UNIVERSITÀ DELLA CALABRIA

Department of Pharmacy, Health and Nutritional Sciences

GRADUATE SCHOOL IN TRANSLATIONAL MEDICINE
CICLE XXXV

Diagnostic challenges for genetic approaches in Amyotrophic Lateral Sclerosis

Dott.ssa Benedetta Perrone

Scientific Disciplinary Sector: MED/03

Coordinator

Prof.ssa Stefania Catalano

CATALANO
STEFANIA
28.02.2023
16:49:23 UTC

Supervisor

Prof.ssa Francesca Luisa Conforti

FRANCESCA LUISA
CONFORTI
28.02.2023 11:01:01
GMT+00:00

Academic Year 2022/2023

Quiet people are those who have the loudest minds.

Stephen Hawking

Table of contents

ABSTRACT

Over the past years, our understanding of the genetic mechanisms involved in complex diseases, such as Amyotrophic Lateral Sclerosis, has increased dramatically. ALS is a fatal and devastating motor neuron disease for which there is no truly effective cure. In 1993, the first gene associated with ALS was identified (1). Since then, our knowledge of the genetic mechanisms of disease has expanded significantly. Diagnostic tools have followed these research insights and Sanger DNA sequencing has been routinely used for many years. The emergence of next-generation DNA sequencing (NGS) approaches in the same decade allowed high throughput approaches to DNA sequencing, enabling the identification of new genes and pathways that highlight the heterogeneity of ALS disease, providing exciting opportunities for the identification of biomarkers useful for patient stratification and helping the development of targeted therapies.

Despite our increased understanding of the mechanisms of this disease, the majority of patients remain undiagnosed, and the remaining cases have no successful treatments. The absence of an effective cure can be well explained by the complex and heterogeneous nature of ALS, with patients displaying distinct clinical characteristics and distinct molecular mechanisms. In this context, the molecular profiling of patients into clinically meaningful subgroups can be extremely valuable for the development of new precision diagnostics.

In this thesis project, we provide an overview on the genetic investigation of ALS patients using different diagnostic approaches highlighting the importance of each methodology and their integrative use for the study of the disease, with the aim of providing a more comprehensive characterization of patients useful for the development of new-targeted strategies in clinical practice and personalized medicine.

List of abbreviations

ALS: Amyotrophic Lateral Sclerosis

fALS: Familial ALS

sALS: Sporadic ALS

MND: Motor Neuron Disease

LMN: Lower Motor Neuron

UMN: Upper Motor Neuron

rEEC: revised El Escorial Criteria

ALSFRS-R: ALS Functional Rating Scale–Revised

FDA: Food and Drug Administration

SOD1: Superoxide Dismutase 1

C9orf72: Chromosome 9 Open Reading Frame 72

ATXN2: Ataxin2

FUS: Fused in Sarcome

AD: Autosomal Dominant

AR: Autosomal Recessive

TARDBP: TAR DNA-binding protein 43

CNS: Central Nervous System

NGS: Next-Generation Sequencing

dNTP: Dideoxynucleotide

MPS: Massive Parallel Sequencing

HTS: High-Throughput Sequencing

WES: Whole Exome Sequencing

WGS: Whole Genome Sequencing

TS: targeted sequencing

SNV: Single Nucleotide Variant

CNA: Copy Number Alteration

ATXN1: Ataxin1

ATXN2: Ataxin2

ANG: Angiogenin

VCP: Valosin-containing protein

VAPB: VAMP associated protein B

UBQLN2: Ubiquilin 2

SV: Structural Variant

VUS: Uncertain Significance Variant

SB: Southern Blotting

PCR: Polymerase Chain Reaction

OGM: Optical Genome Mapping

HMW – DNA: High Molecular Weight DNA

RP-PCR: Repeat-Primed PCR

FTD: Frontotemporal Dementia

PCR: Polymerase Chain Reaction

WES: whole exome sequencing

WGS: whole genome sequencing

eQTL: Expression Quantitative Trait Loci

GQ: Genotype quality

FDP: Flow space read depth

QUAL: Variant quality

STB: strand bias ratio

FAO: flow-space alternate allele observations

MAF: Minor Allele Frequency

eQTLs: Expression Quantitative Trait Loci

GTEx: Genotype-Tissue Expression

ExAC: Exome Aggregation Consortium

SB: Southern Blot

Chapter 1

BACKGROUND & AIM

BACKGROUND

Amyotrophic Lateral Sclerosis

Amyotrophic lateral sclerosis (ALS) is the most frequent disorder of Motor Neuron Disease (MND). The term MND indicates a group of different progressive neurodegenerative pathologies of different aetiology and clinical variability, but a common final event: the loss of upper and/or lower Motor Neurons (MNs; Upper Motor Neuron, UMN; Lower Motor Neuron, LMN). Upper motor neurons travel from the brain along the spinal cord and lower motor neurons branch outward to supply the muscles of the face, throat, arms, chest and legs. There are diseases within the spectrum that have great similarities in clinical features and pathological findings, or even both. In particular, clinical motor presentations exist on a continuum, with a spectrum that extends from LMN to UMN signs (Figure 1).

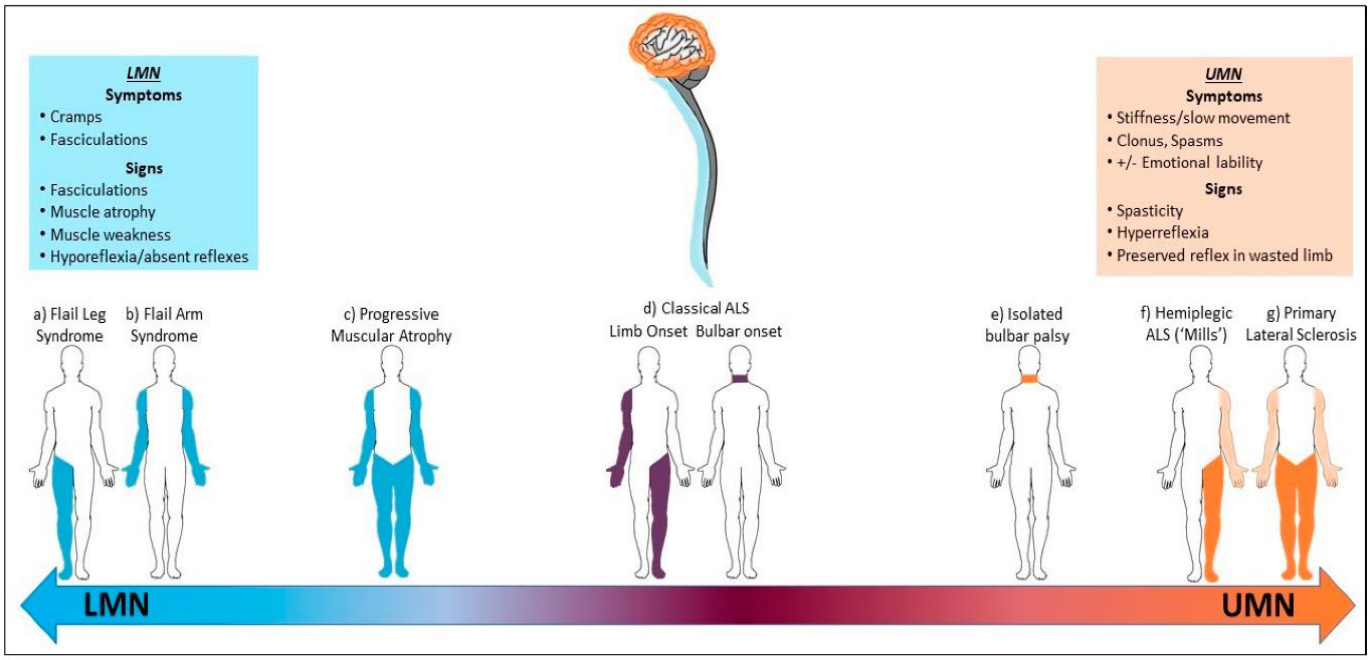


Figure 1. Pattern of motor involvement across the ALS clinical phenotypes. https://doi.org/10.3390/brainsci11060715

