

**Genetic investigations in Hungarian
patients affected by amyotrophic
lateral sclerosis**

Summary of the Ph.D. thesis

Kornélia Tripolszki



Szeged

2017

**Genetic investigations in Hungarian
patients affected by amyotrophic
lateral sclerosis**

Summary of the Ph.D. Thesis

Kornélia Tripolszki M.Sc.

Graduate School of Clinical Medicine

University of Szeged

Supervisor:

Nikoletta Nagy MD, Ph.D.

Department of Medical Genetics

University of Szeged

Szeged

2017

LIST OF PUBLICATIONS

Publications providing the basis of the dissertation

- I. **Tripolszki K**, Csányi B, Nagy D, Ratti A, Tiloca C, Silani V, Kereszty É, Török N, Vécsei L, Engelhardt IJ, Klivényi P, Széll M, Nagy N. Genetic analysis of the SOD1 and C9orf72 genes in Hungarian patients with amyotrophic lateral sclerosis. *Neurobiol Aging* 2017. Published online: 2017.01.29. **IF: 5.153**
- II. **Tripolszki K**, Török D, Goudenege D, Farkas K, Sulák A, Török N, Engelhardt J, Klivényi P, Procaccio V, Nagy N, Széll M. High-throughput sequencing revealed a novel SETX mutation in a Hungarian ALS patient. *Brain Behav* 2017. Accepted manuscript **IF: 2.128**
- III. **Tripolszki K**, Nagy ZsF, Nagy D, Török N, Klivényi P, Engelhardt IJ, Vécsei L, Nagy N, Széll M. *ANG* and *TARDBP* mutations in Hungarian patients with amyotrophic lateral sclerosis. Manuscript in preparation.

Publications indirectly related to the subject of the dissertation

- I. Kinyo A, Valyi P, Farkas K, Nagy N, Gergely B, **Tripolszki K**, Torok D, Bata-Csorgo Z, Kemeny L, Szell M. A newly identified missense mutation of the EDA1 gene in a Hungarian patient with Christ–Siemens–Touraine syndrome. *Arch Derm Res* 2014; 306(1):97-100. **IF:2.270**
- II. Nagy N, Valyi P, Csoma Z, Sulak A, **Tripolszki K**, Farkas K, Paschali E, Papp F, Toth L, Fabos B, Kemeny L, Nagy K, Szell M. CTSC and Papillon–Lefèvre syndrome: detection of recurrent mutations in Hungarian patients, a review of published variants and database update. *Molecular Genetics & Genomic Medicine* 2014; 2(3):217-228.
- III. Vályi P, Farkas K, **Tripolszki K**, Sulák A, Széll M, Nagy N, Nagy K. Rekurrens európai misszensz mutáció egy magyar Papillon-Lefèvre szindrómában szenvedő családban. *Fogorvosi Szemle* 2014; 107(3):87-92.
- IV. Nagy N, Farkas K, **Tripolszki K**, Sulák A, Kemény L, Széll M. A cylindromatosis gén mutációi által okozott genodermatosisok. *Bőr Vener Szemle* 2014; 90:(5) 185-193.

- V. Sulak A, Toth L, Farkas K, **Tripolszki K**, Fabos B, Kemeny L, Valyi P, Nagy K, Nagy N, Szell M. One mutation, two phenotypes: a single nonsense mutation of the CTSC gene causes two clinically distinct phenotypes. *Clin Exp Dermatol* 2016; 41(2):190-195. **IF: 1.092**
- VI. **Tripolszki K**, Knox R, Parker V, Semple R, Farkas K, Sulák A, Horváth E, Széll M, Nagy N. Somatic mosaicism of the PIK3CA gene identified in a Hungarian girl with macrodactyly and syndactyly. *Eur J Med Genet* Apr;59(4):223-6. **IF: 1.81**
- VII. **Tripolszki K**, Farkas K, Sulák A, Duga B, Melegh B, Knox R, Parker V, Semple R, Kemény L, Széll M, Nagy N. Atypical neurofibromatosis type 1 with unilateral limb hypertrophy mimicking overgrowth syndrome. Accepted manuscript. *Clin Exp Dermatol*. Accepted on 28 June 2016. **IF: 1.315**

1 INTRODUCTION

1.1 *Amyotrophic lateral sclerosis*

1.1.1 Clinical symptoms

Amyotrophic lateral sclerosis (ALS; ORPHA803) is a fatal, neurodegenerative disorder characterized by the death of motor neurons in the brain, brainstem and spinal cord. The characteristic clinical course of ALS is a progressive loss of voluntary movement, with symptoms spreading to more distant locations, resulting in paralysis and death from respiratory failure.

