4 min
Step2	10°C	∞

2.6 Agarose gel electrophoresis

In order to investigate the specificity of the designed primers and to separate DNA fragments, PCR products were separated by electrophoresis on an 1.5% agarose gel at 100 V/cm gel for 45 minutes. 1.5% SeaKem LE agarose (Bio Whittaker Molecular Applications, Rockland, ME, USA) was added to 50µL of 1X Tris-Acetate-EDTA (TAE) buffer and heated in a microwave until agarose was completely dissolved in the buffer. After cooling at room temperature, Sybersafe (0.5µg/mL) was added. After completion of the run DNA fragments in the gel were visualized on an ImageQuant Las 4000 instrument (GE Healthcare Life Sciences, Marlborough, USA).

2.7 Isolation and Quantification of RNA

Total RNA was extracted from HEK293T cells using PARIS™ Kit Protein and RNA Isolation System (ThermoFisher, Vilnius, Lithuania) according to the manufacturer's instructions. Briefly, cells were washed with 2mL PBS, and after disrupting cells with 300µL Disruption buffer (supplied with the kit) on ice, equal amount of 2x Lysis/Binding Solution (supplied with the kit) was added to the cell lysate. Next, the same volume of 100% ethanol as 2x Lysis/Binding Solution was added. The sample mixture was applied to a Filter Cartridge assembled on a Collection Tube (supplied with the kit). Finally, after washing steps with Wash Solution 1 and 2/3, RNA was first eluted with 40 µL, and then 10 µL of 95°C Elution Solution (supplied with the kit) at the time. The RNA solution was stored at -80°C.

The concentration and purity of the extracted RNA was measured by using NanoDrop ND-1000 Spectrophotometer (Marshall Scientific, Montchanin, USA). NanoDrop measures the absorbance values, concentrations and the ratios at 230/260 nm and 260/280 nm.

2.8 cDNA synthesis and amplification

Total RNA extracted from HEK293T cells underwent reverse transcription to produce cDNA using High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Vilnius, Lithuania). Random primers were used to amplify the cDNA, and the amplification was performed using a 2720 Thermal Cycler (Applied Biosystems, East Lyme, USA). Thermal cycle parameters and components of reverse transcription reactions are shown in Table2.4. 250ng of RNA was applied and nuclease free H_2O was added to reach the total amount of 20 µL per reaction.

Table 2.4 Thermal cycle parameters and components of reverse transcription. 10x RT Buffer, 25x dNTP Mix (100 nM), 10x RT Random Primers, and MultiScribe® Reverse Transcriptase are supplied with the kit.

Component	Volume per reaction (µL)
10x RT Buffer	2.0
25x dNTP Mix (100 nM)	0.8
10x RT Random Primers	2.0
MultiScribe® Reverse Transcriptase	1.0
Nuclease-free H_2O and RNA	4.2
Total per reaction	10.0

Parameter	Step1	Step2	Step3	Step4
Temperature(°C)	25	37	85	4
Time (minute)	10	120	5	∞

2.9 Gel purification of DNA fragment

PCR amplified cDNA fragments were excised from agarose gel and purified using QIAquick Gel Extraction Kit (QIAGEN, Hilden, Germany) following the manufacturer's protocol. Briefly, equal amount of isopropanol to the excised agarose with the DNA fragment was added. Then, samples were applied to QIAquick columns. After washing steps with the buffers supplied with the kit, samples were eluted in 40 µL of nuclease free water. DNA fragments purified from the gel were Sanger sequenced.

2.10 Protein extraction

To investigate the protein expression of transfected cells, proteins from 500,000 HEK293T cells were extracted for Western blotting. Cells were detached after incubation for 1-2 minutes with 1µL Gibco™ Trypsin-EDTA (0.05%) Phenol red (ThermoFisher, Leicestershire, UK). Next, 5 µL of Dulbecco's Modified Eagle Medium (1X-DMEM, ThermoFisher) with 10% of fetal bovine serum (FBS, ThermoFisher) and 1% of penicillin-streptomycin (PS) 5,000(U\mL) (ThermoFisher) was added to

deactivate the Trypsin, and cells were pipetted into a 15mL tube, and centrifuged for 7 minutes at 1500 rpm using Centrifuge 5810R Eppendorf (Hamburg, Germany). The cell pellet was washed twice with DPBS 1x Dulbecco's Phosphate-Buffered Saline (ThermoFisher), and centrifuged for 8 minutes using Centrifuge 5810R Eppendorf at 4000 rpm. The pellet was resuspended in 150µL of Radioimmunoprecipitation assay buffer, RIPA-I (containing RIPA buffer, Sigma-Aldrich, USA, supplied with 0.1% protease inhibitor and phosphatase inhibitor cocktail, Thermo Scientific, USA). Cells were rotated using HulaMixer Sample Mixer, Life technologies (AS, Norway) for 15 minutes at 4°C to allow for the lysis. Cell debris was spun down using Beckman Coulter Microfuge 22R Centrifuge (California, USA) at 14 000 rpm at 4°C for 15 minutes. Finally, the supernatant was transferred to a new tube and stored at -80°C.

2.10.1 Measurement of protein concentration

The concentration of protein was measured using Pierce™ BCA Protein Assay Kit (Thermo Scientific, USA). To draw a standard curve, two-fold dilutions series of Bovine Serum Albumin (BSA) starting at 2 mg/mL in RIPA-I buffer were used. Samples were diluted 1:3 in RIPA-I buffer. 5 µL of each diluted sample and standard were added to a 96-well plate in triplicates. Next, 200 µL of WR reagent (supplied with the kit) was added to each well and the plate was covered by aluminum foil to be protected from unwanted reactions due to exposure to the light, and mixed on a plate mixer. The plate was incubated at 37°C for 30 minutes. Absorbance at 562 nm was recorded and the concentration was back-calculated based on the standard curve using VersaMax Tunable Microplate Reader (Molecular Devices, San Jose, USA).

2.11 Western blotting

Western blotting was in this thesis performed to study SZT2 and MTOR levels in proteins extracted from HEK293T cells as follows:

1. One volume whole lysate containing 20µg proteins was mixed with one third volume of loading buffer (Laemmli sample buffer (Bio-Rad, California, USA)). Samples were heated for 5 minutes at 97°C before being loaded on the gel.
2. The extracted proteins were separated by 1-D polyacrylamide Mini-PROTEAN® TGX™ Precast Gel electrophoresis (Bio-Rad). 5µL of Precision Plus Protein Dual color standards (Bio-Rad) were used as size markers. Gel was run at 4°C in cold TGS buffer (Tris/Glycine/SDS, Bio-Rad, Munich, Germany) for 50 minutes at 185 V/cm gel using PowerPac™ Basic Power Supply (Bio-Rad).

3. Blotting was performed in approximately 600 mL blotting buffer (Table2.5), in the presence of ice, with magnetic stirring, at 110V for exactly 20 minutes. Western blotting membrane 0.45 µm (Bio-Rad, Germany) and 4 thin Filter Papers (Bio-Rad, USA), for each gel, were used for blotting. Then, the membranes were incubated with Ponceau S solution (Sigma-Aldrich, Steinheim, USA) to stain proteins, and the membranes were washed with tap water.
4. Proteins on the membranes were blocked in 25mL blocking buffer (5% bovine serum albumin (BSA) in 1x TBST) for 60 minutes at room temperature to prevent unspecific binding when antibody is added. Next, the membranes were washed with 25mL TBST three times for 5 minutes.
5. The membranes were incubated with 5mL primary antibody (Table2.6) for 1 hour at room temperature and followed by washing in 25mL of TBST three times for 10 minutes.
6. The membranes were incubated with appropriate secondary antibody (Table2.6) for 1 hour at room temperature and were washed with 25mL of TBST three times for 10 minutes.
7. The signals on the membrane were detected by Amersham ECL Primer Western Blotting Detection Reagent (GE Healthcare Life Sciences, Buckinghamshire, UK) according to the manufacturer's instruction and visualized on an ImageQuant Las 4000 instrument (GE Healthcare Life Sciences, Marlborough, USA).

Table 2.5 Western blotting buffers

Buffer type	**Components**
1X Washing buffer (TBST)	100 mL of 10x Tris Buffered Saline (TBS) Bio-Rad (USA) + TWEEN® 20 Polyoxyethylenesorbitan monolaurate (Sigma-Aldrich, Louis, U.S.A). + 900 mL H_2O
Blotting buffer	3.03g Trizma® base (Sigma-Aldrich, Louis, U.S.A) + 14.4g glycine (Bio-Rad, USA) + 800 mL H_2O + 200 mL methanol

Table 2.6 Antibodies used in Western blotting

Primary antibodies	**Host**	**Provider**	**Catalog number**	**Dilution**
mTOR Monoclonal Antibody (215Q18)	Mouse	ThermoFisher scientific, Rockford, USA	AHO1232	1:100
SZT2 polyclonal antibody	Rabbit	LIftespan Biosciences, Washington, USA	LS-c178753	1:500
GAPDH (D16H11) XP®	Rabbit	Cell signaling technology, Massachusetts, USA	5174	1:1000
Secondary antibodies	**Host**	**Provider**	**Catalog number**	**Dilution**

Goat Anti-Rabbit Ig, Human ads-HRP	Rabbit	Southern Biotech, Birmingham, USA	4010-05	1:1000
Goat Anti-Mouse Ig, Human ads-HRP	Mouse	Southern Biotech, Birmingham, USA	1010-05	1:1000

2.11.1 Stripping and Loading control

In order to reuse the Western blotting membrane, stripping of the membrane was performed between the use of different primary and secondary antibodies. In order to do that, the membranes were washed three times for 5 minutes with 25mL TBS. Then, the membranes were incubated with 25mL Restore™ western blot Stripping Buffer (ThermoFisher, USA) for 15 minutes at room temperature. Next the membranes were washed three times for 5 minutes with 25mL TBS to remove stripping buffer. The membranes were blocked with 25mL of 5% BSA in TBST for 60 minutes at room temperature. The final washing step was performed three times for 5 minutes with 25mL TBST.

2.12 siRNA Transfection

Synthetic siRNAs are double-stranded RNA molecules (21–27 nucleotides in length), with 2 nucleotides overhang at 3' ends, that reduce expression of a gene. By incorporating one strand of a siRNA into the multi-subunit ribonucleoprotein complex (RISC), it is directed to a target mRNA and may cause degradation of the mRNA resulting in reduced gene expression. To introduce siRNAs into cells, lipids that form complexes with the siRNAs can be used. 50,000 HEK293T cells were seeded per well in 6-well plates and were incubated for approximately 24 hours. 9µL of Lipofectamine™ RNAiMAX Transfection Reagent (ThermoFisher, Vilnius, Lithuania) was diluted in 150µL of Opti-MEM™ -reduced serum medium (ThermoFisher, Grand Island, USA). 2,3,4, or 5 µL of 10µM siRNA (Table2.7) was diluted in 150µL of Opti-MEM™ and mixed well by several pipetting. Diluted siRNA was added to diluted Lipofectamine RNAiMAX (1:1ratio) and incubated for 5 minutes at room temperature. 250µL of siRNA-lipid complex was added to the cells, and the cells were incubated at 37°C (on parallel plates).

Table 2.7 siRNA used for knocking down FK506-binding protein 12-rapamycin-associated protein 1 (*FRAP1*)/*MTOR*, and *GAPDH* genes

siRNA
Silencer® Select Negative Control #1 siRNA (Ambion, USA)
Silence® Select GAPDH Positive Control siRNA (Ambion)
FRAP1 Silence® Select Validated siRNA (Ambion)
SZT2 Silence® Select Validated siRNA (Ambion)

2.13 Real time quantitative PCR

Real-time qPCR is a technique in which nucleic acid sequences (DNA or RNA) are quantified. To detect the amount of DNA, nonspecific fluorescent dyes such as SYBR Green can be used in which incorporation of the dye into the PCR product (double-stranded DNA) produces fluorescent signals. By increasing the PCR product, the fluorescent signals will increase.

2.13.1 Primer validation

Prior to performing comparative gene expression experiment by real-time qPCR, efficiency of the qPCR primers was tested on three primer pairs for each gene with standard curves using 5 dilution points (two-fold dilutions) starting at 8 ng per reaction in triplicates using QuantStudio™ 12K Flex Real-Time PCR System (Applied biosciences). Reagents are displayed in Table2.8. Primers were designed based on parameters showed in section2.2. Plate setup is shown in Figure 2.2. The best primer pairs were chosen based on their optimal correlation coefficient $(R)^2$ value, slope and configuration of dissociation curves, to apply in gene expression experiments. Efficiency of the primer pairs is the rate at which a PCR amplicon is generated, and it is calculated based on the slope of the standard curve and R^2 value which was analyzed automatically on QuantStudio 12K Flex software. Slope, R^2 value and efficiency in the range of -3.1 and -3.6, 90% and 110%, 0.99 and 0.999, respectively are considered satisfactory. Single peak in the dissociation curves determines the high specificity of the primer pairs without producing nonspecific products(47). Dissociation curves and standard curves were obtained using QuantStudio 12K Flex software.