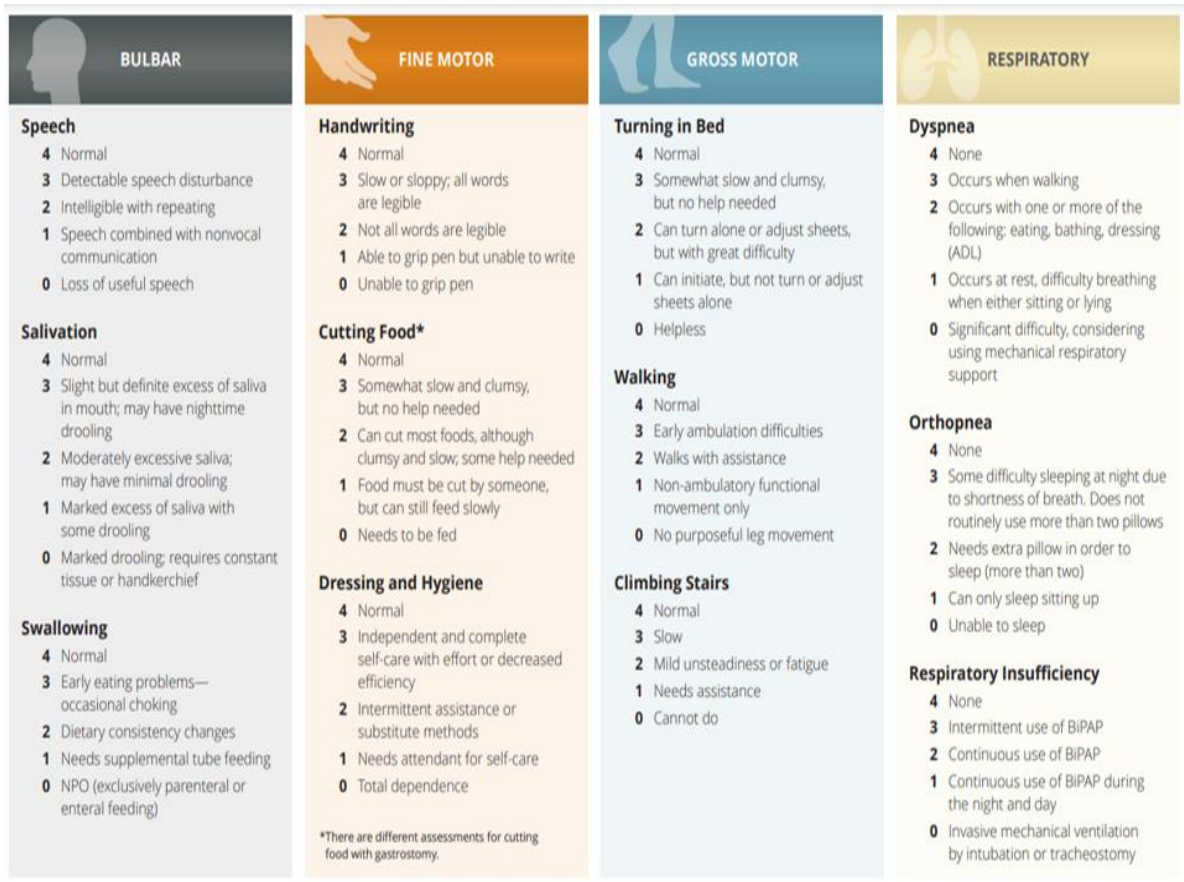
ALS is characterized by progressive degeneration of LMNs and UMNs in the cerebral cortex, brainstem, and spinal cord, resulting in muscle atrophy, fasciculation, and spasticity (https://raredisease.org). The first symptoms of ALS usually include muscle weakness or

stiffness in a limb or muscles in the mouth or throat (so-called bulbar muscles). Gradually almost all muscles under voluntary control are affected and individuals lose strength and the ability to speak, eat, move and even breathe (https://www.als.org). The onset and early progression of ALS are frequently insidious, and symptoms may go unrecognized and undiagnosed for up to 12 months. Precisely, the age of onset is between 40 and 65, with an average age of 55 at the time of diagnosis and most ALS people die of respiratory failure, usually within 3 to 5 years of the onset of the symptoms (https://www.als.org). However, about 10% of people with ALS survive 10 or more years.

Globally, ALS incidence is about 1-3 cases per 100,000 inhabitants per year and in Italy, at least 6,000 patients and 2,000 new cases are estimated each year (https://www.als.org/).

There are some specific criteria for the diagnosis of ALS known as the Escorial World federation of Neurology criteria (so called after a conference center in Spain, where they were developed), first published in 1994 (2) and revised in 2000 (3). Criteria were established because ‘‘variety of clinical features which may be present early in the course of ALS makes absolute diagnosis difficult and compromises the certainty of diagnosis for clinical research purposes and therapeutic trials” (2). The revised El Escorial Criteria (rEEC) classifies ALS patients into 4 levels of diagnostic certainty, namely clinically defined, probable, laboratory-supported probable, and possible. However, they were soon shown to have very low sensitivity in the early stages of the disease, especially in patients with bulbar-onset ALS (3). For this reason, other electrodiagnostic criteria (Awaji-Shima criteria) were introduced in 2006 (4) eliminating the probable category supported by the laboratory. The introduction of the latter criteria has shown an increase in diagnostic sensitivity compared to the El Escorial (2000) clinical diagnostic criteria. The increase in diagnostic sensitivity was nearly doubled for bulbar-onset ALS and clinically possible ALS (5) (6). However, both the clinical criteria of El Escorial (2000) and the electrodiagnostic criteria of Awaji-Shima (2006) are mainly used for the stratification of ALS patients into subgroups in order to justify their inclusion in large experimental multicenter clinics (7).

In addition, the highly variable symptom severity and rate of progression are determined by ALS Functional Rating Scale - Revised (ALSFRS-R) (8) (Figure 2).



BULBAR

Speech
- 4 Normal
- 3 Detectable speech disturbance
- 2 Intelligible with repeating
- 1 Speech combined with nonvocal communication
- 0 Loss of useful speech

Salivation
- 4 Normal
- 3 Slight but definite excess of saliva in mouth; may have nighttime drooling
- 2 Moderately excessive saliva; may have minimal drooling
- 1 Marked excess of saliva with some drooling
- 0 Marked drooling; requires constant tissue or handkerchief

Swallowing
- 4 Normal
- 3 Early eating problems—occasional choking
- 2 Dietary consistency changes
- 1 Needs supplemental tube feeding
- 0 NPO (exclusively parenteral or enteral feeding)

FINE MOTOR

Handwriting
- 4 Normal
- 3 Slow or sloppy; all words are legible
- 2 Not all words are legible
- 1 Able to grip pen but unable to write
- 0 Unable to grip pen

Cutting Food*
- 4 Normal
- 3 Somewhat slow and clumsy, but no help needed
- 2 Can cut most foods, although clumsy and slow; some help needed
- 1 Food must be cut by someone, but can still feed slowly
- 0 Needs to be fed

Dressing and Hygiene
- 4 Normal
- 3 Independent and complete self-care with effort or decreased efficiency
- 2 Intermittent assistance or substitute methods
- 1 Needs attendant for self-care
- 0 Total dependence

*There are different assessments for cutting food with gastrostomy.