Originally, the spectrum of adult motor neuron diseases (MND) included ALS with the combined degeneration of upper (cortical) and lower (pontobulbar and spinal) motor neurons, primary lateral sclerosis (PLS) with only upper motor neuron lesion and lower MND with the damage of pontobulbar and/or spinal motor neurons. In a recent work, Finsterer and Burgunder (2014) suggested that ALS can manifest clinically as a continuum ranging from exclusive impairment of the upper motor neurons to exclusive impairment of the lower motor neurons.

Familial forms account for about 10% of ALS cases, although higher levels have been reported in certain geographical regions. The lifetime risk for developing the disease is approximately 1/400. ALS has a very poor prognosis, with a median survival period of three years after the onset of the initial symptoms. There is no cure for this relentless disease, although the drug riluzole, which blocks the release of glutamate, has been shown to slow disease progression.

1.1.2 Genetic background

The genetics of ALS are of research interest because it helps to uncover the mechanism of cell death in ALS. Neuronal cytoplasmic protein aggregation and defective RNA metabolism show to be frequent pathogenic mechanisms involved in ALS. Regarding its genetic background, more than 20 major ALS genes have been implicated in the Mendelian ALS forms and further about 100 genes have been associated as predisposing factors with the non-Mendelian variants. ALS can be inherited in an autosomal dominant, autosomal recessive or X-linked manner.

Familial forms are mainly transmitted in a Mendelian pattern of autosomal dominant inheritance.

1.1.2.1 Superoxide dismutase 1 (*SOD1*)

Among the ALS causative genes, superoxide dismutase 1 (*SOD1*) is one of the most commonly mutated genes and accounts for approximately 12–23% of the familial and up to 7% of the sporadic ALS forms. *SOD1* gene encodes the Cu/Zn superoxide dismutase enzyme, which catalyzes the inactivation of superoxide into oxygen and hydrogen peroxide, providing antioxidant defense. To date, more than 170 mutations have been reported for *SOD1* in the Amyotrophic Lateral Sclerosis Online Genetics Database since the gene was firstly associated to ALS in 1993. *SOD1* mutations occur in all the five exons of the gene.

1.1.2.2 TAR DNA binding protein (*TARDBP*)

Human TDP-43 was discovered in a screen for transcriptional repressors of the transactive response (TAR) DNA binding element of the HIV-1 virus in 1995. Since the TDP-43 association in ALS was initially described in 2006 more than 40 mutations have been found in familial and sporadic ALS cases. The *TARDBP* gene contains 6 exons, and with one exception (p.D169G in exon 4), all the identified mutations are located in the last exon.

1.1.2.3 Angiogenin (*ANG*)

The *ANG* gene encodes angiogenin, a 123-residue, 14.1-kDa protein and it is located on chromosome 14q11.2. The ANG protein belongs to pancreatic ribonuclease superfamily, and it plays a role in rRNA biogenesis, cellular proliferation and has a crucial role in inhibiting protein translation by cleaving tRNA.

Greenway *et al.* (2004) identified *ANG* as a candidate gene for amyotrophic lateral sclerosis in Irish and Scottish populations. ALS caused by mutations in the *ANG* gene is named as ALS9, an autosomal dominant adult onset disease. To date, 29 variants have been described for *ANG* in the ALSod database.