Table 2.8 Reagents used in qPCR and cycle parameters. PCR steps went for 40 cycles.

Component	**Volume per reaction**
SYBR® Green JumpStart™ Taq ReadyMix™ (Sigma-Aldrich Merck, New Jersey, USA)	5 µL
Primer mix(2µM)	1 µL
cDNA	3 µL
H_2O	1 µL
Final volume	10 µL

Step	Temperature	Time
Hold step	50°C	2 min
	95°C	10 min
PCR step	95°C	15 sec
	60°C	1 min
Melt step	95°C	15 sec
	60°C	1 min
	95°C	15 sec

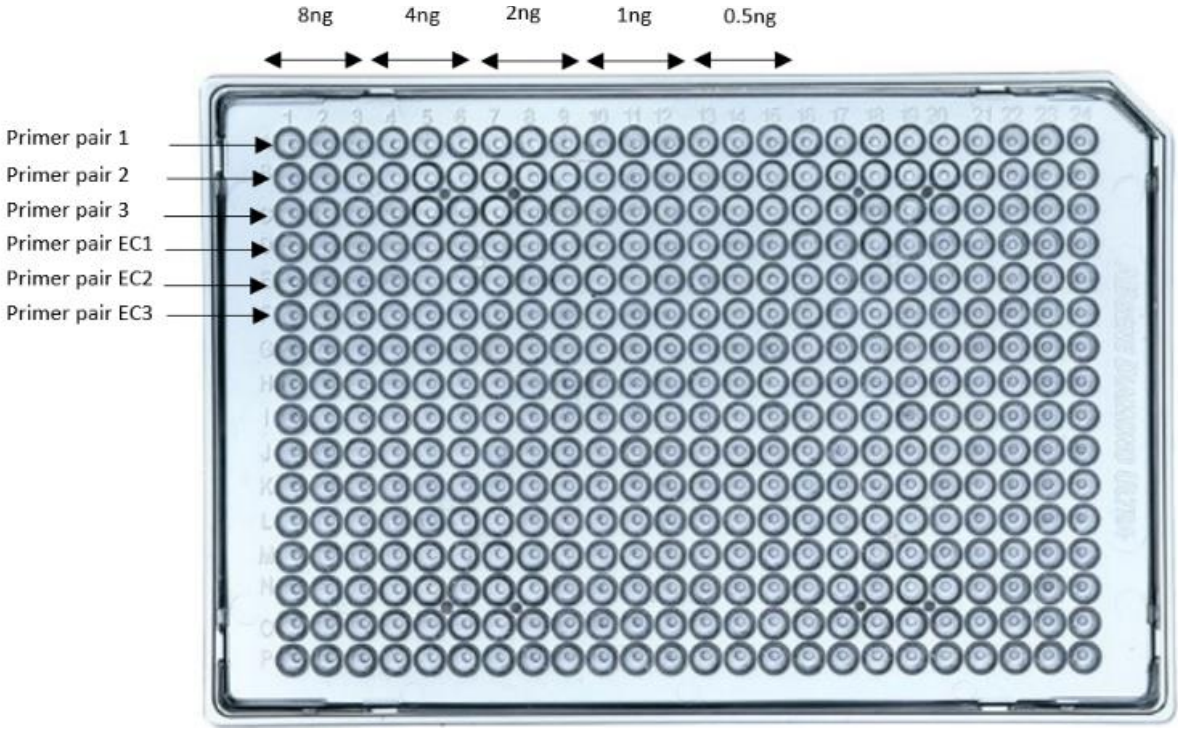


Figure 2.2 Setup of a 384-well RT-qPCR plate in primer validation experiment. EC: Endogenous Control. Primer pairs 1,2 and 3 used in primer validation experiments.

2.13.2 Comparative gene expression experiment

Comparative gene experiment was performed using 4 ng of cDNA per reaction in triplicates from HEK293T cells to observe the mRNA expression using QuantStudio 12K Flex Real-Time PCR System machine (Thermo Fisher Scientific, USA). Reagents used and cycle parameters are given in Table2.8. Chosen primer pairs from primer validation experiment were used. Amplification levels were calculated using the ΔΔCt method, normalizing with an endogenous control gene.

2.14 Immunofluorescence (IF) staining

100,000 fibroblast cells were seeded on sterilized coverslips in a 6-well plate, and incubated at 37°C overnight with the DMEM medium containing 10% FBS and 1% PS. The manufacturer protocol was followed for producing the amino acid-free medium. In brief, 8.59 g of RPMI medium w/o amino acids, sodium phosphate powder (USBiological, Salem, Massachusetts, USA) was dissolved in 800-900 mL of ddH_2O, and 2g of sodium bicarbonate was added. Then, pH was adjusted to 7.3, using 50mM Sodium hydroxide (NaOH). After that, additional ddH_2O was added to the final volume of 1000 mL, and it was filtered using Pall Acrodisc 32mm syringe filter with 2μm Supor membrane (Pall corporation, Newquay Cornwall, UK). Then, 10 % FBS and 1% PS were added. On the following day, cells were washed once with 2mL of the amino acid- free medium and incubated in 2 mL of it (per well in 6-well plates) for 60 minutes. For the re-stimulation of cells after 60 minutes amino acid starvation, cells were incubated in the 1x DMEM with 10% FBS and 1% PS for 10 or 20 minutes. Then, cells were washed once with 2 mL of Dulbecco's phosphate-buffered saline (DPBS) (Life Technologies, Paisley, UK), and were fixed in 1 mL of paraformaldehyde solution 4% in PBS (Santa Cruz Biotechnology, Dallas, TX) for 10 min at room temperature. Then, the cells were washed three times with 2 mL DPBS. Coverslips were placed on a parafilm in a humidified chamber and each coverslip was washed with 1 mL of DPBS. Cells were blocked and permeabilized with DPBS-AT (1% Sodium acetate buffer (Sigma-Aldrich, USA), 0.5% Triton™ X-100 (Sigma-Aldrich, USA) in DPBS) for 15 min at room temperature. Afterwards, cells were incubated with 30μL primary antibody diluted in DPBS-AT for 2 hours at room temperature (antibodies are indicated in Table 2.9). Coverslips were washed with 1 mL DPBS and incubated with 30 μL of an appropriate secondary antibody for 45 minutes at room temperature (antibodies are shown in Table 2.9). Coverslips were washed with 1mL DPBS. DNA was stained with 40μL of 0.6μg/mL Hoechst 33258 in PBS (Invitrogen, Eugene, Oregon, USA) for 2 minutes at room temperature. Each coverslip was washed with 1 mL DPBS and rinsed by dipping 10 times in 50 mL ddH_2O. When the coverslips were

dried, by approximately 20 minutes incubation at room temperature, 3µL ProLongTM Gold antifade reagent was added to the slides.

Table 2.9 Antibodies used in immunofluorescence staining

Primary antibodies	**Host**	**Provider**	**Catalog number**	**Dilution**
mTOR Monoclonal Antibody (215Q18)	Mouse	ThermoFisher scientific, Rockford, USA	AHO1232	1:100
Anti-LAMP2 antibody	Mouse	Abcam, Cambridge, UK	Ab25631	1:200
LAMP1(D2D11) XP	Rabbit	Cell signaling technology, Massachusetts, USA	9091	1:200
mTOR (7c10)	Rabbit	Cell signaling technology, Massachusetts, USA	2983	1:100
Secondary antibodies	**Host**	**Provider**	**Catalog number**	**Dilution**
Donkey Anti-Rabbit Ig, H&L	Rabbit	Abcam, Cambridge, UK	Ab150075	1:1000
Goat Anti-Mouse Ig, H&L	Mouse	Abcam, Cambridge, UK	Ab150117	1:1000

2.14.1 Imaging

The imaging of the stained cells was performed on a Zeiss LSM 700 microscope (Cambridge, UK) and Zen imaging software used for Acquisition and Analysis.

3 Result

3.1 Whole exome sequencing, variant calling and filtering

Whole exome sequencing was performed on DNA of the patient and her parents to identify potential disease-causing variants. Samples were prepared by Asbjørn Holmgren (Frengen's research group) before sequencing at the Norwegian Sequencing Centre (NSC). In short, library preparation, exome capture and quantification of captured exome library were performed according to manufacturer's protocol, and then submitted for paired-end sequencing using Illumina HiSeqX platform. Alignment and variant calling of the data were performed by Dulika Sumathipala (Frengen's research group). Filtering of the data was performed in collaboration with Dulika Sumathipala on the variant calling format (VCF) files to focus on rare variants predicted to be pathogenic using FILTUS (http://folk.uio.no/magnusv/FILTUS/)(48). The data were filtered for homozygous, compound heterozygous variants and also for *de novo* variants. Following variants were retained:

1. Variants with a Combined Annotation Dependent Depletion (CADD) score greater than 10. The CADD score is a pathogenicity estimate of single nucleotide variants and insertion or deletion variants.
2. Variants with a PhyloP score greater than 1.5. PhyloP score measures evolutionary conservation at individual alignment sites.
3. Variants with a Minor allele frequency (MAF) less than 0.01 according to the Genome Aggregation Database (gnomAD) (https://gnomad.broadinstitute.org/). GnomAD is a database containing exome and genome data of 141,456 unrelated individuals.

Using this cut-off, no candidate variants were detected among the *de novo* and homozygous variants (results not shown). The results when filtering for compound heterozygous variants are shown in Table3.1.

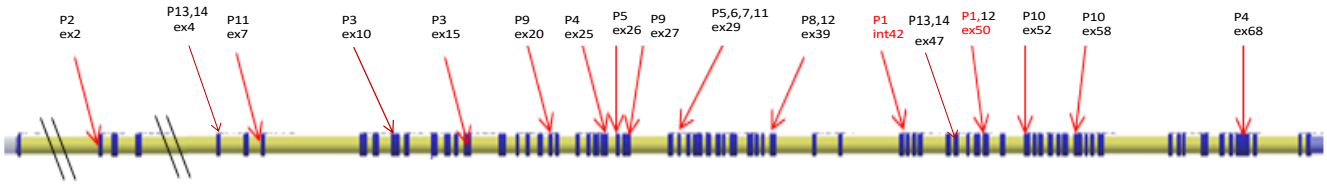
Table 3.1 Compound heterozygous variant detected after filtering the WES data

Gene name	Chromosome	Chromosome position	Reference allele	Observed allele	Effect	Read depth
MACF1	1	39757678	G	T	Non-synonymous	40
SZT2	1	43902997	A	T	Non-synonymous	52
SZT2	1	43905659	G	A	Non-synonymous	49
TEKT4	2	95541427	G	T	Non-synonymous	49

TEKT4	2	95541442	C	T	Non-synonymous	46
FBXL7	5	15936996	G	A	Non-synonymous	79
SLCO1C1	12	20868135	G	A	Non-synonymous	178
SLCO1C1	12	20868136	G	A	Non-synonymous	177
PABPC3	13	25670797	C	G	Non-synonymous	70
PABPC3	13	25671150-25671152	CGG	-	In-frame deletion	249
PABPC3	13	25671155-25671163	GACGGAACT	-	In-frame deletion	256
PABPC3	13	25671210	C	T	Non-synonymous	290
PABPC3	13	25671292	C	T	Non-synonymous	254
PABPC3	13	25671311-25671315	TATGA	-	Frame-shift deletion	235
PABPC3	13	25671369	G	T	Non-synonymous	195
PABPC3	13	25671429	G	T	Non-synonymous	157
PABPC3	13	25671451	A	G	Non-synonymous	140
RNASEH2A	19	12917491	G	T	Non-synonymous	23
RNASEH2A	19	12917495	T	C	Non-synonymous	23

Searches in OMIM (https://www.omim.org/) and PubMed (https://www.ncbi.nlm.nih.gov/pubmed/) were performed to identify plausible connections between the variants detected and the patient's clinical features. The compound heterozygous variants in the *SZT2* gene (ENSG00000562955.1), Chr1: g.43902997A>T (NM_015284:c.6016+3A>T), dbSNP build 151: rs376516957 and Chr1: g.43905659G>A (NM_015284:c.6979G>A), NP_056099.3p: Asp2327Asn, dbSNP build 151: rs201622088 variants were concluded to be possible pathogenic variants in the patient , for the following reasons: First, *SZT2* variants Chr1: g.43902997 A>T and Chr1:g.43905659 G>A are not homozygous in GnomAD, and have MAF of 0 and 0.005 in GnomAD, respectively. Second, the CADD scores of 14.13 and 23.3, and PhyloP scores of 2.067 and 8.631 of the two variants are considered as satisfactory scores for a potential disease-causing variant. Third, literature searches showed 13 previously reported patients with mutations in *SZT2* (positions are shown in Figure3.1) and they all had recessive loss of function mutations in *SZT2* gene (compound heterozygous or homozygous) and phenotypes overlapping the phenotype in the patient in the current work (displayed in Table3.2)

A



B

Patient	Variants in SZT2	Zygosity	Reference
P1	c.6979G>A: p.Asp2327Asn, c.6016+3A>T	Compound heterozygous	This work
P2	c.73C>T: p.Arg25*	Homozygous	(49)
P3	c.1496G>T: p.Ser499Ile, c.2092C>T:p.Gln698*	Compound heterozygous	(49)
P4	c.3509_3512delCAGA: p.Thr1170ArgfsTer22, c.9703C>T: p.Arg3235*(50)	Compound heterozygous	(50)
P5-P7	c.4202_4204delTTC: p.Phe1401del	Homozygous	(51)
P8	c.3700_3716del: p.Asn1234Alafs*35, c.5482del:p.Gly1829Valfs*52	Compound heterozygous	(52)
P9	c.3947dup: p.Glu1317Glyfs*4, c.2929+1G>A: p.Leu939Aspfs*19	Compound heterozygous	(52)
P10	c.7303C>T: p.Arg2435Trp, c.8162C>G: p.Ser2721Cys	Compound heterozygous	(52)
P11	c.8596dup: p.Tyr2866Leufs*42, c.4181C>T: p.Pro1394Leu)andc.2930-17_ 2930-3delinsCTCGTG(53)	Compound heterozygous	(53)
P12	c.5499delC: p.Phe1834Serfs*47, c.6916G>A: p.Gly2306Arg	Compound heterozygous	(54)
P13 & P14	c.6553C>T: p.Arg2185Trp, c.498G>T: p.Gln166His	Compound heterozygous	(55)

Figure 3.1 Positions of the *SZT2* mutations previously described in patients and the variants identified in the patient in the current work (shown in red). Schematic illustration of the *SZT2* gene with the positions of the *SZT2* mutations described in patients (A). Described variants of the patients with mutations in *SZT2* (B). Del: deletion, fs: frameshift, *: stop codon.