GROSS MOTOR

Turning in Bed
- 4 Normal
- 3 Somewhat slow and clumsy, but no help needed
- 2 Can turn alone or adjust sheets, but with great difficulty
- 1 Can initiate, but not turn or adjust sheets alone
- 0 Helpless

Walking
- 4 Normal
- 3 Early ambulation difficulties
- 2 Walks with assistance
- 1 Non-ambulatory functional movement only
- 0 No purposeful leg movement

Climbing Stairs
- 4 Normal
- 3 Slow
- 2 Mild unsteadiness or fatigue
- 1 Needs assistance
- 0 Cannot do

RESPIRATORY

Dyspnea
- 4 None
- 3 Occurs when walking
- 2 Occurs with one or more of the following: eating, bathing, dressing (ADL)
- 1 Occurs at rest, difficulty breathing when either sitting or lying
- 0 Significant difficulty, considering using mechanical respiratory support

Orthopnea
- 4 None
- 3 Some difficulty sleeping at night due to shortness of breath. Does not routinely use more than two pillows
- 2 Needs extra pillow in order to sleep (more than two)
- 1 Can only sleep sitting up
- 0 Unable to sleep

Respiratory Insufficiency
- 4 None
- 3 Intermittent use of BiPAP
- 2 Continuous use of BiPAP
- 1 Continuous use of BiPAP during the night and day
- 0 Invasive mechanical ventilation by intubation or tracheostomy

Figure 2. ALS Functional Rating Scale–Revised (ALSFRS-R) Guide (https://www.alspathways.com).

Genetics of Amyotrophic Lateral Sclerosis

ALS is distinguished into familial (fALS, 5-10%) cases affects by monogenic disease and sporadic (sALS, 95-90%) cases determined by the interplay between environmental (i.e. muscle exercise, life, education and smoking) and genetic risk factors (9). Despite the list of ALS-related genes rapidly increasing, mainly due to the application of new technologies (10), many individuals do not receive a molecular diagnosis yet and, there is still substantial missing heritability. Therefore, a molecular diagnosis may be a valuable tool for dissecting out ALS heterogeneity and for identifying the causative molecular mechanisms of the disease and outlining precision therapeutic approaches.

There is currently no effective cure for ALS. Only two drugs, have been approved by the Food and Drug Administration (FDA) for treatment: Riluzole, a glutamate receptor antagonist,

which prolongs survival for a few months and, Edaravone, a free radical scavenger available in only a few countries and also a small effect on disease progression (11).

On the other hand, gene therapy clinical trials are currently underway for ALS patients with Superoxide dismutase (*SOD1)* - mutations, Chromosome 9 Open Reading Frame 72 (*C9orf72)* hexanucleotide repeat expansions, Ataxin 2 (*ATXN2)* trinucleotide expansions, and Fused in sarcoma (*FUS)* mutations, as well as sporadic disease without known genetic cause (12).

Notably, the pattern of ALS inheritance varies depending on the gene involved. Most cases show an autosomal dominant (AD) pattern. Less frequently and very rarely, ALS is inherited in an autosomal recessive (AR) and X-linked dominant mode, respectively (13). However, there are also instances of AR inheritance of AD genes in specific populations, both recessive and dominant mutations (i.e. p.D91A, *SOD1*), as well as evidence of reduced penetrance of disease-associated mutations (including p.I114, *SOD1* and G_4C_2, *C9orf72*) and oligogenic inheritance, illustrating that ALS is a highly complex genetic disorder (13). There are four "major" genes associated with the disease (*SOD1*, *C9orf72*, *TARDBP*, *FUS*) but through molecular genetic techniques increasingly applied to ALS research, over 40 genes have been discovered (Figure 3).

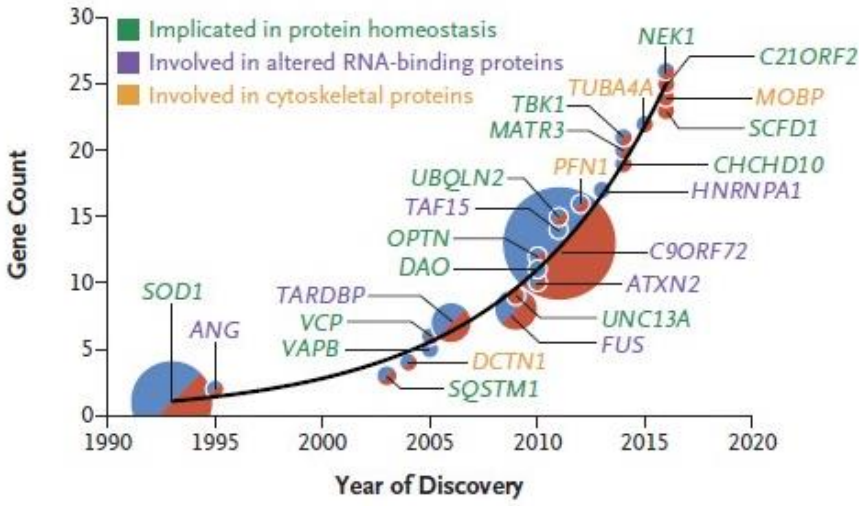


Figure 3. ALS Gene Discovery since 1990. PMID: 28700839.

Interestingly, mutations in *SOD1* are the second most recurrent known cause of ALS after *C9orf72* alterations and account for almost 20% of familial and 2-4% of apparently sporadic ALS cases (14). SOD1 mutations, mostly characterized by an autosomal dominant and highly penetrating inheritance pattern, determine a variable age of onset from the second decade (13). Despite rare cases of ALS with cognitive dysfunction have been reported, in general the clinical presentation of the *SOD1* mutation carriers is that of classical ALS (15) (Figure 1). Until now, more than 180 different mutations have been discovered in *SOD1* including missense mutations, deletion, insertion or splicing events (Figure 4).