1.1.2.4 Chromosome 9 open reading frame 72 (*C9ORF72*)

A frequently mutated ALS gene is chromosome 9 open reading frame 72 (*C9orf72*), which – in addition to the *SOD1* mutations – is now recognized as the main cause of familial and sporadic ALS. A hexanucleotide (GGGGCC) repeat expansion (RE) located in the non coding region of the gene that can reach up to 4400 units (normal range: 2-23 units) has been identified in patients with ALS and/or frontotemporal dementia. Apart from the repeat expansion a splice site mutation has been identified in the *C9orf72* gene.

1.1.2.5 Senataxin (*SETX*)

Mutations in the senataxin (*SETX*) gene have been identified at a lower frequency than in the *SOD1* gene in ALS. *SETX* encodes a helicase protein involved in DNA repair and RNA production.

1.1.2.6 Fused in sarcoma (*FUS*)

The fused in sarcoma (*FUS*) gene is also associated with the Mendelian forms of ALS. *FUS* is located on chromosome 16p11.2 and encodes for a 526 amino acid protein. *FUS* is a nucleoprotein that plays a role in transcription, splicing, and shuttling of RNA from the nucleus to the cytoplasm. Fused in sarcoma also acts as a transcriptional regulatory sensor of DNA damage signals and therefore is required in maintaining the integrity of the genome.

1.2 Aims

In my thesis, the primary aim was to summarize the results of my genetic investigations in a group of Hungarian patients (n=66) suffering from ALS. I investigated the contributions of six commonly mutated ALS genes, *SOD1*, *TARDBP*, *ANG*, *FUS*, *SETX* and *C9orf72*.

This study represents the first genetic screening of ALS in Hungary, which adds novel data to the genetic and phenotypic diversity of this disease.

2 PATIENTS AND METHODS

2.1 Patients

The unrelated patients (n=66) included in this study were recruited from the Department of Neurology, University of Szeged, Szeged, Hungary, between 2010 and 2016. All patients fulfilled the revised El Escorial and the Awaji-shima criteria for ALS. The investigation was approved by the Internal Ethical Review Board of the University of Szeged. Written informed consent was obtained from patients and healthy controls, and the study was conducted according to the Principles of the Declaration of Helsinki.

2.2 Methods

2.2.1 DNA isolation

Blood samples from all the enrolled individuals (n=66) were collected after the individuals had signed an informed consent document. Genomic DNA was isolated using a QIAamp DNA Blood Mini Kit (QIAGEN; Hilden, Germany).

2.2.2 Mutation Screening – Sanger Sequencing

The entire coding region and the flanking introns of the *SOD1*, *TARDBP* and *ANG* genes were amplified. Direct sequencing of the PCR products was performed on an ABI 3100 sequencer and compared with the wild-type gene sequences. Fragments with identified mutations were independently re-amplified and resequenced from both ends.

2.2.3 Repeat expansion analysis

2.2.3.1 Repeat-primed PCR and Fragment length analysis

A two-step protocol was followed for the detection of the GGGGCC hexanucleotide RE in the *C9orf72* gene. The presence of the GGGGCC RE was observed as a saw-tooth pattern with 6-base pair periodicity.

2.2.3.2 Genotyping

To determine whether the single individual carrying the GGGGCC RE identified in this study also carried the “risk” haplotype, I selected the rs3849942 variant to be used as a marker for the “risk” haplotype for the patient and control genotypes.

Rs3849942 genotyping was based on allelic discrimination assays using TaqMan chemistry following the manufacturer's instructions.

2.2.4 Targeted High-throughput sequencing

2.2.4.1 Library preparation and sequencing

Amplicons (n=56) were designed to cover the coding regions and the flanking introns of the investigated *FUS*, *SETX* and *C9orf72* genes, and amplicon libraries were prepared according to the Amplicon Library Preparation Manual for the Roche Junior 454 next generation sequencing system. Emulsion PCR and next generation sequencing were performed according to the manufacturers' protocols.

2.2.4.2 Bioinformatic analysis

To improve the efficiency of the Roche pipeline, composed of the Roche 454 GS Reference Mapper for mapping (on UCSC human reference genome hg19) and Amplicon Variant Analyzer (v2.5p1) for variant calling, an additional in house pipeline was used.