Table 3.2 Clinical features in the patient in the current work compared to published patients with *SZT2* mutations

Main features in the patient in the current work (P1)	P2	P3	P4	P5-7	P8	P9	P10	P11	P12	P13 & P14	Total
Intellectual disability (ID)	_	_	_	+	+	+	+	+	_	+	9\13
Epilepsy	+	+	+	_	+	+	+	+	+	+	10\13
Macrocephaly	_	_	+	+	_	_	_	+	+	+	8\13
Dysmorphic features	+	+	+	_	+	+	+	+	+	_	8\13
Regression	+	+	+	+(2/3)	+	+	+	+	+	+(1/2)	11\13

These observations support that the variants are pathogenic, however because these variants have not previously been described in patients, further analyses were necessary to assess their pathogenicity.

3.1.1 *SZT2* variant Chr1: g.43905659 G>A verification by Sanger sequencing

In order to verify the *SZT2* variant Chr1: g.43905659 G>A, DNA was extracted from fibroblasts from the patient, her mother and father, PCR amplified using SZT2-F and SZT2-R primers (Table2.1), and sequenced. The alignment of the sequencing results from both reverse and forward reads of the patient to the reference genome (GRCh37 (homo_sapiens_ensembl_hg19.gff) 17), using SeqScape software is shown in Figure3.2. This result shows two sequencing peaks (G and A), which identify the heterozygosity at position Chr1: g.43905659 in the patient. Sanger sequencing on the PCR product from the father showed the same results which confirms that the variant is paternally inherited (result not shown).

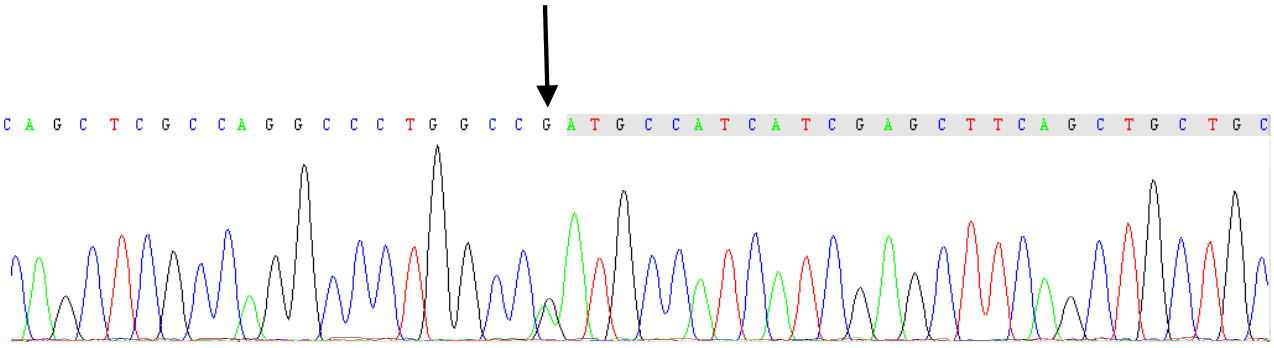


Figure 3.2 Sequencing result of patient's amplified *SZT2* DNA verified Chr1:43905659 G>A. The black arrow points at the heterozygous position, which is illustrated with two peaks (black for G and green for T). The reference sequence from GRCh37 (homo_sapiens_ensembl_hg19.gff)17 reference genome is shown on top of the sequence.

3.1.2 SZT2 variant Chr1: g.43902997 A>T causes skipping of exon 42

The variant Chr1: g.43902997 A>T is located in the splicing donor site of intron 42 and may cause skipping of exon 42. In order to assess this, total RNA was extracted from fibroblasts from the patient, her mother and father. Reverse transcription was performed (section 2.8) and a cDNA fragment was amplified with cDNA-specific primers (SZT2-ex40-41-F and SZT2-ex43-R, Table 2.1), which target the junctions of exon 40 and 41 to avoid amplification of contaminating genomic DNA in the RNA sample. The amplified cDNA fragment was separated by gel electrophoresis. The two bands

observed in the samples from patient and her mother (Figure3.3), indicate that they are heterozygous for the variant. Only one band was detected in cDNA from her father, which indicate homozygosity for the wild type variant. All five cDNA fragments were excised from the gel and sequenced.

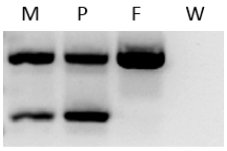


Figure 3.3 Gel electrophoresis result showing two *SZT2* cDNA-fragments from the mother (M) and patient (P), and one fragment in the sample from the father (F). A negative control sample was included using water instead of cDNA as a template (W).

Sanger sequencing was performed on excised cDNA fragments from the family trio. Results from both reverse and forward reads were aligned to the reference genome (GRCh37 (homo_sapiens_ensembl_hg19.gff) 17) using SeqScape software. Figure 3.4 displays the heterozygosity in the patient for the variant. In the transcript from the affected allele, the sequence of exon 41 is fused directly with the exon 43, demonstrating the skipping of exon 42 in the shortest cDNA fragment (Figure 3.4A). The same results were obtained from forward and reverse reads and in the fragments obtained from the sample from the mother (result not shown). Exon 42 was retained in the longer cDNA fragment from the patient, mother and father (Figure 3.4B).

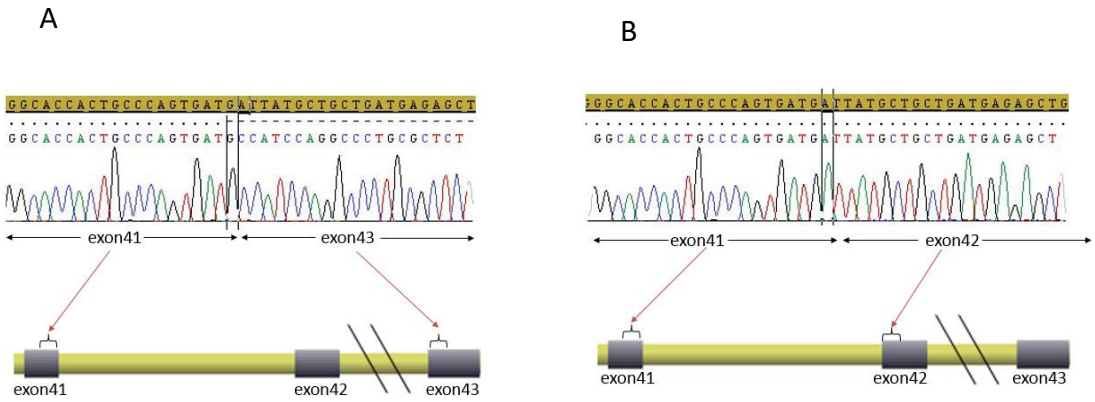


Figure 3.4 Sanger sequencing result of a *SZT2* cDNA fragment from the patient verified skipping of exon 42. The sequencing of the shortest cDNA fragment from the patient shows skipping of exon 42 (A), while the longer cDNA fragment retained exon 42 (B). The reference sequence from GRCh37 (homo_sapiens_ensembl_hg19.gff)17 reference genome is shown in brown boxes on top of each patient sequence.

In summary, these results document that the maternally inherited variant Chr1: g.43902997 A>T caused skipping of *SZT2* exon 42.

3.2 The effect of the compound heterozygous *SZT2* variants in cell signaling pathways

To study the effect of the patient variants in *SZT2* in cell signaling pathways, RNA sequencing of fibroblasts of the patient and four healthy controls were performed. The samples were prepared for RNA sequencing by Dulika Sumathipala using the Illumina TruSeq stranded RNA-seq sample prep (Illumina), and then sequenced on Illumina NextSeq 500 (Illumina) at the Norwegian Sequencing Centre (NSC). Reads were aligned to the Ensemble gene model GRCh37 (homo_sapiens_ensembl_hg19.gff) 17 with HISAT2 2.0.1.18. Unbiased differential expression analysis was performed by Dulika Sumathipala on the transcriptome of the patient versus the controls with False Discovery Rate (FDR) <0.1 using a bioinformatics analysis described in (56)which led to identification of 57 differentially expressed (DE) genes (Table 7.1, Supplementary data). The 57 DE genes were analyzed for pathway enrichment using R package PathfindR as described in (57), and protein-protein interaction (PIN) data of Reactom was used for identification of subnetworks. After searching for active subnetworks, identified subnetworks were filtered with following criteria; i) The lowest P. value threshold of 0.05 (over 10 iterations). ii) Number of DE gene threshold of 1. The resulting enriched pathways are shown in Table 3.3.

Table 3.3 Enriched pathways detected with the list of 57 differentially expressed genes

Pathway	Occurrence	Lowest P.	Highest P.	Up-regulated	Down-regulated
Cell surface interaction at the vascular wall	3	0.00013	0.00013	CD74, PSG3, PSG5, PSG8	
MHC class-II antigen presentation	4	0.00328	0.00328	CD74, HLA-DMB, HLA-DOA, HLA-DRA, HLA-DRB1	
Hemostasis	3	0.00999	0.00999	CD74, PSG3, PSG5, PSG8	GATA6, MRVI1, PCDH7, TGFB2, ZFPM2
ERBB2 activates PTK6 signaling	4	0.01044	0.01044		EREG
ERBB2 regulates cell motility	4	0.01377	0.01377		EREG
GRB2 events in ERBB2 signaling	4	0.01561	0.01561		EREG
PI3K events in ERBB2 signaling	4	0.01561	0.01561		EREG

Translocation of ZAP-70 to immunological synapse	10	0.02014	0.04024	HLA-DRA, HLA-DRB1	
Phosphorylation of CD3 and TCR zeta chain	10	0.02496	0.04986	HLA-DRA, HLA-DRB1	
PD-1 signaling	10	0.02668	0.05329	HLA-DRA, HLA-DRB1	
SHC1 events in ERBB2 signaling	4	0.02901	0.02901		EREG
RAF/MAP kinase cascade	4	0.03755	0.03755		EREG, FGF7, SHC3
MAPK1/MAPK3 signaling	4	0.04037	0.04037		EREG, FGF7, SHC3
Generation of second messenger molecules	10	0.04703	0.09389	HLA-DRA, HLA-DRB1	
Downregulation of ERBB2 signaling	4	0.04986	0.04986		EREG

Lowest and highest P. value were calculated over 10 iterations. Occurrence indicates how many times each pathway is enriched over 10 iterations. The two last columns show up-regulated and down-regulated genes from DE genes list, which are involved in the enriched pathways.

3.3 The cellular effect of compound heterozygous *SZT2* variants

To study the cellular effects of compound heterozygous *SZT2* variants identified in the patient, we aimed to perform immunofluorescence (IF) microscopy studies on fibroblasts. Prior to IF staining, antibody validation experiments were carried out to investigate the specificity of MTOR and SZT2 antibodies (Table 2.6). To do that, the *MTOR* and *SZT2* expressions were knocked down in HEK293T cells using MTOR-siRNA and SZT2- siRNA, respectively (Table 2.7, section 2.12). Next, total RNA was extracted (section 2.7), and cDNA synthesis was performed to produce total cDNA (section 2.8). Then, in order to investigate differences between transfected and un-transfected cells on mRNA expression level of target genes, relative gene expression experiment, using real-time qPCR, was performed on cDNA from the cells (section 2.13.2).