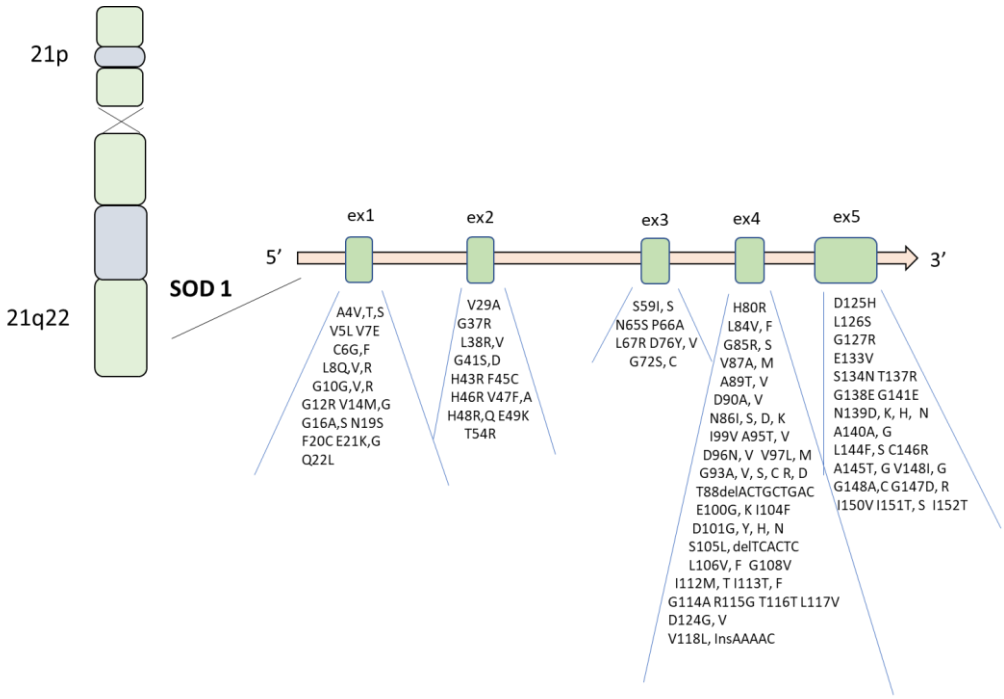


Figure 4. Schematic representation of SOD1 mutations since 2020. PMID: 32497448.

A different mutation causing the repeated anomaly of GGGGCC (G4C2) sequences is known in the *C9orf72* gene and is the most common genetic cause of the pathological spectrum of fALS (40%). Age of onset is extremely variable (20 to 90 years), commonly with repetitive expansion and, is generally related to expansion size (16). In addition to repeat size heterogeneity among *C9orf72* carriers, mosaicism and somatic instability may be part of the mechanisms that support the clinical variety observed in patients, as well as mechanisms anticipating the age of onset from generation to generation, usually associated with increased

disease severity. Indeed, in the same patient, it is possible to measure different dimensions of expansion in the Central Nervous System (CNS) and in the blood (17, 18).

Pathogenic mechanisms

Although the exact mechanisms underlying selective motor neuron death in ALS are still not fully understood, we know that an association of distinct mechanisms (glutamate excitotoxicity, oxidative stress, inflammation, abnormal protein aggregation, abnormalities in energy metabolism) rather than a single mechanism, provides neurodegeneration in ALS. S*OD1*, initially associated with oxidative stress and excitotoxicity in motor neurons in ALS, is an example among others, where studies of gene mutations have involved the protein in other mechanisms such as protein misfolding, insolubility and decay in disease. Despite the lack of knowledge of the pathogenetic mechanisms by which SOD1 works, it would appear that pathogenicity acts through a toxic increase in function (19), supported by the *SOD1* knockout mouse model (20) (21) and a loss of nuclear function (22), altering the antioxidant response. Furthermore, RNA dysregulation as a key pathological mechanism in ALS has been described following the identification of disease-causing variations in the RNA-binding protein, *TARDBP* and *FUS* genes. Moreover, the discovery of *C9orf72* hexanucleotide expansion also provided evidence of the crucial role of RNA processing for ALS. Pathogenic mechanisms associated with this variation include loss of function and gain of function of *C9orf72* caused by toxic RNA molecules and repeat protein toxicity from the dipeptide (23). The normal function of C9orf72 is unknown, but recent studies suggest that it may be important for endosomal trafficking.

Diagnostic approaches: advantages and limitations

Today, the diagnosis of ALS rests on clinical and neurophysiological findings. In research laboratories, genetic testing was limited to Sanger sequencing of a few genes, until not too long ago. This methodology disallowed diagnosis in patients with no family history of the disease who presumably carried mutations in new or undiscovered genes. Now, with the advent of new sequencing technologies (ALS multigene panels, whole exome sequencing and assays for *C9orf72* hexanucleotide expansion), there are several genetic testing options improving the chances of a definitive diagnosis in a considerable number of ALS cases (24). Genetic tests

can provide a molecular-level diagnosis in 70% of fALS cases and in 10% of sALS patients (13). However, family history pedigree analysis, and risk assessment remain crucial for the correct application of genetic tests and counselling for each patient. Therefore, a molecular diagnosis may help to understand the basis of their condition, allow for an accurate risk assessment and, provide participation in clinical trials but otherwise, does not affect treatment, and therefore is a complex ethical question.

Routine diagnostic testing includes *SOD1, TD-43, FUS* and, *C9orf72*, representing the most frequently mutated disease genes (major genes).

Briefly, in the first diagnostic step, i.e. before multigene sequencing approaches are applied, a repeat expansion in *C9orf72* should be excluded. If all major genes are negative, a second-level-genetic analysis can be performed, assessing other ALS-related genes (minor genes), according to the country-specific genetic epidemiology, if available (13).

Major genes testing

First-generation DNA sequencing methods

Sanger sequencing, commonly referred to as chain termination or dideoxynucleotide sequencing, Sanger sequencing, first emerged in 1977 and, has been the most widely used sequencing method for the last four decades (25). This "first generation" DNA sequencing method, represents the most reliable and accurate method for variants detection and is extensively used in ALS diagnosis.

Sequencing uses one strand of the double-stranded DNA as a template to be sequenced and is carried out using oligonucleotide primers and elongated by a mixture of deoxynucleotide triphosphates (dNTPs), which provide the needed arginine (A), cytosine (C), tyrosine (T), and guanine (G) nucleotides to build the new double-stranded structure. In addition, a small quantity of chain-terminating dideoxynucleotide triphosphates (ddNTPs) for each nucleotide is included, combined with fluorescent dyes with different emission wavelengths (13). Automated machines are able to reveal the DNA sequence from the intensity profiles of the four fluorophores by a fluorescence-sensing laser, from which a peak capable of detecting variants within the sequence is developed. Overall, this methodology is a robust testing strategy able to determine whether a point mutation or small/deletion/duplication is present. Sanger sequencing, although too laborious and expensive when compared with other multiplex testing

systems, remains routinely when sequencing of specific genes or gene fragments is required. Indeed, screening for known common mutations in cases of familial ALS is largely done by Sanger sequencing (13).

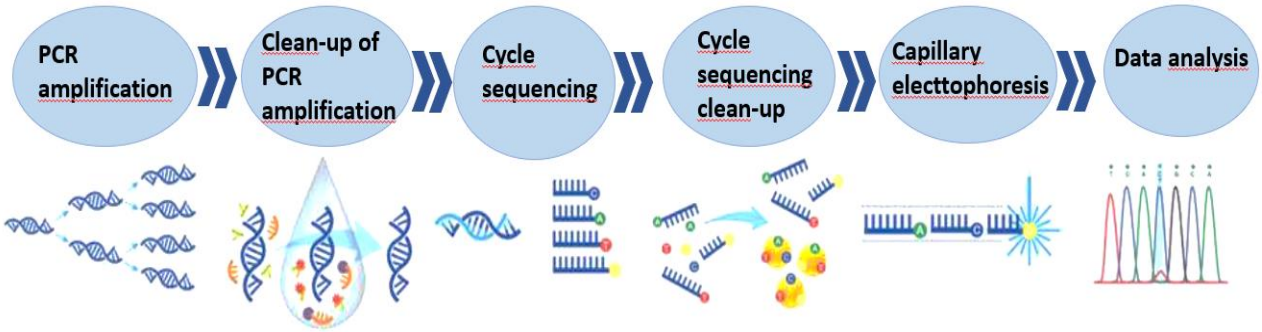


Figure 5. Sanger sequencing workflow from sample to data. (Source: https://www.thermofisher.com/).