3 RESULTS

3.1 *SOD1* gene

The direct DNA sequencing approach identified four different mutations of the *SOD1* gene in 5 ALS patients: three known heterozygous missense mutations (c.43G>A p.Val14Met; c.272A>C p.Asp90Ala; c.435G>C p.Leu144Phe) and one novel mutation (c.275_276delAA, p.Lys91ArgfsTer8).

3.1.1 Novel mutation detected in the *SOD1* gene

The detected novel heterozygous mutation (c.275_276delAA, p.Lys91ArgfsTer8) is located in the fourth exon of the *SOD1* gene and led to a frameshift with the insertion of 8 novel amino acids and the formation of premature stop codon at the new amino acid position 99. The novel p.Lys91ArgfsTer8 *SOD1* mutation was not present in 110 healthy controls of Hungarian ancestry that I investigated neither it is represented in mutation databases.

3.1.2 Detected known mutations of the *SOD1* gene

The *SOD1* p.Leu144Phe mutation is located in the fifth exon of the gene and was identified in two female ALS patients.

The p.Val14Met mutation is located in the first exon of the *SOD1* gene and was present in an affected female patient who reported no family history of ALS. ALS symptoms first appeared at the age of 62.

The p.Asp90Ala mutation, which is the most prevalent *SOD1* mutation in Europe, is located in the fourth exon of the *SOD1* gene and was present in a female patient. Interestingly, the patient also carried the rs111273304 splice-donor variant (c.239+2T>A), which is of unknown significance, in heterozygous form. This patient had clinical features typically associated with this genotype, including a relatively long survival after onset.

3.2 *TARDBP* gene

Direct sequencing of the *TARDBP* gene revealed a previously described missense mutation in exon 6 (c.931A>G, p.Met311Val). The variant was first described by Lemmens *et al.* (2009) in a family of Belgian origin.

3.3 *ANG* gene

Mutation analysis of the angiogenin gene revealed two heterozygous mutations (c.3G>A, p.Met-24Ile; c.169C>T, p.Arg33Trp).

The p.Met-24Ile mutation is located in the signal peptide region of the protein, and it was initially described by Conforti *et al.* (2008) in a sporadic patient from Italy. This mutation was detected in the same Hungarian patient who carried the *SOD1* p.Val14Met mutation.

The second *ANG* mutation detected in this cohort (c.169C>T, p.Arg33Trp) affects the nuclear translocation signal of the angiogenin. The mutation (p.Arg33Trp) is not reported in the ALSod database and the only information about this variant was that according to the ExAc database it has been found in one individual (out of 121.410 individuals).

3.4 *C9orf72* gene

Analysis of the *C9orf72* gene identified GGGGCC RE in one out of 66 ALS patients. The average repeat number based on fragment-length analysis was 5 (range 2–17

repeats) in the remaining 65 patients, none of whom carried repeat expansion. The patient with the repeat expansion also carried the rs3849942 risk allele, which was previously described as a part of the Finnish “risk” haplotype.

3.5 *SETX* gene

A novel heterozygous missense mutation (c.791A>G, p.Asn264Ser) of the *SETX* gene was identified in a Hungarian female patient. The novel c.791A>G, p.Asn264Ser *SETX* mutation identified in this study was confirmed with direct sequencing. The mutation was absent in the patient’s clinically healthy 45-year-old son and in the age- and gender-matched healthy control individuals. The heterozygous p.Asn264Ser mutation affects the N-terminal region of the SETX protein and is located in a region which is conserved among mammals.

4 DISCUSSION

In this study, I analyzed *SOD1*, *TARDBP*, *ANG*, *C9orf72*, *SETX* and *FUS* genes in a cohort of 66 Hungarian ALS patients, including one single case with a reported familial history for the disease.

4.1 *Mutations in the SOD1 gene*

I identified a novel and likely disease-causing heterozygous frameshift mutation (p.Lys91ArgfsTer8) in the *SOD1* gene. Three other heterozygous recurrent *SOD1* missense mutations (p.Val14Met, p.Asp90Ala and p.Leu144Phe) were also detected. *SOD1* mutations were detected in 7.5%

The two-base-pair deletion of the p.Lys91ArgfsTer8 *SOD1* mutation leads to a frameshift with the formation of a premature stop codon after the insertion of eight novel amino acids, causing a severe truncation of the protein. This truncation abolishes the integrity of the intrachain C57–C146 disulfide bridge.