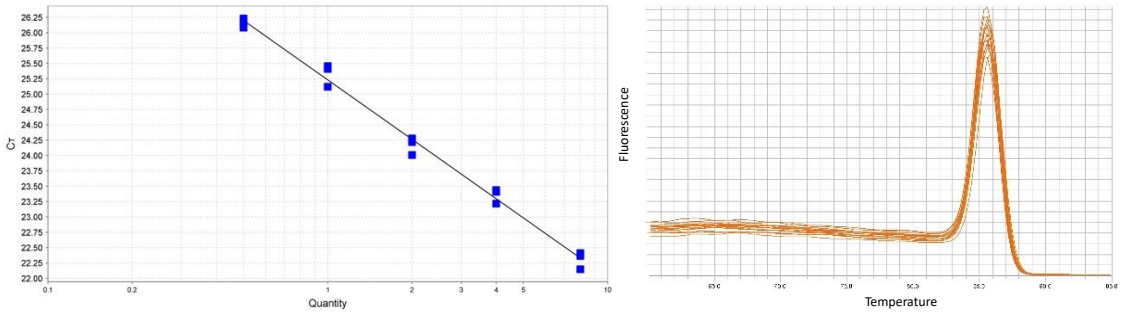
3.3.1 Validation of MTOR and SZT2 siRNAs using RT-PCR

We aimed to knock down *MTOR* and *SZT2* genes through using MTOR- and SZT2-siRNAs (Table 2.7), but first the efficiency and specificity of the primers were assessed.

3.3.1.1 Validation of primers for use in qPCR

Reverse transcribed RNA extracted from HEK293T cells, 8ng, 4ng, 2ng, 1ng and 0.5ng of cDNA, were used for each primer pairs as the template for real-time qPCR. The amplification of cDNA by the designed primer pairs was performed to analyze the efficiency of the primer pairs using real-time qPCR. Syber green was used in this experiment, which is an unspecific fluorescence color, and binds to any double-stranded DNA molecule. By increasing the temperature (at the last step of real-time qPCR cycle (melting step)), DNA molecules are starting to dissociate which leads to the desertion of a fluorescence color. Then, the color is detected by a real-time qPCR machine and results are shown by a dissociation or melting curve. Based on the content and length of DNA molecules, they begin to dissociate at different temperatures. Therefore, dissociation curve analysis is used for characterization of DNA molecules, and existence of more than one peak in a dissociation curve (or melting curve) can be due to the formation of a primer dimer, or amplification of a nonspecific product during real-time qPCR. Efficiency of amplification is calculated by $E = [10^{(-1/S)}] - 1$ formula in which S is the slope of a standard curve. Figure 3.5 shows the selected primer pairs' dissociation curves with a single peak which determines the high specificity of the primer pairs without producing nonspecific products, and standard curves which slope, R^2 and efficiency values are shown Table 3.4.

A



B

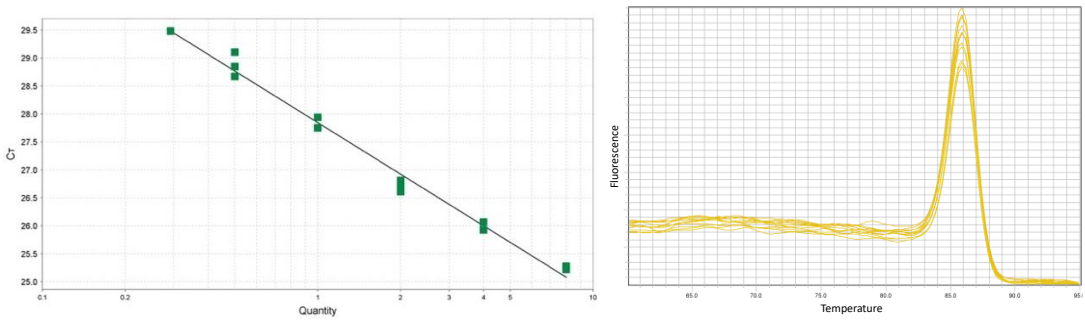


Figure 3.5 Standard curves and melting curves of primer pairs targeting exons 30, 31 and 32 of *MTOR* (MTOR-cDNA-exon30-31F and MTOR-cDNA-exon31-32R, Table 2.1) (A) and exons 56,57 of *SZT2* (SZT2-cDNA-exon57-56F and SZT2-cDNA-exon56R, Table 2.1) (B) used in comparative gene expression quantification experiment. Slope, R2 value and %efficiency in the range of -3.1 and -3.6, 90% and 110%, 0.99 and 0.999, respectively are considered satisfactory.

Table 3.4 Slope, R^2 and Eff% obtained from standard curves of the primer pairs used in comparative gene expression experiment.

Primer pair	Slope	R^2	Eff%
MTOR-cDNA-exon30-31F and MTOR-cDNA-exon31-32R	-3.212	0.991	104.778
SZT2-cDNA-ex55-56-F and SZT2-cDNA-ex56-R	-3.06	0.982	112.241

Thus, according to the parameters for primer efficiency, described in section 2.13.1 , primer pairs targeting exons 30, 31 and 32 of *MTOR* (MTOR-cDNA-exon30-31F and MTOR-cDNA-exon31-32R, Table 2.1), and exons 56,57 of *SZT2* (SZT2-cDNA-exon57-56F and SZT2-cDNA-exon56R, Table 2.1) were chosen for comparative gene expression.

3.3.1.2 Transfection validation experiment

Next, the transfection conditions were validated in HEK293T cells transfected with siRNA targeting the glyceraldehyde 3-phosphate dehydrogenase (GAPDH) transcript. In order to analyze transfection efficiency, HEK293T cells were transfected with four concentrations of siRNA targeting the glyceraldehyde 3-phosphate dehydrogenase (GAPDH) transcript, which is used as a reference

transcript. The cells were transfected with siRNAs at 64nM, 94nM, 128nM and 160nM concentrations, and incubated for 24 hours and 48 hours (section 2.12). In parallel, the cells were also transfected with 94nM negative control siRNA (Table 2.7), which is designed to have no targets in human cells. Comparative C_T method ($2^{-\Delta\Delta CT}$) was performed for analyzing the gene expression, in which C_T value is the number of cycles where the fluorescence signal reaches the threshold amount for being detectable, and since each sample was run in triplicate, the average C_T values were used in calculations (note that outliers were omitted from the data to avoid inaccurate assessments).

ΔC_T = Average C_T value of target gene - Average C_T value of endogenous control

$\Delta\Delta C_T$ = ΔC_T of sample - ΔC_T of calibrator

$\Delta\Delta C_T$ for calibrator sample is 0, therefore $2^{-\Delta\Delta CT}$ for calibrator sample is 1, and fold changes in the gene expression of other samples were compared to that. The Peptidylprolyl isomerase B (PPIB) transcript was used as the endogenous control in the RT-PCR experiments. The primer pair already validated, was used in the experiment (Table 2.1). The un-transfected sample was used as a calibrator sample, meaning cDNA copy number of the target gene (*GAPDH*) in the transfected samples were compered to cDNA copy number of the un-transfected sample. The results in Figure 3.6 show that cells transfected with GAPDH-siRNA have a reduced *GAPDH* expression level in comparison to un-transfected cells, and cells transfected with negative control-siRNA. Samples from un-transfected cells with no reverse transcriptase added, which were included in the cDNA synthesis experiments, did not show any amplification during the real-time qPCR experiment indicating that genomic DNA contamination was avoided (data not shown).

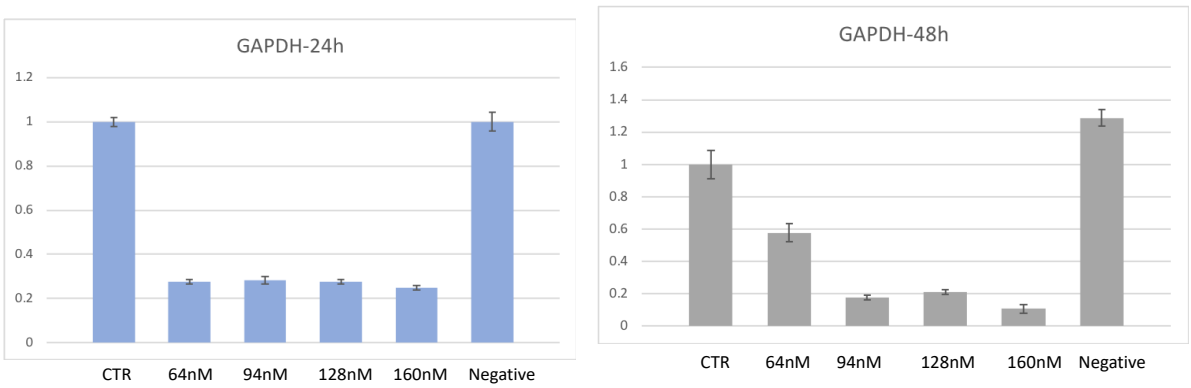


Figure 3.6 Relative quantification of GAPDH gene expression result using real-time qPCR. Bars show the ratio of the *GAPDH* expression levels in siRNA-transfected versus un-transfected cells after incubation for 24 hours (A) or 48 hours (B). Abbreviations: CTR: control (un-transfected sample), 64nM: 64nM GAPDH-siRNA, 94nM: 94nM GAPDH-siRNA, 128nM: 128nM GAPDH-siRNA, 160nM: 160nM GAPDH-siRNA, Negative: 64nM negative control-siRNA. Error bars indicating the standard deviation for each sample are shown on the top of each bar.

In conclusion, these results indicate that; first, *GAPDH* expression levels were reduced in all samples transfected with GAPDH-siRNA compared to un-transfected samples. Second, 64nM siRNA seems sufficient for interfering with the *GAPDH* expression, decreasing its expression level. Third, 24 hours incubation time seems sufficient for the siRNA transfection experiment to reduce the expression level.

3.3.1.3 Comparative gene expression quantification

In order to assess the efficiency of the MTOR-siRNA and SZT2-siRNA, HEK293T cells were transfected with the two siRNAs, separately for 24 hours (section 2.12). In parallel, cells were transfected with GAPDH-siRNA and negative control siRNA. Cell pellets from the experiment were divided into two fractions. Then, total RNA was extracted from one fraction and protein lysates were obtained from the other (section 2.10).

Next, quantitative real-time qPCR was performed on synthesized cDNAs using validated primer pairs in section 2.13.1 (MTOR-cDNA-exon30-31F, MTOR-cDNA-exon31-32R and SZT2-cDNA-exon57-56F, SZT2-cDNA-exon56R, Table 2.1).The un-transfected sample was used as a calibrator sample, which is the basis of the comparison. In Comparative gene expression experiment, the *PPIB* transcript was used as the endogenous control with validated primer pair. Figure 3.7 shows that in all MTOR-siRNA and SZT2-siRNA transfected samples (shown in yellow and green bars, respectively), the expression of the target gene was reduced in comparison to the un-transfected samples (CTR, calibrator samples). The negative control siRNA does not result in a decrease for the target genes in the experiments.