Second-generation DNA sequencing methods

The last twenty years have been characterized by significant advances in the nucleic acid sequencing technologies field, collectively known as "Next-Generation Sequencing" (NGS) (26). These methodologies allow the identification and analysis of many genetic variations present in the human genome, at a high level of resolution and with extremely significant data throughput, offering a valid alternative method for the detection of disease alleles. NGS, also called Massively Parallel Sequencing (MPS) or high-throughput sequencing (HTS), provides to sequence target genes and regions, exome or whole genome, and is revolutionizing molecular diagnosis as it now permits large-scale parallel sequencing. This has led to a wider analysis of genes less frequently associated with ALS and to the discovery of patients with variants in more than one ALS-associated gene (oligogenic cases) (27).

Through NGS is possible to explore more genes at the same time and in more samples in the same experiment through a DNA marking method (barcoding), which in the computerized analysis phase, allows to distinguish and separate the sequences of each sample. Therefore, these technologies can produce large amounts of data with relative advantages in terms of time and cost compared to the Sanger method. In NGS, sequence reads need to cover each base many times to increase data accuracy while lower sequencing depth leads to a non-negligible amount of genotyping uncertainty (28). Individual sequencing read errors are

statistically insignificant when they are outnumbered by correct reads. Coverage is variable within a sample, and typical coverage ranges from 30 or fewer to >1000 reads for typical human genetics and cancer applications, respectively (https://irepertoire.com/). Another typical feature of NGS sequencing is the ability to accurately detect the presence of low-frequency alleles, allowing the identification of variations or deletions in heterozygosity.

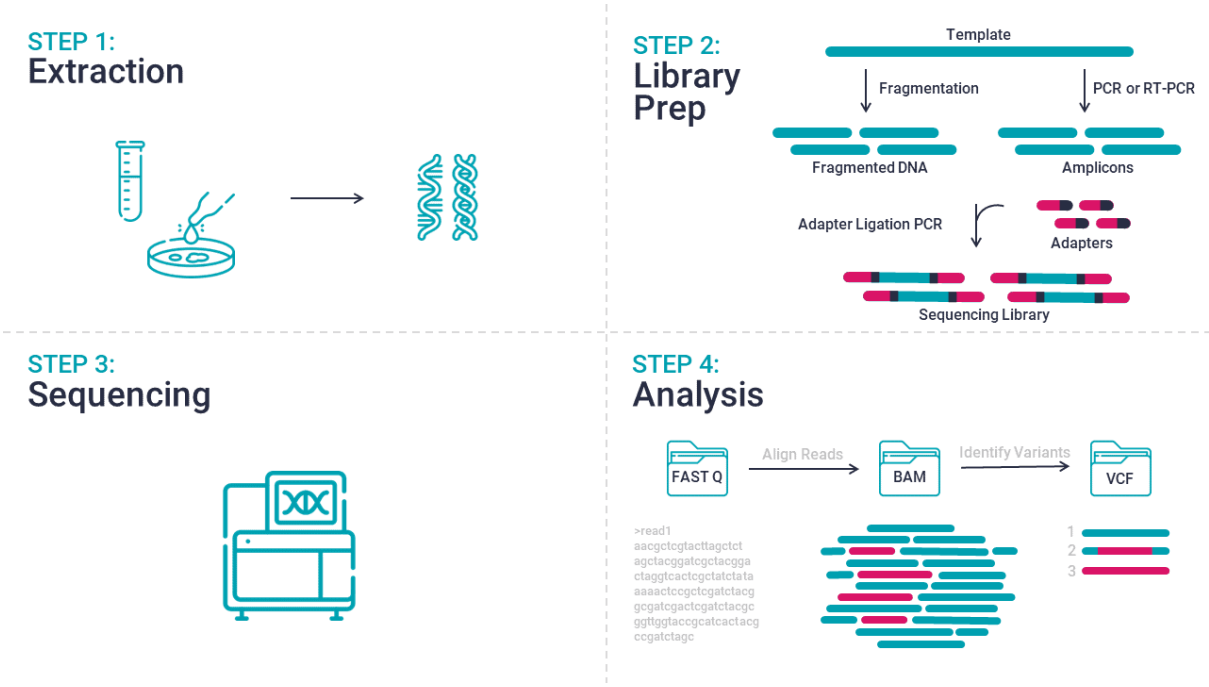


Figure 6. The general NGS workflow (https://irepertoire.com/).

Three main sequencing approaches, following the general workflow shown in figure 6, are used in clinical settings and indicated for the detection of rare variants and differ according to the size of the analysed portion of the genome (Figure 7). Whole Exome Sequencing (WES) involves the analysis only of the coding part of the genome (about 1.5% of the total) since it is widely demonstrated that at least 80% of pathogenic mutations are localized in exons. Whole Genome Sequencing (WGS) is used to sequence the entire genome and, to have complete genetic information, both of the coding and non-coding parts of the genome. WES execution is particularly indicated in cases with high genetic heterogeneity where an ever-increasing number of genes are involved, but each is responsible for a low percentage of cases. Targeted sequencing (TS) or Panel NGS testing, available in clinical services, allows sequencing key

genes or regions of interest at high depth (500-1000X or greater), allowing the identification of rare variants as well as providing convenient results for studies on the disease-related genes (29).

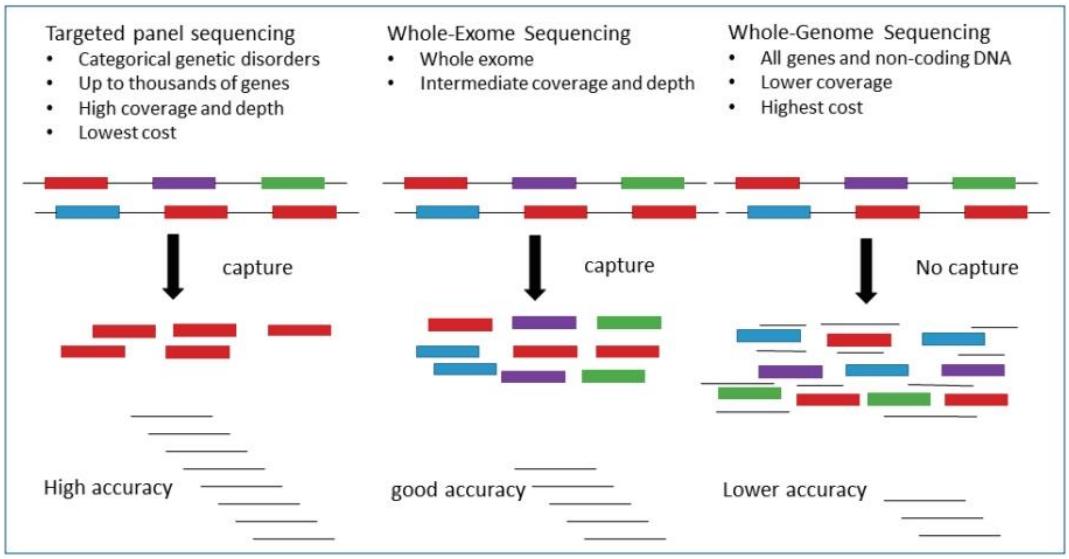


Figure 7. Comparison of targeted gene panel, whole exome sequencing, and whole genome sequencing approaches. PMID: 32024334.