The p.Asp90Ala mutation is the most common *SOD1* mutation in Europe, and it can be inherited in either a dominant or recessive manner.

The p.Leu144Phe missense mutation, which is the most prevalent mutation in the Balkan region (ALSoD), was detected in a familial and in a sporadic case. The rare p.Val14Met mutation was detected in an apparently sporadic case in a female patient

with upper and lower motor neuron signs. None of the *SOD1* mutations detected in these ALS patients was identified in 110 healthy Hungarian controls.

4.2 Mutations in the *TARDBP* gene

A known disease-causing mutation (p.Met311Val) has been detected in the last exon of the *TARDBP* gene. With the exception of one mutation (p.Asp169Gly in exon 4), all the identified mutations are localized across the entire C-terminal sequence of TDP-43 with a tendency of clustering in specific positions. The p.Met311Val mutation is located in the C-terminal part of the protein, the glycine-rich region that may mediate interactions with other proteins, including heterogenous ribonucleoproteins.

4.3 Mutations in the *ANG* gene

Two heterozygous mutations were detected in the *ANG* gene (c.3G>A, p.Met-24Ile, p.Met1Ile* ; c.169C>T, p.Arg33Trp, p.Arg57Trp*).

The p.Met-24Ile mutation, located in the signal peptide region of the protein, was first described by Conforti *et al.* (2008). This mutation was detected in the same Hungarian patient who carried the *SOD1* p.Val14Met mutation. The Met-24I mutation affects the start codon (ATG) of the gene, which may influence the correct translation of the protein.

The p.Arg33Trp (p.Arg57Trp*) affects the nuclear translocation signal of the angiogenin. Amino acids ²⁹IMRRRGL³⁵ represent the nuclear localization signal and helps in nuclear translocation of angiogenin. The arginine residues ³¹RRR³³ are critical in governing the nuclear translocation activity of angiogenin. The R³³ is essential to nuclear translocation and residues ³¹RR³² modulate this process.

4.4 Repeat expansion of the *C9orf72* gene

C9orf72 repeat expansion was detected only in one patient of 66. The patient carrying the RE variant also carried the previously described rs3849942 risk allele in heterozygous form. The presence of the risk allele suggests that the common founder is also shared by the Hungarian patient with the RE. These results suggest that the frequency of RE observed in the Hungarian ALS patients (1,5%) is significantly lower than in Western European populations, further demonstrating that the frequency of genetic factors for ALS varies among different geographic regions.

4.5 Mutations of the SETX gene

A novel, disease-causing heterozygous missense mutation has been identified (p.N264S) in the *SETX* gene. To date, only 17 pathogenic *SETX* mutations have been implicated in the development of the autosomal dominant juvenile form, ALS4. The typical phenotype, present in the majority of the ALS4 patients, presents as a combination of lower and upper motor neuron impairments.

The identified novel p.Asn264Ser *SETX* mutation is associated with an unusual ALS phenotype characterized by lower motor neuron impairment only.

This study further widens the geographic range for the origin of disease-causing heterozygous missense mutations of the *SETX* gene, which have already been implicated in ALS in patients from different countries. To my knowledge, my study is the first demonstrating a novel *SETX* mutation in the Hungarian ALS population.

4.6 Mutations in the FUS gene

No mutations in *FUS*, were detected in the Hungarian patients.

5 SUMMARY

Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disease characterized by the degeneration of the motor neurons. Approximately 90% of ALS cases are sporadic the remaining 10% are familial. Due to the complex genetic background of the disease, the underlying disease-causing variant is rarely established for individual cases. More than 20 major ALS genes are implicated in the Mendelian ALS forms and about 100 genes are associated with increased susceptibility for the development of the disease.

The aim of this study was to investigate Mendelian-disease causing genes in a group of Hungarian ALS patients (n=66). Considering the onset, the symptoms and the course of the disease, six genes (superoxide dismutase, *SOD1*; TAR DNA binding protein, *TARDBP*; angiogenin, *ANG*; senataxin, *SETX*; fused in sarcoma, *FUS* and chromosome 9 open reading frame 72, *C9orf72*) were selected for mutation screening.