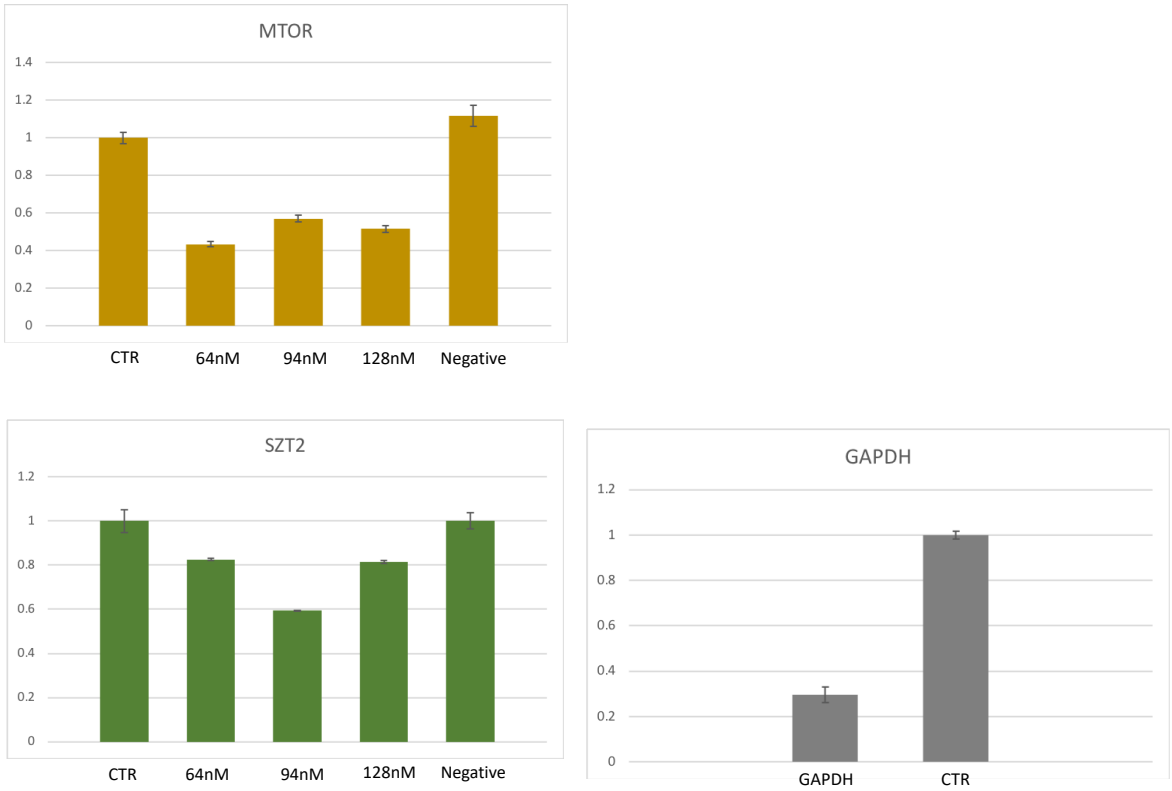


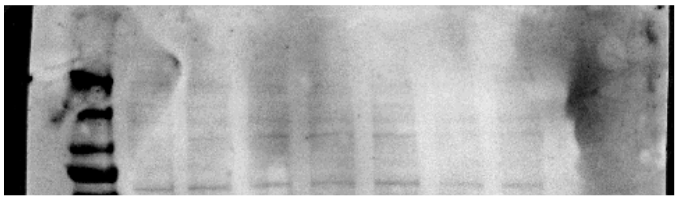
Figure 3.7 Relative quantification (RQ) Relative quantification of the *MTOR* expression (yellow bars), *SZT2* expression (green bars) and *GAPDH* expression (gray bars) of the samples transfected with GAPDH-siRNA (GAPDH) as a positive control for the transfection experiment compared to the un-transfected sample (CTR) using real-time qPCR**.** Abbreviations: 64nM: 64nM siRNA, 94nM: 94nM siRNA, 128nM: 128nM siRNA, Negative: 64nM negative control-siRNA. Error bars indicating the standard deviation for each sample are shown on the top of each bar.

These results show that the expression of *SZT2* and *MTOR* genes were decreased in cells transfected with the siRNAs targeting these genes. Thus, further analyses were performed to measure protein levels to assess SZT2 and MTOR antibody specificities.

3.3.2 Assessment of MTOR and SZT2 antibodies

Western blotting was performed on the protein lysates described in section 4.3.1.3 using the MTOR(215Q18) and SZT2 antibodies (Table 2.6), and loading control was performed on them, using GAPDH antibody (Table 2.6). No obvious bands were detected when the SZT2 antibody was used (Figure 3.8A). However, the bands detected with MTOR and GAPDH loading control were quantified using ImageJ software, and the MTOR to GAPDH ratios were calculated for each sample. Next, all samples were normalized against the un-transfected sample by dividing the value of un-transfected sample by MTOR to GAPDH ratios of each sample. Figure 3.8B shows reduction of MTOR/GAPDH in all transfected samples with MTOR-siRNA in comparison to un-transfected samples.

A



B

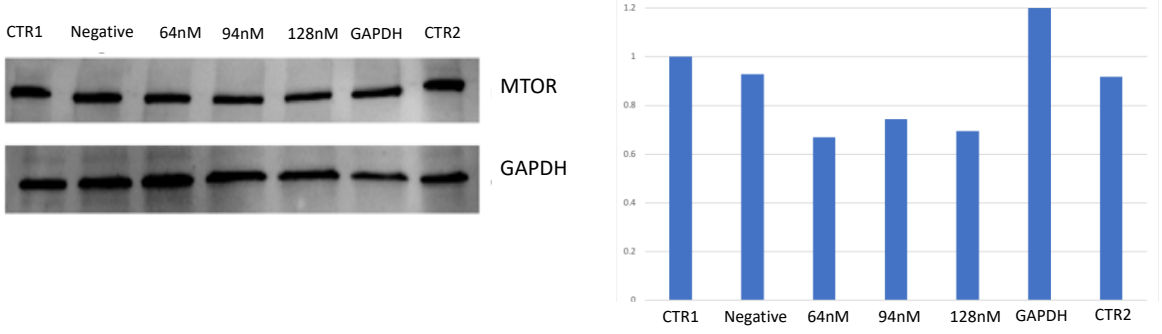


Figure 3.8 Western blot results from protein lysates from transfected and un-transfected HEK293T cells. The signals were detected with antibodies targeting SZT2 (A), MTOR and GAPDH as indicated (B). The bars (C) show the comparison of MTOR to GAPDH ratios of the quantified western blot signals, which were normalized against the un-transfected sample (CTR1). Abbreviations: CTR1 and 2: control (un-transfected), Negative: negative siRNA control, 64nM: 64nM siRNA, 94nM: 94nM siRNA, 128nM, 128nM siRNA, GAPDH: GAPDH-siRNA (positive control).

Therefore, these results indicate that MTOR antibody detects MTOR, whereas with SZT2 antibody no bands were detected. Thus, the MTOR antibody was further used in IF analyses.

3.4 immunofluorescence (IF) study

After validation of the antibody targeting MTOR protein, it was used for IF. MTOR has an important contribution in nutrient sensing, and the main activation site for MTOR is a lysosomal membrane. Thus, in the presence of amino acids MTOR should be active and located on the lysosomal membrane, whereas under the amino acid deprivation, MTOR is expected to be detached from the lysosome(58) . MTOR localization was analyzed by IF in amino acids starved fibroblasts. The study was carried out with two different treatments on fibroblasts from the patient and a healthy control:

1. Amino acid starved fibroblast cultures, in which fibroblasts were incubated in amino acids-free media for 60 minutes at 37°C.
2. Fibroblast cultures, when cells were first incubated in amino acids-free media for 60 minutes, and then incubated in normal media for 10 minutes at 37°C to stimulate cells with amino acids.

Antibodies targeting two lysosomal markers LAMP1 and LAMP2 were used to investigate the co-localization with MTOR. Next MTOR(7c10) antibody is also tested on cells. Following dilutions of the primary antibodies in PBS-AT were used to gain the optimal condition: 1:100- 1:200- 1:500. Best captured pictures are shown in Figure 3.9 and 3.10 after treatment 1 and 2, respectively. The merged photos display both signals from MTOR antibody and a lysosomal marker.

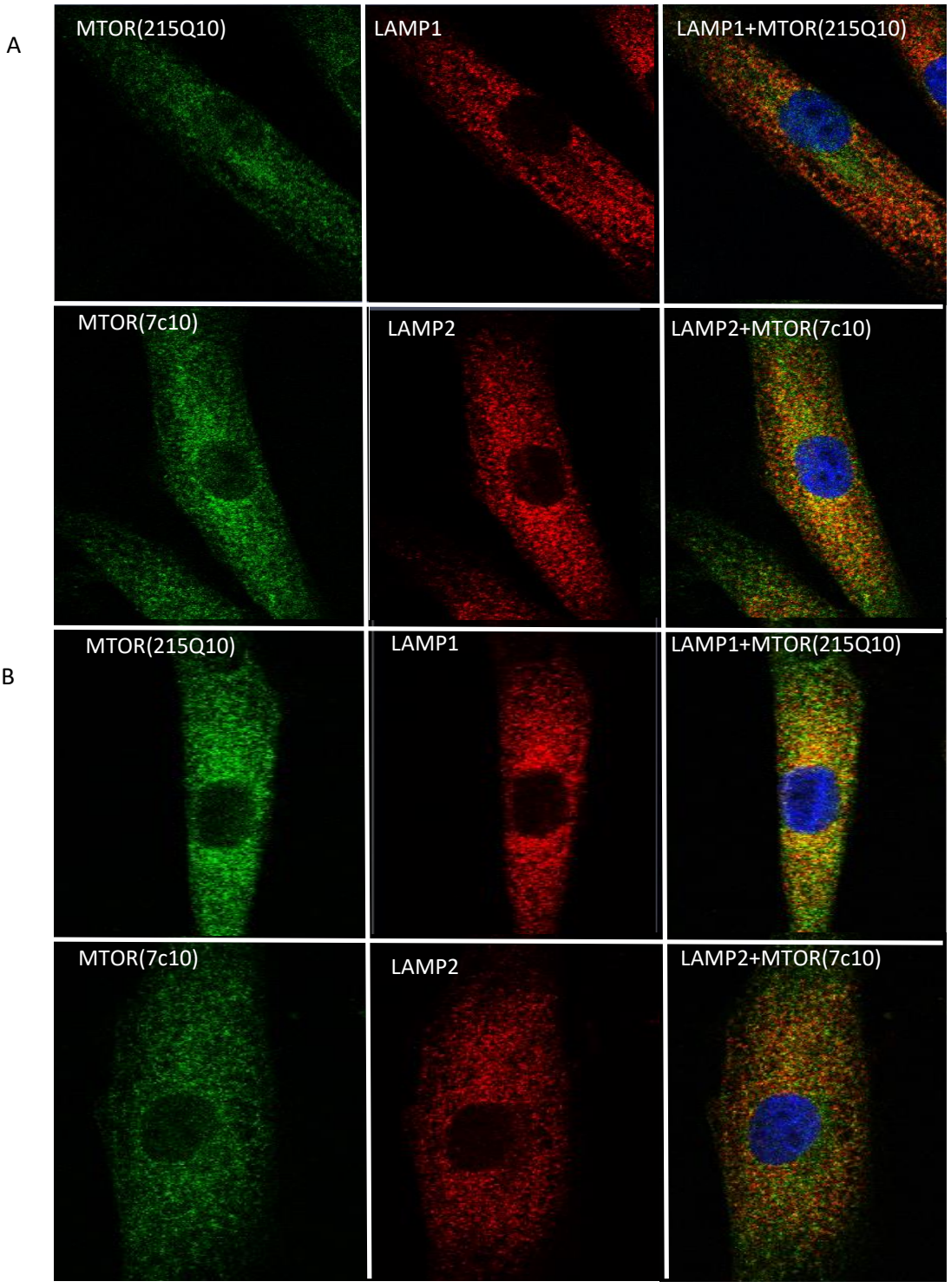


Figure 3.9 IF captured images with 63x Oil, when control fibroblasts (A) and patient fibroblasts(B) were grown in an amino acid-free medium for 60 minutes.

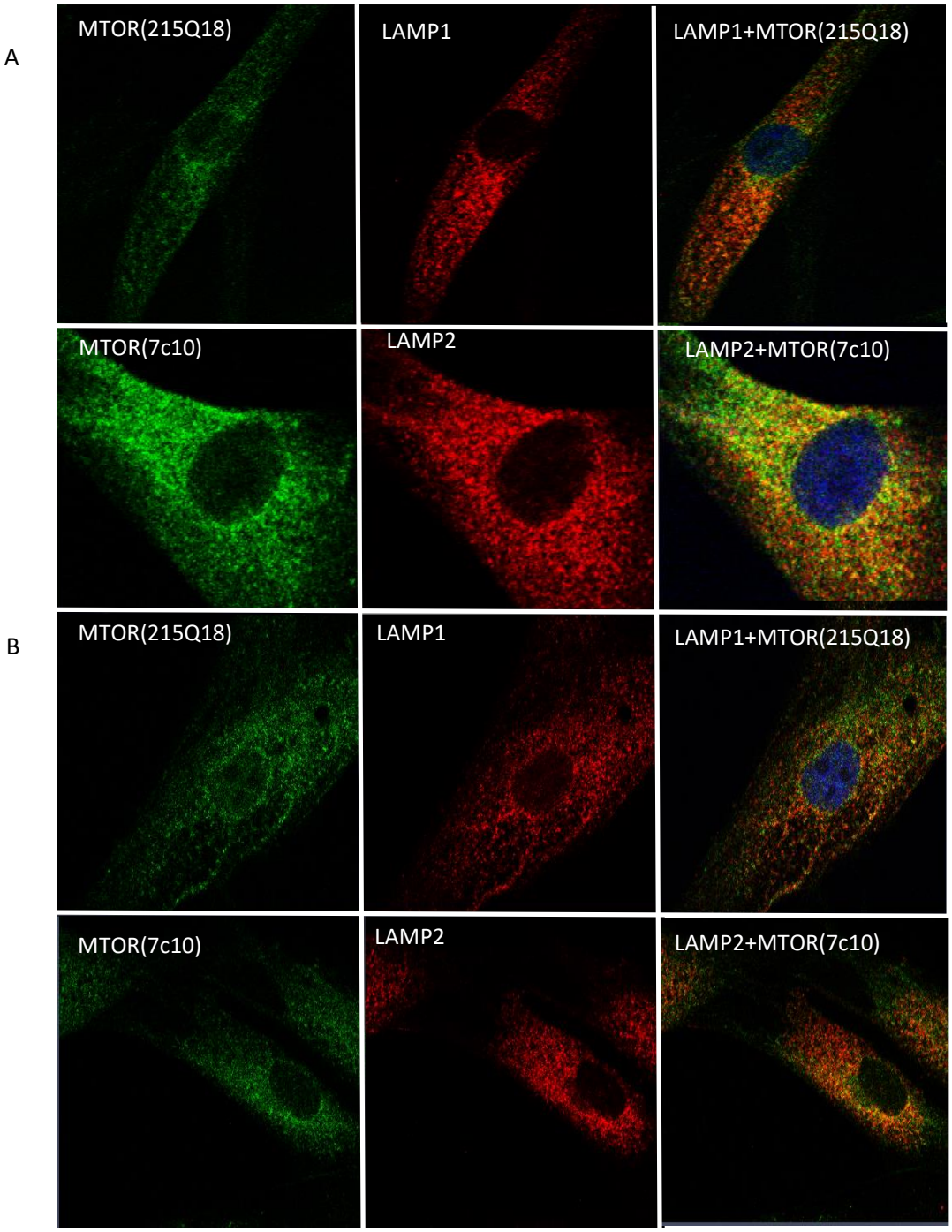


Figure 3.10 IF captured images with 63x Oil, when control's fibroblasts (A) and patient's fibroblasts(B) were deprived in an amino acid-free medium for 60 minutes and then re-stimulated with amino acids for 10 minutes.