Target panel analysis is usually applied for different heterogeneous groups of genetic disease, and the number of the analysed genes varies across laboratories (30). In addition to being cost-effective, using targeted panels is advantageous due to the high coverage in the region of interest and the high probability of detecting any mosaicisms. On the other hand, ALS targeted panel results show a high rate of variants of uncertain significance (VUS). In disorders, such as ALS, these variants need to be approached with caution in the clinical setting, as the complex genetic architecture may consist of combinations of genes variants that differ in frequency and penetrance. The number of ALS patients reported to have more than one ALS risk variant varies by studies but has been reported to be between 1 and 4% (27, 31). Despite the multiple advantages of next-generation sequencing techniques, Sanger sequencing has a fundamental place in clinical genomics both as a method to confirm the sequence variants identified and to fill the 'coverage gap' of poorly coverage regions by NGS (32).

However, integration of NGS and Sanger data is not always possible. Measures of sequence quality, depth of coverage, and allele frequencies are all values not directly and

precisely comparable between NGS and Sanger sequencing data (32). Also, NGS fails to identify most of the large insertions, deletions, and copy number changes in the 2/3 of the genome that is repetitive. Furthermore, NGS does not reliably detect balanced structural variants such as inversions and translocations.

C9orf72 test

Knowing the correct size of the most common repetitions in the population is essential in clinical diagnostics to guide the genetic test. Due to the complexity of *C9orf72* expansion with high GC content, large repeat size, and high insertion/deletion rate and sequence variation in flanking regions, molecular genetic analysis of the locus is challenging (33). Among the known techniques, Southern Blotting (SB) and repeat primed (RP)-PCR have always been used to detect size repeat large expansions. SB is considered the gold standard for expansion detection, but this test is onerous, and a high quality and quantity of DNA, often unavailable, is required for a single experiment (34). RP-PCR uses a locus-specific flanking primer along with a paired repeat primer that amplifies from multiple sites within the repeat, generating a characteristic ladder of fragments after capillary electrophoresis. Although this is a rapid and inexpensive method, it does not provide an accurate estimate of the number of repeats (>50-60 repeats) and does not reveal the actual DNA composition of the expansion (34). The incorrect diagnosis is also determined by the extensive somatic mosaicism at the C9orf72 locus implies that repeat lengths estimated from blood may not accurately reflect repeats in vulnerable brain tissue. Based on these premises and taking into account that traditional sequencing and NGS technologies have some limitations in detecting expanded repeats and/or mosaicism, it is necessary to investigate the potential of other techniques for the diagnosis of expansion disorders.

Optical genome mapping (OGM) developed by Bionano Genomics (Figure 8) is an imaging and mapping tool that can provide de novo maps and resolve complex repetitive regions, identify CNV, and elucidate whole-genome structural variations such as balanced/unbalanced translocations, inversions, and indels. It uses linearized strands of high molecular weight (HMW) DNA that are far longer than the DNA sequences analysed in current sequencing methods. This allows repetitive regions and other difficult-to-map regions to be more easily

extended than short molecules, detecting structural variants, ranging from 500 bp to megabase pairs in length, with sensitivity up to 99% and allele fractions down to 1 % (35).

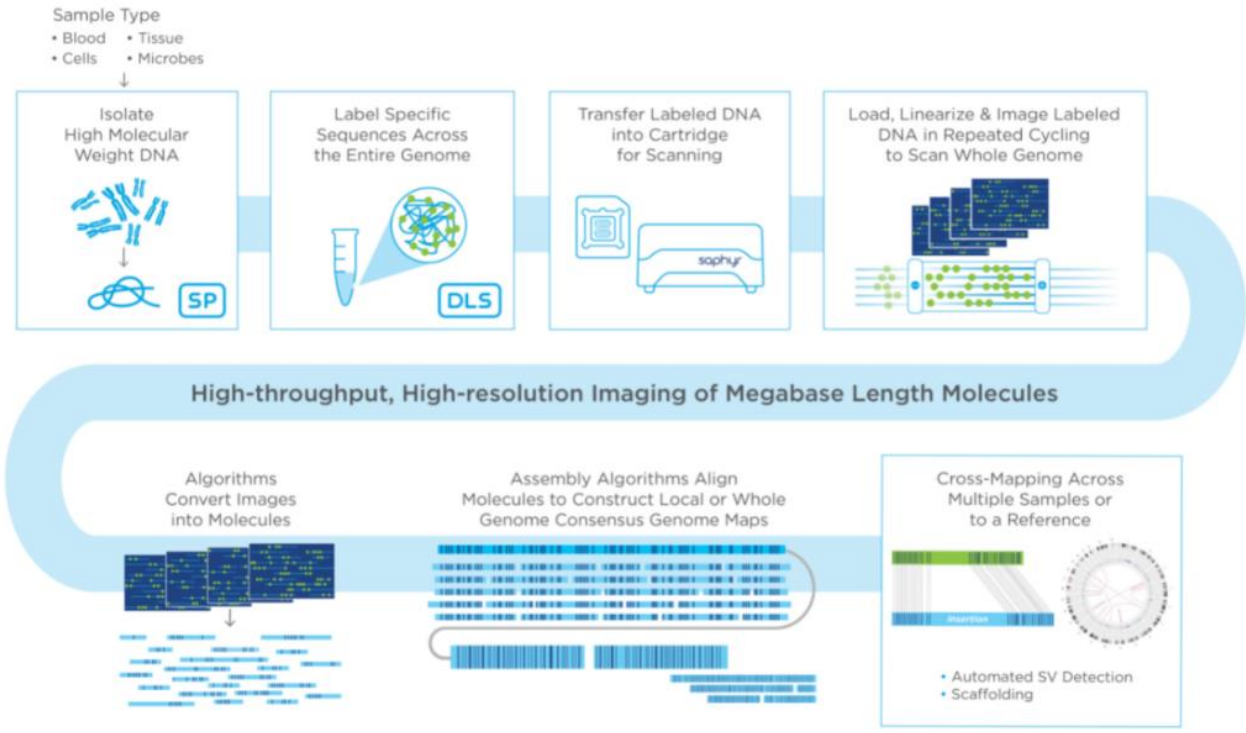


Figure 8. The Bionano Genomics Optical Genome Mapping workflow from mega-base size DNA to automatically detected, genome wide, structural variants using the Saphyr® Genome Imaging Instrument (https://bionano.com/).

AIMS OF THE PhD WORK

Given the multifactorial nature, researching ALS has proven challenging. In the last few years, genomics technologies, including next-generation sequencing (NGS), paved the way for achieving a greater understanding of the aetiology of the disease and hopefully one day the correct diagnosis, treatment, and prevention.

Aware of the importance of a detailed and accurate molecular investigation of the various genetic aberrations associated with complex diseases, in recent years our research group has been interested in the knowledge and use of different molecular genetic approaches to better identify causative genes for ALS and understand underlying altered mechanisms.

The present PhD project, aims to describe low-and high-throughput sequencing techniques applied to ALS diagnosis and, to evaluate the utility and the clinical use of the new whole genome-wide approach, the Optical Genome Mapping.