The patients (n=66) participating in this study were recruited from the Department of Neurology, University of Szeged. Genomic DNA was isolated from blood. *SOD1*, *TARDBP* and *ANG* genes were analysed by Sanger sequencing. The GGGGCC hexanucleotide repeat in *C9orf72* was analyzed by a two-step protocol, including a first step of polymerase chain reaction amplification using genotyping primers. The fragment length analysis was performed using GeneScan. Only samples presenting with a single peak/amplification product were further analyzed in the second step by repeat-primed polymerase chain reaction. Targeted high-throughput sequencing was used for the analysis of the coding regions of the *SETX*, *FUS* and *C9orf72* genes.

Direct sequencing revealed a novel heterozygous frameshift mutation (c.275_276delAA, p.Lys92ArgfsX8) and three recurrent missense mutations (p.Val14Met; p.Asp90Ala and p.Leu144Phe in two patients) in the *SOD1* gene in five patients. The novel heterozygous *SOD1* mutation (p.Lys91ArgfsTer8) is located in the fourth exon of the gene and led to a frameshift with the insertion of 8 novel amino acids and the formation of premature stop codon at the new amino acid position 99. A known missense mutation (p.Met311Val) was detected in the *TARDBP* gene, which is located in the C-terminal part of the protein, the glycine-rich

region that may mediate interactions with other proteins. Two heterozygous mutations were identified in the *ANG* gene (c.3G>A, p.Met-24Ile; c.169C>T, p.Arg33Trp). The p.Met-24Ile mutation was detected in the same Hungarian patient who carried the SOD1 p.Val14Met mutation, and is located in the signal peptide region of the protein. The second *ANG* mutation (c.169C>T, p.Arg33Trp) affects the nuclear translocation signal of the angiogenin. *C9orf72* repeat expansion was detected in one sporadic female patient. She also carried the previously described rs3849942 risk allele in heterozygous form. Targeted high-throughput sequencing revealed a novel heterozygous missense mutation (c.791A>G, p.Asn264Ser) in the *SETX* gene. The mutation affects the N-terminal region of the SETX protein and is located in a region which is evolutionary conserved among mammals.

In conclusion, I performed the first genetic analysis of *SOD1*, *TARDBP*, *ANG*, *C9orf72*, *SETX* and *FUS* genes in a cohort of Hungarian ALS patients. My study further widens the geographic range for the origin of disease-causing heterozygous missense and frameshift mutations of the *SOD1*, *TARDBP*, *ANG*, *C9orf72* and *SETX* genes, which have already been implicated in ALS patients from different countries of origin.

6 ACKNOWLEDGEMENTS

I would like to thank to Dr. Nikoletta Nagy for her great supervising activity.

I would like to thank to Prof. Dr. Márta Széll for the opportunity to perform the genetic investigations of this study in the molecular laboratory of the Department of Medical Genetics, University of Szeged.

I am grateful to Dr. József I Engelhardt, Dr. Péter Klivényi, Nóra Török and Dr. László Vécsei for the neurological examinations and for enrolling the patients into the study.

Special thanks to Dr. Antonia Ratti, Dr. Cinzia Tiloca and Dr. Vincenzo Silani (Department of Neurology and Laboratory of Neuroscience, IRCCS Istituto Auxologico Italiano, Milan, Italy) for providing the *C9orf72* repeat expansion positive samples.

I am extremely grateful to Dr. Bernadett Csányi for her help in the *C9orf72* repeat expansion analysis. I would like to thank to Dr. Éva Kereszty for the opportunity to perform the *C9orf72* repeat expansion analysis in the laboratory of the Department of Forensic Medicine, University of Szeged.

I would like to express my gratitude to Dr. Vincent Procaccio and Dr. David Goudenege (University of Angers, France) for their guidance and help in the bioinformatic analysis of the next generation sequencing data.

Special thanks to all my colleagues for their kind help at the Department of Medical Genetics, University of Szeged.