Antibodies against LAMP1 and LAMP2 are supposed to bind to the LAMP1 and LAMP2 proteins in lysosomal membranes to detect lysosomes, and the presence or absence of amino acids does not influence in their localizations since they are part of the lysosomal membrane. However, MTOR in the presence of amino acids should be active and localized on the lysosomal membrane and in the absence

of amino acids it is inactive and detached from the lysosomal membrane in healthy control cells. Therefore, in the re-stimulated cells with amino acids we expected to see following patterns;

1. In amino acid-deprived control cells, MTOR is expected not to be located on the lysosomes and in the merged images MTOR may be detected in green separately from LAMP1 and LAMP2 in red. While, when cells were re-stimulated with amino acids both lysosomal markers and MTOR are expected to localize on the lysosomes resulting in yellow signals (combination of red for a lysosomal marker and green for MTOR) in the merged images.
2. In the patient cells, because MTOR may be constitutively active in both treatments, lysosomal markers and MTOR may be localized on the lysosomes in both conditions.

However, no differences were observed between starved and re-stimulated cells and between the patient and control cells.

4 Discussion

4.1 Identification of genetic variants using whole exome sequencing

Due to the lack of obvious matching in the clinical features of the patient studied in this project and known syndromes, hypothesis-free WES on family trio was performed, which has been increasingly successful in discovery of disease-causing variants in pediatric neurological disorders (59, 60)

Whole exome sequencing on DNA extracted from fibroblast samples of the patient and the parents was carried out to identify potentially disease-causing variants. Data filtering resulted in a list of 19 variants (Table 3.1). Literature searches were performed concerning the obtained variants using Pubmed and OMIM databases in order to identify potential association between variants in these genes and previously reported clinical phenotypes. *SZT2* mutations have previously been shown to cause a disorder presenting with epilepsy and macrocephaly (55). So far, 13 patients have been reported with *SZT2* mutations and phenotypes overlapping with the phenotype of the patient in the current study (Table 3.2). All these patients had loss of function mutations. Thus, the *SZT2* variants, Chr1: g.43902997 A>T and Chr1: g.43905659 G>A were suspected to be the cause of disease in the patient in the current work.

Variant Chr1: g.43902997 A>T was shown to cause skipping of exon 42 and was maternally inherited, while variant Chr1: g.43905659 G>A is paternally inherited and leads to an amino acid substitution from Aspartic acid (Asp) to Asparagine (Asn). The missense variant has an un-known impact on protein function. While, the other variant leads to skipping of *SZT2* exon 42, which is 153 bp. This in-frame deletion may lead to a shortening of the SZT2 protein. To examine whether variant Chr1: g.43902997 A>T causes in a smaller protein or not, Western blotting can be carried out on the extracted protein from the patient. However, due to the lack of validated SZT2 antibodies, we could not study this.

The CADD score of a variant indicates the pathogenicity (61), and a cutoff between 10-20 is proposed for the determination of pathogenicity of a variant, where 99 is the most deleterious degree (with the scale of 1 to 99)(62), meaning that variants with CADD score below 10 are most probably benign variants. Based on this cutoff the SZT2 variants Chr1: g.43902997 A>T and Chr1: g.43905659 G>A which have CADD scores of 14.13 and 23.3, respectively, are considered as potential disease-

causing variants. However, the CADD score alone does not determine whether a variant is pathogenic. The minor allele frequency (MAF) were 0 and 0.005 in GnomAD for variants Chr1: g.43902997 A>T and Chr1: g.43905659 G>A, respectively, showing that the *SZT2* variants are rare variants, and they are not found homozygous in GnomAD.

These observations support that the variants are pathogenic, but because these variants have not previously been described in patients, further analyses were necessary to assess their pathogenicity.

4.2 Study the patient's fibroblast

The transfection of HEK293T cells was performed using siRNAs to knock down the *MTOR* and *SZT2* genes. The comparative expression experiments using qPCR showed that the mRNA expression level of both target genes was decreased by the siRNA transfections. Then, Western blotting was performed on the protein lysate of the transfected cells using SZT2 and MTOR antibodies. The Western blotting when using MTOR antibody showed a reduction in the MTOR protein expression level of cells transfected with MTOR-siRNA. The SZT2 antibody, however, did not detect a band of the expected size in the Western blotting experiment. Nevertheless, loading control for both experiments showed that proteins are loaded in equal quantity in the gels prior to the Western blotting steps. Therefore, the results indicate that the MTOR antibody has sufficient specificity, and the patient's cells were studied using only this antibody by immunofluorescence staining.

Immunofluorescence staining with the MTOR antibody was conducted on fibroblasts from the patient and a healthy control to assess the localization of MTOR, which is the central core of the mechanistic target of rapamycin complex1 (MTORC1). MTOR is a protein kinase, and it plays an important role in the regulation of cell growth, as well as organismal homeostasis through sensing and integrating different environmental signals in mammalian cells. The signals affecting the MTORC1 can be parameters such as cellular stress, energy source, oxygen levels, amino acid levels and growth factor signals. As it is indicated in Figure 4.1, boosting in the level of oxygen, energy source, amino acids and growth factors may lead to the activation of the MTORC1 pathway, while increasing in the cellular stress level may result in the inactivation of the MTORC1 signaling. As a result, the activated MTORC1 promotes cell growth through inducing anabolic processes such as metabolism, growth, macromolecule biosynthesis and cell cycle progression, while inhibiting catabolic ones such as autophagy(63-66). Fibroblasts have been shown as an efficient cell type for studying MTORC1 pathway and its association with lysosomes(67). Therefore, in order to study the localization of MTOR in

fibroblasts from patient and a healthy control inhibition and re-stimulation was pursued by withdrawal and reintroduction of amino acids (section 3.4). Our results show no obvious difference in the MTOR localization in the presence and absence of amino acids and between cells from the patient and healthy controls. The results do not show co-localization of MTOR and the lysosomal markers, LAMP1 and LAMP2, which may be explained by one of the following reasons; i) In our study fibroblasts were amino acid starved in the medium containing 10% FBS, and FBS has a small amount of amino acids, which can impact on the MTORC1 activation. Therefore, other treatments such as glucose starvation on cells before staining may be pursued. ii) Even though the MTOR antibody seemed to detect a signal in Western blotting that is in agreement with the expected band, we do not know whether it specifically detect the MTOR protein in the cells by immunofluorescence staining. iii) LAMP1 and LAMP2 are two proteins associated with late endosomal and lysosomal membranes(68), which are used as lysosomal markers. We have not validated the LAMP1 and LAMP2 antibodies for using in immunofluorescence study. Thus, further tests may be required to prove that the LAMP1 and LAMP2 antibodies detect lysosomes.

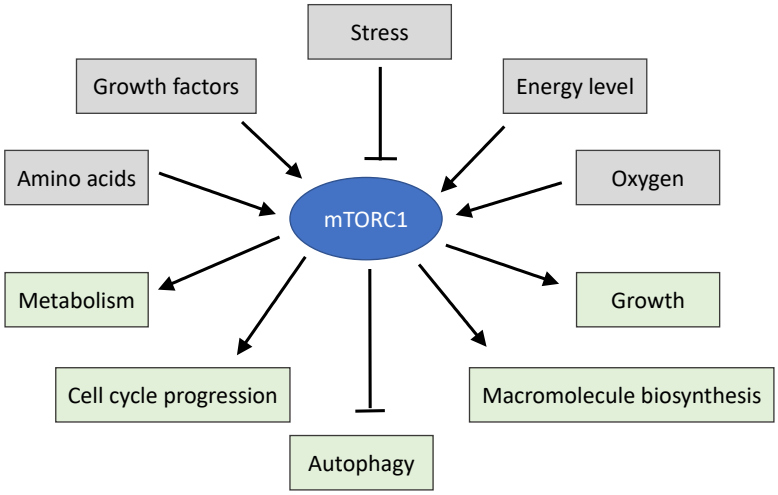


Figure 4.1 Illustration of environmental factors affecting MTORC1 and downstream events of activated MTORC1 (modified figure from(63) permission has been retrieved from [Elsevier]).

The main site of activation for MTORC1 is the lysosomal membrane (Figure4.2)(58, 69-71), and in the presence of amino acids this complex is recruited to the lysosomal membrane(Figure 9)(72). MTORC1 is inhibited through the interference of GATOR1 (GTPase-activating protein activity toward Rags) complex (73). GATOR1 needs to be recruited to a lysosomal surface for inhibiting MTORC1 in the absence of nutrients. This transmission is performed by SZT2, which is part of a larger protein-complex

named KICSTOR(composed of four proteins KPTN, ITFG2, C12orf66, and SZT2) (73, 74). Therefore, deficiency in SZT2 may result in no recruitment of GATOR1 on the lysosome and finally abolishing the inhibition of MTORC1 in fasting conditions (73). As a result, independent of the presence of amino acids, MTORC1 is constitutively active and located on a lysosomal membrane(72) (Figure 4.2). A validated SZT2 antibody is required to further study these aspects.

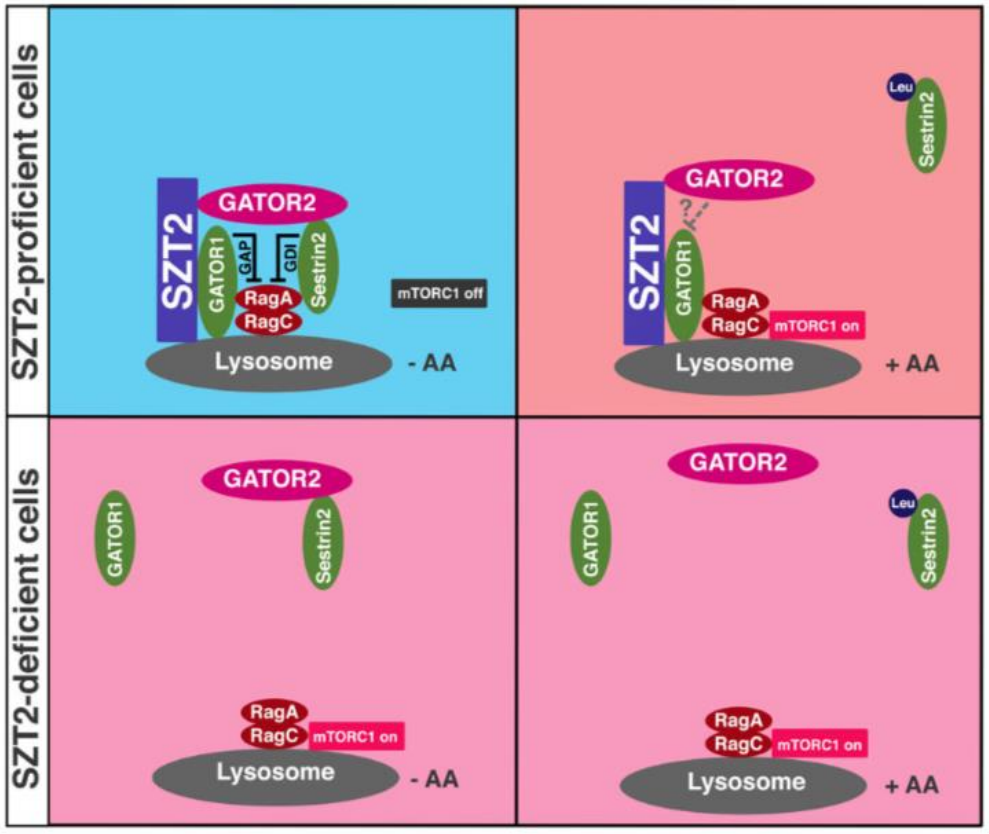


Figure 4.2 mTORC1 localization in SZT2-deficient and proficient cells in the presence and absence of amino acids (figure obtained from (73) with permission from [Springer Nature]).