After a short introduction (*Chapter 1*), the specific topics discussed in the next pages can be summarized as follows:

In *Chapter 2,* we analysed a large cohort of ALS patients referred to our institution over the past 2 decades, by using Sanger sequencing technique.

In *Chapter 3* and *4*, using Next Generation Sequencing (NGS), we:

i) investigated the oligogenic nature of ALS in p.D91A- *SOD1* patients.

ii) analysed a new enrolled cohort of ALS patients.

The last Chapter is devoted to the investigation of the size of *C9orf72* expansions and to the occurrence of *C9orf72* mosaicism in ALS patients, by using Optical Genome Mapping (*Chapter 5*).

Chapter 2

Published in Neurobiology of Aging Volume 99, March 2021, Pages 99.e7e99.e14

Genetic investigation of amyotrophic lateral sclerosis patients in south Italy: a two-decade analysis

ABSTRACT

Amyotrophic lateral sclerosis (ALS) is a multifactorial disease characterized by the interplay of genetic and environmental factors. In the majority of cases, ALS is sporadic, whereas familial forms occur in less than 10% of patients. Herein, we present the results of molecular analyses performed in a large cohort of Italian ALS patients, focusing on novel and already described variations in ALS-linked genes. Our analysis revealed that more than 10% of tested patients carried a mutation in one of the major ALS genes, with C9orf72 hexanucleotide expansion being the most common mutation. In addition, our study confirmed a significant association between ALS patients carrying the *ATNX-1* intermediate repeat and the pathological C9orf72 expansion, supporting the involvement of this risk factor in neuronal degeneration. Overall, our study broadens the known mutational spectrum in ALS and provides new insights for a more accurate view of the genetic pattern of the disease.

1. Introduction

Amyotrophic lateral sclerosis (ALS) is an adult, fatal neurodegenerative disease affecting primarily both upper and lower motor neurons and leading to muscular denervation, atrophy, and, ultimately, paralysis of skeletal muscles. The incidence of ALS is reported to be around 5 cases per 100,000 population/year in most countries (Chiò et al., 2013; McCauley et al., 2019). Approximately 5%-10% of patients newly diagnosed with ALS report a positive family history and are classified as familial ALS (FALS), often with an autosomal-dominant pattern of inheritance, while the remaining 90%-95% of cases are considered as sporadic ALS (SALS) (Rowland et al., 2001). Although useful, there is a consensus that this classification is unreliable because of incomplete penetrance in family histories, unclear relatedness, early death of close relatives, and since every established FALS gene has also been implicated in SALS (Gibson et al., 2017; Turner et al., 2017). Indeed, an SALS patient with an FALS mutation is very often an FALS patient with a non- recognized family history.

Although more than 100 genes have been associated with ALS (https://alsod.ac.uk/), only a few of them are linked to a significant percentage of ALS cases. Together, *SOD1, C9orf72*, *TARDBP*, and *FUS* genes account for about 50% of FALS and 6% of SALS in the world (major genes), while frequencies of single gene mutations in other genes are very rare (≤ 1% of

patients) (Lamp et al., 2018; Müller et al., 2018; Zou et al., 2017). Nevertheless, many genetic variants not directly causing ALS could enhance susceptibility to the disease, modifying the clinical phenotype (Chiò et al., 2012a; Millecamps et al., 2012). Among these, the independent contribution of *ATXN1* and *ATXN2* as ALS risk factors has been proposed (Conforti et al., 2012; Elden et al., 2010), and the action of these genes supports the theory by which several variants strictly drive the interaction between genes, in a way that promotes disease onset and progression (Renton et al., 2014).

Achieving a detailed and accurate molecular investigation of the various genetic aberrations associated with ALS may help broaden our vision on the role of genetics in ALS pathogenesis. With this aim, here we present the results of our own experience in the molecular genetic testing of ALS-related genes performed in a large cohort of ALS patients referred to our institute during the past 2 decades.

2. Patients and methods

2.1 Patients

Informed consent was obtained from each study subject or from a close relative if the subject was too severely disabled to give written consent. Nine hundred ninety-seven patients of Italian descent, except for 2 single SALS cases done of French origin and the second of Arabian origin were prospectively and randomly recruited at the Institute of Neurological Sciences-CNR, Mangone (CS), and DNA samples were collected from January 1999 to December 2018. All patients underwent a full neurological evaluation to establish the clinical diagnosis of ALS according to the El Escorial criteria (Brooks et al., 2000) and the recently proposed guidelines for FALS classification (Byrne et al., 2012). A population characterized by 296 age- and sex-matched Italian individuals without neurodegenerative disease was used as control sample. The characteristics of ALS patients and controls are reported in Table 1.

Table 1. Main clinical characteristics of ALS patients and controls.

Overall subjects(n)	Gender (female), n (%)	Median age at onset[a]/inclusion[b] y (range)	Form (fALS, n (%)	Site of onset (bulbar), n (%)	FTD, n (%)
Patients (997)	421 (42.2%)	59.3 (17e89)	66 (6.6%)	182 (22%)	16 (1.9%)
Controls (296)	125 (42.2%)	60.5 (31e80)	-	-	-

Key: fALS, familial ALS; FTD, frontotemporal dementia.
[a] Data not available for site of onset for 174 patients. [b] Age at onset of ALS and age at inclusion for controls.

2.2 Genetic analysis

Mutational analysis of *C9orf72* (mendelian inheritance in man [MIM]: 614260), *SOD1* (MIM: 147450), *TARDBP* (MIM: 605078), *FUS* (MIM: 137070), *ANG* (MIM: 105850), *VAPB* (MIM: 605704), *VCP* (MIM: 601023), and *ATXN1* (MIM: 164400) was performed according to standard procedures. Purified amplicons were directly sequenced on an ABI Prism 3130XL genetic analyzer (Applied Biosystems, Foster City, CA), using the BigDye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems). Analysis of repeat expansion in *C9orf72* and *ATXN1* was performed using fluorescent-labeled primer PCR with capillary electrophoresis on an ABI Prism 3130XL genetic analyzer and analyzed with GeneMapper Software, version 4.0 (Applied Biosystems).

2.3 *Statistical analysis*

Statistical differences between mutated cases and ALS patients without genetic mutations were evaluated with a 2 tailed t-test for continuous variables (such as age at symptom onset) and chi-squared test for discrete variables (such as gender distribution, family history, and site of onset). We also assessed associations of polyQ repeats in *ATXN1* gene and *C9orf72* hexanucleotide repeat expansions in different groups of ALS patients using a chi-squared test. All p-values were computed using the R software (R Foundation for Statistical Computing) and adjusted using Welch's correction in a 2-tailed t-tests and Yates' continuity correction in a chi-squared tests. A p-value <0.05 was considered statistically significant.