4.3 Do the genetic variants affect gene expression?

Total mRNA from fibroblast from the patient and 4 healthy controls were sequenced to detect differentially expressed (DE) genes between the patient and the controls. Analyzing DE genes may identify genes that are dysregulated and being involved in development of the phenotype of the patient. Therefore, assessing the biological pathways involving the DE genes may allow the identification of genes that may be affected contributing in the patient's conditions.

Pathway analyses on a detected DE genes list showed enriched pathways (Table 3.3), which indicated that signaling through receptor tyrosine kinases (RTK) may be affected. Epiregulin (*EREG*)

gene is one of the down-regulated genes in cells from the patient compared to controls. EREG is a member of the epidermal growth factor (EGF) family(75). Fibroblast growth factor7 (FGF7), which is likewise down regulated in the patient's fibroblast, and EGF are ligands for the two RTKs fibroblast growth factor receptor (FGFR) and epidermal growth factor receptor (EGFR)(76). Binding the ligand (growth factor) to its RTK, through autocrine mechanisms(77, 78), leads to the activation of down-stream signaling pathways including RAS/RAF/MEK/ERK and PI3K/AKT/MTOR, and it results in the regulation of several cellular events, such as cell proliferation, migration, apoptosis inhibition, etc.(79-83). The SHC adaptor protein3 (SHC3), which mediates the transduction of cell signaling through RTKs similarly is down regulated in the patient's cells (84) (Figure 4.3).

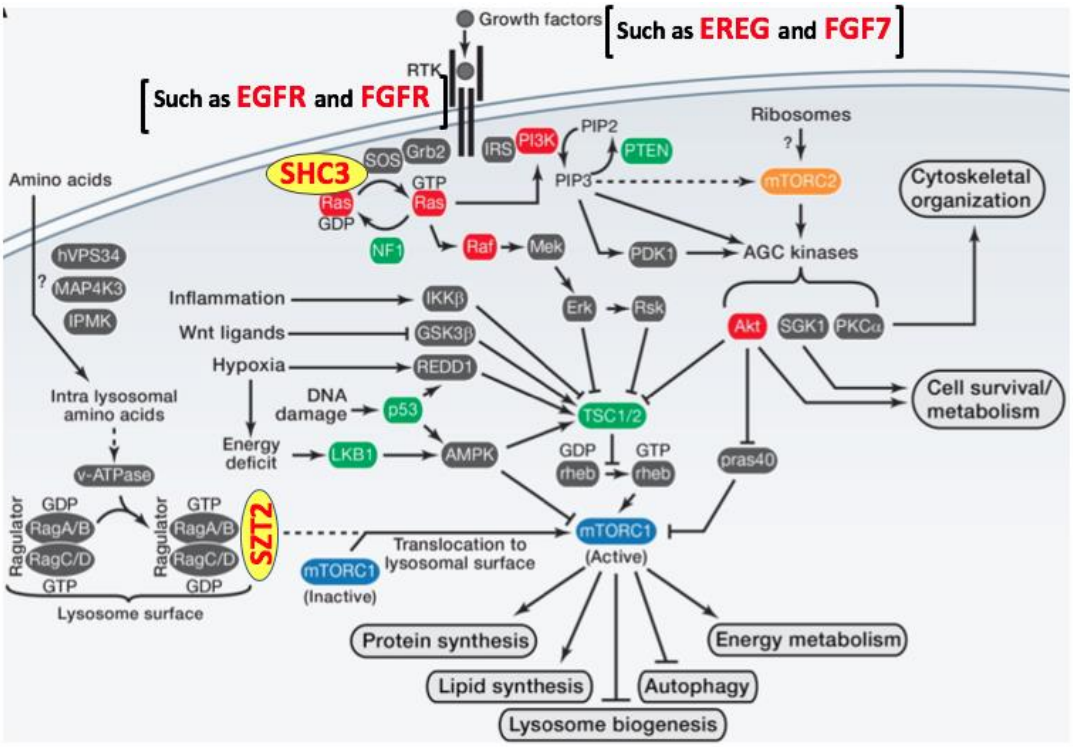


Figure 4.3 Illustration of the MTOR cell signaling pathway, showing SZT2, SHC3, FGF7, EREG, FGFR and EGFR (receptor of EREG). Further details about SZT2 are described in Figure 4.2 (figure is modified from (63). Permission has been obtained from [Elsevier].

As described in section 4.2, SZT2 deficiency may result in constitutive activation of MTORC1 independent of nutrient state. On the other hand, down regulation in the EREG, FGF7 and SHC3 may inhibit the initiation of cell signaling balancing the results of constitutively activated MTORC1. Therefore, these observations suggest that cell regulation mechanisms in the patient's cells may attempt to reduce the activation of MTORC1, which is constitutively active during fasting, due to the deficient SZT2.

One challenge in pathway enrichment analysis is the insufficient annotation knowledge in data bases. Alternative processing on primary RNA transcripts from a single gene may lead to the production of mRNA and protein isoforms, which can differ in function, localization, structure, etc. (85). Current data bases only specify whether a gene is active in a particular pathway, irrespective of the gene-related type of isoform(s) associated to that pathway. Therefore, pathway enrichment analyses may lead to detection of some enriched pathways which are not necessarily related to the identified isoform(s), but they might be resulted from contribution of other isoform(s) of the same gene. In addition, some transcript variants are exclusively expressed in specific tissues, consequently lack of the knowledge about tissue-specific gene expression in data bases may result in a misinterpretation of data(86).

Limitations in the transcriptome analyses in this work consist of: i) The RNA sequencing was performed on the mRNA extracted from fibroblast cells, whereas the patient's clinical features mainly involved the nervous system. ii) In our study limited number of samples may also have an effect in the results, in a way that small differences among the patient and 4 controls may be neglected by chance due to the small number of samples. Therefore, some DE genes may have been detected purely by chance. For example, the HLA-DRA, HLA-DRB1 and CD74 genes may have been detected in the list of DE genes irrespective of the patient's phenotype. Additional patient samples would significantly improve these analyses.

5 Conclusion

We aimed to characterize genetic causes of the clinical features of the patient with a neurological disorder. Whole exome sequencing (WES) was performed to find putative disease-causing variants. The main outcomes in this work are as follows:

- Identification of potential pathogenic compound heterozygous *SZT2* Chr1: g.43902997 A>T and Chr1: g.43905659 G>A variants which are maternally and paternally inherited, respectively.

- The Chr1: g.43905659 A>T variant was shown to cause skipping of SZT2 exon 42, causing an in-frame deletion in the mature transcript.

- Patient transcriptome analyses may indicate that the RAS/RAF/MEK/ERK cascade was affected in the patient's cells, which acts upstream of MTORC1, in a way to avoid increasing the activation of MTORC1. However, further studies are needed to verify these observations.

- The MTOR antibody has been validated which can be used in future studies to investigate the MTOR protein in patients' cells.

The results of this work could not verify the pathogenicity of the SZT2-variants, so further studies are required to document the cellular effects of the variant in the patient cells.

6 Future aspects

- Analyze MTOR localization on glucose starved patient cells compared to re-stimulated cells with glucose using immunofluorescence staining. The optimization in the amount of glucose being sufficient for cells to remain alive and grow, and at the same time being sufficient to induce inactivation of MTORC1 is necessary to be defined in such experiment.

- The transcriptome investigation in case of having cells from more patients with putative pathogenic SZT2 variants, in order to expand the number of patient's mRNA samples, as well as healthy control samples, to obtain more reliable result for documenting the effects.

- Validate a new SZT2 antibody and investigate the cellular localization of SZT2 in the patient cells.

- Study the MTOR antibody co-localization with lysotracker, since lysotracker penetrates into and detects organelles according to their pH rather than by binding to a specific protein, it can be beneficial to detect lysosomes and explore localization of MTOR based on that.

7 Supplementary data

Table 7.1 List of DE genes

Gene name	Adj. P val.	Gene name	Adj. P val.	Gene name	Adj. P val.
HLA-DRB1	$3{,}32E^{-10}$	HLA-DRA	$7{,}91E^{-03}$	DLX1	$6{,}93E^{-02}$
CD74	$2{,}94E^{-09}$	NMNAT2	$1{,}25E^{-02}$	HLA-DOA	$6{,}93E^{-02}$
HLA-DMB	$4{,}45E^{-09}$	JAG1	$1{,}56E^{-02}$	LOC100996609	$6{,}93E^{-02}$
SLCO2A1	$4{,}45E^{-09}$	MRVI1	$1{,}56E^{-02}$	SHC3	$6{,}93E^{-02}$
MAF	$1{,}35E^{-06}$	FLT1	$1{,}85E^{-02}$	EREG	$8{,}28E^{-02}$
KCNE4	$1{,}76E^{-06}$	FAM198B	$2{,}36E^{-02}$	G0S2	$8{,}37E^{-02}$
EBF2	$4{,}27E^{-05}$	PSORS1C1	$2{,}37E^{-02}$	RNF17	$8{,}37E^{-02}$
KCNJ8	$4{,}27E^{-05}$	SLC40A1	$2{,}47E^{-02}$	CLEC2A	$9{,}16E^{-02}$
ABCC9	$9{,}80E^{-05}$	FGF7	$3{,}04E^{-02}$	PSG8	$9{,}29E^{-02}$
MGST1	$2{,}04E^{-04}$	NES	$3{,}04E^{-02}$	MFGE8	$9{,}82E^{-02}$
HAPLN1	$3{,}91E^{-04}$	PSG3	$3{,}30E^{-02}$	MEGF6	$1{,}05E^{-01}$
PCDH7	$3{,}91E^{-04}$	KRTAP1-5	$3{,}34E^{-02}$	PTGIS	$1{,}06E^{-01}$
RSPO4	$3{,}91E^{-04}$	GATA6	$4{,}82E^{-02}$	CADPS	$1{,}34E^{-01}$
MFAP5	$3{,}96E^{-04}$	FGD4	$5{,}23E^{-02}$	GABBR2	$1{,}34E^{-01}$
EDIL3	$2{,}45E^{-03}$	TXNIP	$5{,}35E^{-02}$	PITPNM3	$1{,}34E^{-01}$
PSG5	$2{,}86E^{-03}$	TGFB2	$5{,}67E^{-02}$	SLC7A11	$1{,}34E^{-01}$
HTATIP2	$3.47E^{-03}$	ZFPM2	$6{,}64E^{-02}$	CD36	$1{,}41E^{-01}$
CDH18	$5.39E^{-03}$	ABCA8	$6{,}93E^{-02}$	COL5A2	$1{,}41E^{-01}$
				NDFIP2	$1{,}47E^{-01}$

8 References

1. Brown TA. The Human Genome. Genomes. 2nd ed. Oxford: Wiley-Liss; 2002.

2. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. Science. 2001;291(5507):1304-51.

3. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature. 2004;431(7011):931-45.

4. Nussbaum RL , Mclnnes R, Willard HF. Thompson & Thompson Genetics in medicine. 8th ed2016.

5. Turnpenny P, Ellard S. Emery's elements of medical genetics. 13th ed: Elisevier Limited; 2007.

6. Lower SS, McGurk MP, Clark AG, Barbash DA. Satellite DNA evolution: old ideas, new approaches. Current opinon in genetics & development. 2018;49:70-8.

7. Garrido-Ramos M. Satellite DNA: an evolving topic. Genes. 2017;8(9).

8. Oliveira GA, Dantas JL, Oliveira EJ. Informativeness of minisatellite and microsatellite markers for genetic analysis in papaya. Genetica. 2015;143(5):613-31.

9. Vergnaud G, Denoeud F. Minisatellites: mutability and genome architecture. Genome research. 2000;10(7):899-907.

10. Okada N. SINEs: short interspersed repeated elements of the eukaryotic genome. Trends in ecology & evolution. 1991;6(11):358-61.

11. Dewannieux M, Esnault C, Heidmann T. LINE-mediated retrotransposition of marked Alu sequences. Nature genetics. 2003;35(1):41-8.

12. Batzer MA, Deininger PL. Alu repeats and human genomic diversity. Nature reviews genetics. 2002;3(5):370-9.

13. Boissinot S, Davis J, Entezam A, Petrov D, Furano AV. Fitness cost of LINE-1 (L1) activity in humans. Proceeding of national academy of sciences. 2006;103(25):9590-4.

14. Stauffer S, Gardner A, Ungu DA, López-Córdoba A, Heim M. Meiosis. 2018 Nov. In: labster virtual lab experiments: Basic biology [Internet]. Springer Spektrum, Berlin, Heidelberg.

15. Zarrei M, MacDonald JR, Merico D, Scherer SW. A copy number variation map of the human genome. Nature reviews genetics. 2015;16(3):172-83.

16. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. Nature Reviews Genetics. 2006;7(2):85-97.

17. Mills RE, Luttig CE, Beauchamp A, Tsui C, Pittard WS, Devine SE. An initial map of insertion and deletion (INDEL) variation in the human genome. Genome research. 2006;16(9):1182-90.

18. Mullaney JM, Mills RE, Pittard WS, Devine SE. Small insertions and deletions (INDELs) in human genomes. Human molecular genetics. 2010;19(R2):R131-6.

19. Jackson M, Marks L, May GH, Wilson JB. The genetic basis of disease. Essays Biochemistry. 2018;62(5):643-723.

20. Cavalli-Sforza LL, Bodmer WF. The genetics of human populations. Courier corporation; 1999.

21. Reece EA , Hobbins JC, Clinical obstetrics: the fetus & mother handbook. 3rd ed: Blackwell publishing Ltd; 2007.

22. Bolk S, Pelet A, Hofstra RM, Angrist M, Salomon R, et al. A human model for multigenic inheritance: Phenotypic expression in Hirschsprung disease requires both the RET gene and a new 9q31 locus. Proceedings of the national academy of sciences. 2000;97(1):268-73.

23. Lin X, Tang W, Ahmad S, Lu J, Cobly CC, Zhu J, et al. Applications of targeted gene capture and next-generation sequencing technologies in studies of human deafness and other genetic disabilities. Hearing Research. 2012;288(1-2):67-76.

24. Grada A , Weinbrecht K. Next-generation sequencing: methodology and application. The jurnal of investigative dermatology. 2013;133(8):e11.

25. Nolan D, Carlson M. Whole exome sequencing in pediatric neurology patients: clinical implications and estimated cost analysis. Journal of child neurology. 2016;31(7):887-94.

26. O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwarts JJ, Girirajan S, Karakoc E, et al. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nature genetics. 2011;43(6):585-9.

27. Sawyer SL, Schwartzentruber J, Beaulieu CL, Dyment D, Smith A, Chardon JW, et al. Exome sequencing as a diagnostic tool for pediatric-onset ataxia. Hum mutation. 2014;35(1):45-9.

28. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, et al. Targeted capture and massively parallel sequencing of 12 human exomes. Nature. 2009;461(7261):272-6.

29. Ghaoui R, Cooper ST, Lek M, Jones K, Corbett A, Reddel SW, et al. Use of whole-exome sequencing for diagnosis of limb-girdle muscular dystrophy: outcomes and lessons learned. Jama neurology. 2015;72(12):1424-32.

30. Lee H, Deignan JL, Dorrani N, Strom SP, Kantarci S, Quintero-Rivera F, et al. Clinical exome sequencing for genetic identification of rare Mendelian disorders. Jama. 2014;312(18):1880-7.

31. Zhu X, Petrovski S, Xie P, Ruzzo EK, Lu YF, McSweeney KM, et al. Whole-exome sequencing in undiagnosed genetic diseases: interpreting 119 trios. Genetics in medicine. 2015;17(10):774-81.

32. O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature. 2012;485(7397):246-50.

33. Veltman JA, Brunner HG. De novo mutations in human genetic disease. Nature reviews genetics. 2012;13(8):565-75.

34. Belkadi A, Bolze A, Itan Y, Cobat A, Vincent QB, Antipenko A, et al. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. Proceedings of the national academy of sciences. 2015;112(17):5473-8.

35. Gilissen C, Hehir-Kwa JY, Thung DT, van de Vorst M, van Bon BW, Willemsen MH, et al. Genome sequencing identifies major causes of severe intellectual disability. Nature. 2014;511(7509):344-7.

36. Turner TN, Hormozdiari F, Duyzend MH, McClymont SA, Hook PW, Iossifov I, et al. Genome sequencing of autism-affected families reveals disruption of putative noncoding regulatory DNA. The American journal of human genetics. 2016;98(1):58-74.

37. Zhang F, Lupski JR. Non-coding genetic variants in human disease. Human molecular genetics. 2015;24(R1):R102-10.

38. Huang Y, Yu S, Wu Z, Tang B. Genetics of hereditary neurological disorders in children. translational pediatric. 2014;3(2):108-19.

39. Thapar A, Cooper M, Rutter M. Neurodevelopmental disorders. The Lancet Psychiatry. 2017;4(4):339-46.

40. Maulik PK, Mascarenhas MN, Mathers CD, Dua T, Saxena S. Prevalence of intellectual disability: a meta-analysis of population-based studies. Research in developmental disabilities. 2011;32(2):419-36.

41. Hirsch E. Childhood epilepsy syndromes with both focal and generalized seizures. Acta Neurologica Scandinavica. 2005;112(s118):52-6.

42. Polanczyk GV, Salum GA, Sugaya LS, Caye A, Rohde LA. Annual research review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. Journal of child psychology and psychiatry. 2015;56(3):345-65.

43. Simonoff E, Pickles A, Charman T, Chandler S, Loucas T, Baird G. Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity, and associated factors in a population-derived sample. Journal of the American academy of child & adolescent psychiatry. 2008;47(8):921-9.

44. Palop JJ, Chin J, Mucke L. A network dysfunction perspective on neurodegenerative diseases. Nature. 2006;443(7113):768-73.

45. Sanger TD. Pathophysiology of pediatric movement disorders. child neurology. 2003;18(1):S9-S24.

46. Orstavik KH, Strømme P, Ek J, Torvik A, Skjeldal OH. Macrocephaly, epilepsy, autism, dysmorphic features, and mental retardation in two sisters: a new autosomal recessive syndrome? Journal of medical genetics. 1997;34(10):849-51.

47. Raymaekers M, Smets R, Maes B, Cartuyvels R. Checklist for optimization and validation of real-time PCR assays. Journal of clinical laboratory analysis. 2009;23(3):145-51.

48. Vigeland MD, Gjøtterud KS, Selmer KK. FILTUS: a desktop GUI for fast and efficient detection of disease-causing variants, including a novel autozygosity detector. Bioinformatics. 2016;32(10):1592-4.

49. Basel-Vanagaite L, Hershkovitz T, Heyman E, Raspall-Chaure M, Kakar N, Smirin-Yosef P, et al. Biallelic SZT2 mutations cause infantile encephalopathy with epilepsy and dysmorphic corpus callosum The American journal of human genetics.. 2013;93(3):524-9.

50. Venkatesan C, Angle B, Millichap JJ. Early-life epileptic encephalopathy secondary to SZT2 pathogenic recessive variants. Epileptic Disorders. 2016;18(2):195-200.

51. Falcone M, Yariz KO, Ross DB, Foster II, Menendez I, Tekin M. An amino acid deletion in SZT2 in a family with non-syndromic intellectual disability. PLoS One. 2013;8(12):e82810.

52. Tsuchida N, Nakashima M, Miyauchi A, Yoshitomi S, Kimizu T, Ganesan V, et al. Novel biallelic SZT2 mutations in 3 cases of early-onset epileptic encephalopathy. Clinical genetics. 2018;93(2):266-74.

53. Nakamura Y, Togawa Y, Okuno Y, Muramatsu H, Nakabayashi K, Kuroki Y, et al. Biallelic mutations in SZT2 cause a discernible clinical entity with epilepsy, developmental delay, macrocephaly and a dysmorphic corpus callosum. Brain and development. 2018;40(2):134-9.

54. Pizzino A, Whitehead M, Sabet Rasekh P, Murphy J, Helman G, Bloom M, et al. Mutations in SZT2 result in early-onset epileptic encephalopathy and leukoencephalopathy. American journal of medical genetics part A. 2018;176(6):1443-8.

55. Domingues FS, König E, Schwienbacher C, Volpato CB, Picard A, Cantaloni C, et al. Compound heterozygous SZT2 mutations in two siblings with early-onset epilepsy, intellectual disability and macrocephaly. Seizure. 2019;66:81-5.

56. Love MI, Anders S, Kim V, Huber W. RNA-Seq workflow: gene-level exploratory analysis and differential expression. F1000Research. 2015;4(1):1070.

57. Ulgen E, Ozisik O, Sezerman OU. pathfindR: an R package for pathway enrichment analysis utilizing active subnetworks. bioRxiv. 2018:272450.

58. Sancak Y, Bar-Peled L, Zoncu R, Markhard AL, Nada S, Sabatini DM. Ragulator-Rag complex targets mTORC1 to the lysosomal surface and is necessary for its activation by amino acids. Cell. 2010;141(2):290-303.

59. Srivastava S, Cohen JS, Vernon H, Barañano K, McClellan R, Jamal L, et al. Clinical whole exome sequencing in child neurology practice. Annals of neurology. 2014;76(4):473-83.

60. Ropers HH. On the future of genetic risk assessment. Journal community genetics. 2012;3(3):229-36.

61. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nature genetics. 2014;46(3):310-5.

62. Itan Y, Shang L, Boisson B, Ciancanelli MJ, Markle JG, Martinez-Barricarte R, Scott E, et al. The mutation significance cutoff: gene-level thresholds for variant predictions. Nature methods. 2016;13(2):109-10.

63. Laplante M, Sabatini DM. mTOR signaling in growth control and disease. Cell. 2012;149(2):274-93.

64. Jewell JL, Guan KL. Nutrient signaling to mTOR and cell growth. Trends in biochemical sciences. 2013;38(5):233-42.

65. Shimobayashi M. , Hall M. N. . Making new contacts: the mTOR network in metabolism and signalling crosstalk. Nature reviews molecular cell biology. 2014;15(3):155-62.

66. Kim SG, Buel GR, Blenis J. Nutrient regulation of the mTOR complex 1 signaling pathway. Molecules and cells. 2013;35(6):463-73.

67. Manzoni C, Mamais A, Dihanich S, McGoldrick P, Devine MJ, Zerle J, et al. Pathogenic Parkinson's disease mutations across the functional domains of LRRK2 alter the autophagic/lysosomal response to starvation. Biochemical and biophysical research communications. 2013;441(4):862-6.

68. Eskelinen EL, Tanaka Y, Saftig P. At the acidic edge: emerging functions for lysosomal membrane proteins. Trends in Cell Biology. 2003;13(3):137-45.

69. Sancak Y, Peterson TR, Shaul YD, Lindquist RA, Thoreen CC, Bar-Peled L, et al. The Rag GTPases bind raptor and mediate amino acid signaling to mTORC1. Science. 2008;320(5882):1496-501.

70. Bar-Peled L, Schweitzer LD, Zoncu R, Sabatini DM. Ragulator is a GEF for the rag GTPases that signal amino acid levels to mTORC1. Cell. 2012;150(6):1196-208.

71. Zoncu R, Bar-Peled L, Efeyan A, Wang S, Sancak Y, Sabatini DM. mTORC1 senses lysosomal amino acids through an inside-out mechanism that requires the vacuolar H(+)-ATPase. Science. 2011;334(6056):678-83.

72. Yao Y, Jones E, Inoki K. Lysosomal regulation of mTORC1 by amino acids in mammalian cells. Biomolecules. 2017;7(3).

73. Peng M, Yin N, Li MO. SZT2 dictates GATOR control of mTORC1 signalling. Nature. 2017;543(7645):433-7.

74. Wolfson RL, Chantranupong L, Wyant GA, Gu X, Orozco JM, Shen K, et al. KICSTOR recruits GATOR1 to the lysosome and is necessary for nutrients to regulate mTORC1. Nature. 2017;543(7645):438-42.

75. Cao W, Luo LL, Chen WW, Liang L, Zhang RR, Zhao YL, et al. Polymorphism in the EREG gene confers susceptibility to tuberculosis. BMC medical genetics. 2019;20(1):7.

76. Schlessinger J. Receptor tyrosine kinases: legacy of the first two decades. Cold Spring Harbor perspectives in biology. 2014;6(3).

77. Jing Q, Wang Y, Liu H, Deng X. FGFs: crucial factors that regulate tumour initiation and progression. Cell proliferation. 2016;49(4):438-47.

78. Auf G, Jabouille A, Delugin M, Guérit S, Pineau R, North S. High epiregulin expression in human U87 glioma cells relies on IRE1α and promotes autocrine growth through EGF receptor. BMC Cancer. 2013;13(1):597.

79. Herbst RS, Bunn PA. Targeting the epidermal growth factor receptor in non-Small cell lung cancer. clinical cancer research. 2003;9(16):5813-24.

80. Lund-Johansen M, Bjerkvig R, Humphrey PA, Bigner SH, Bigner DD, Laerum O. Effect of epidermal growth factor on glioma cell growth, migration, and invasion in vitro. Cancer research. 1990;50(18):6039-44.

81. Huang PH, Xu AM, White FM. Oncogenic EGFR signaling networks in glioma. Science signaling. 2009;2(87):re6-.

82. Dikic I. Mechanisms controlling EGF receptor endocytosis and degradation. Portland press l limited; 2003.

83. Zeng F, Harris RC. Epidermal growth factor, from gene organization to bedside. InSeminars in cell & developmental biology. 2014;28:2-11.

84. Pasini L, Lanfrancone L. SHC4 (SHC (Src homology 2 domain containing) family, member 4). Atlas of genetics and cytogenetics in oncology and haematology. 2011(8).

85. Matlin AJ, Clark F, Smith CW. Understanding alternative splicing: towards a cellular code. Nature reviews molecular cell biology. 2005;6(5):386.

86. Khatri P, Sirota M, Butte AJ. Ten years of pathway analysis: current approaches and outstanding challenges. PLoS computational biology. 2012;8(2):e1002375.