3. Results

Molecular analyses revealed that more than 10% of tested patients carried a mutation of one of the major ALS genes, with C9orf72 hexanucleotide expansion being the most common mutation (Fig. 1). Table 2 summarizes demographic, clinical, and genetic data of ALS patients carrying pathogenic mutations in ALS-related genes. Thirty-seven of the 66 patients (56.1%) with fALS carried a mutation in one of the tested ALS genes. In contrast, only 66 (7.1%) of the 931 apparently sALS cases had a genetic mutation. Mean age at symptom onset was similar among patients carrying SOD1 mutations, *TARDBP* mutations, and C9orf72 hexanucleotide repeat expansion. In contrast, patients with *FUS* mutations manifested symptoms at a much younger age and this difference was found to be statistically significant (unpaired t-test with Welch's correction, $p = 0.0022$). There was a significant difference in gender, family history, and site of onset between C9orf72 positive and negative groups (Table 2), while no statistically significant difference was observed between patients with and without mutations in other ALS-linked genes. The distribution of the C9orf72 repeat in our cohort of ALS patients is shown in Fig. 2. All mutations detected, except for 2 in *SOD1* and *FUS* genes, were previously reported (Table 2). The new *SOD1* missense substitution D83V (according to human genome variation society nomenclature: c.251A>T, p. Asp84Val) was neither found in the Genome Aggregation Database (gnomad.broadinstitute.org/) nor in the 1000G database (http:// www.1000genomes.org/) and it was absent in our control subjects. Yet, the mutation was predicted to be pathogenic by 3 different programs, Sift (https://sift.bii.aestar.edu.sg/), Polyphen2 (http://genetics.bwh.harvard.edu/pph2), and MutationTaster (http://www.mutationtaster.org/). The patient carrying this mutation was a 44-year-old man without a relevant family history, who initially presented with a slowly progressive muscle weakness of the lower extremities with upper motor neuron signs. He showed mild dyspnea but no dysphagia, nor dysphonia. He suffered from mild dysarthria, worsened until he had difficulty with walking, showing impairment of lower motor neurons, and he died 2 years after disease onset. The novel *FUS* mutation Gly246DUPL (c.738_740dupl AGG), identified in exon 6 of the gene, was neither found in the abovementioned databases nor in silico analysis predicted the mutation as "disease-causing." The patient carrying this mutation is a 35- year-old obese man, who noticed fasciculation in both shoulders and arms, together with weakness and atrophy in the left proximal arm muscles. Disease duration is to date 48 months, with a relatively

slow progression. Bulbar functions are normal and he does not show dysphagia nor dysarthria. *TARDBP* analysis identified the previously described G376D mutation (Conforti et al., 2012), in a large family in which all affected individuals showed a rapid progressive disease. Interestingly, the reconstruction of the genealogic tree led us to a large collection of DNAs from family members, either affected or not, and the segregation analysis revealed a dominant pattern of transmission, even though the penetrance appeared incomplete (Fig. 4). The screening of VAMP/synaptobrevine associated membrane protein B (*VAPB*), angiogenin (*ANG*), and valosin containing protein (*VCP*) genes revealed no novel mutations but many single nucleotide polymorphisms and a previously described ANG gene mutation in 1 sporadic patient (Conforti et al., 2008). A list of single nucleotide polymorphisms related to all the genes analyzed in ALS patients is reported in Table 5. Finally, a cohort of 703 ALS patients (49 fALS and 654 sALS) underwent *ATXN-1* repeat analysis to evaluate the frequency of *ATXN-1* expansion in C9orf72 carriers. We considered 33 as the cut-off to discriminate between normal and intermediate repeats. Results showed that 10/51 C9orf72 positive cases (19.6%) had at least 1 allele with a polyQ repeat length 33, revealing a statistically significant association between *ATXN1* and *C9orf72* repeat expansions in ALS patients (fixed-effect model odds ratio = 2.28, 95% confidence interval = 1.12e4.7, p = 0.0446). In particular, 27.8% (5/18) of fALS patients and 15.15% (5/33) of sALS patients with C9orf72 repeat expansions showed 33 polyQ repeats in the *ATXN1*, with no significant differences between the 2 groups (fALS, p = 0.1123; sALS,
p = 0.3624).

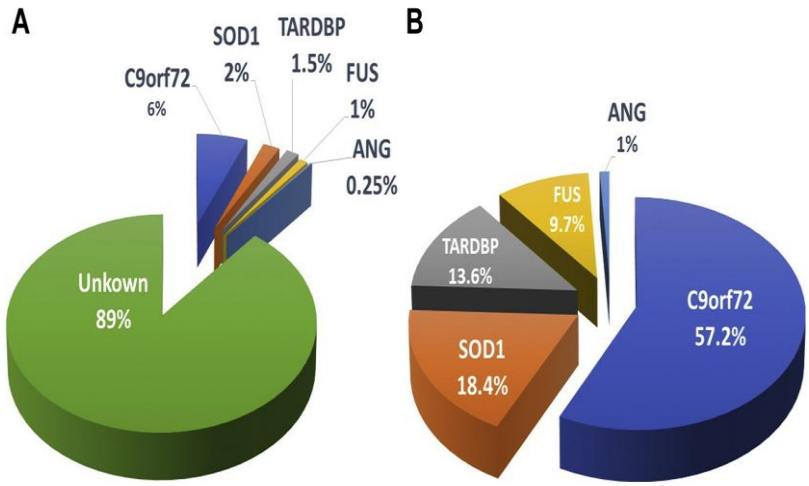


Figure 1. Percentage of mutations in ALS. (A) Distribution of mutated gene in whole ALS cohort and (B) prevalence of mutations in the 103 ALS cases. Abbreviations: ALS, amyotrophic lateral sclerosis.

Table 3. Descriptive statistics of amyotrophic lateral sclerosis patients with pathogenic C9orf72 expansion (C9orf72 positive) and without expansion (C9orf72 negative).

	C9orf72 positive (n=59)	**C9orf72 negative (n=919)**	**p-value**
Gender, n (%)			
Female	34 (57.63%)	382 (41.57%)	0.0224
Male	25 (42.37%)	537 (58.43%)	
Family history, n (%)			
fALS	22 (37.29%)	43 (4.68%)	<0.0001
sALS	37 (62.71%)	876 (95.32%)	
Site of onset, n (%)			
Bulbar	18 (36.73%)	163 (21.28%)	0.0190
Spinal	31 (63.27%)	603 (78.72%)	
Mean age at onset (y)	57.75	59.48	0.2247

Values in bold show statistically significant differences.
Key: fALS, familial amyotrophic lateral sclerosis; fALS, sporadic amyotrophic lateral sclerosis.

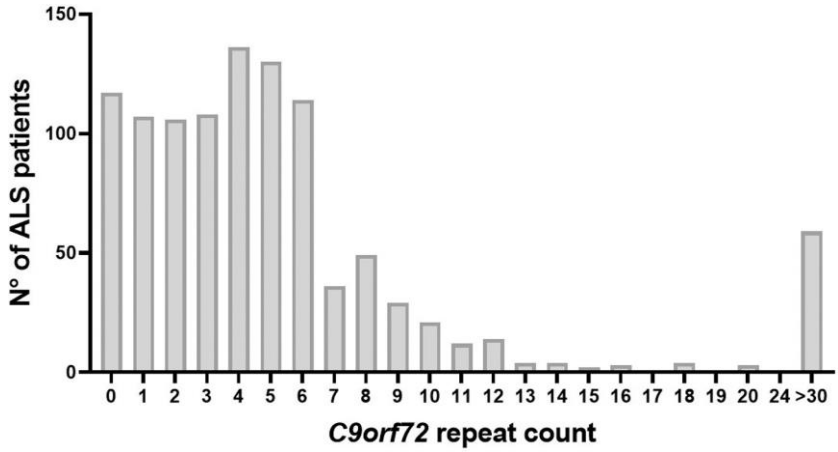


Figure 2. Histogram of C9orf72 repeat sizes in ALS patients (n=978). Total number of patients is shown as regards the repeat number on the x-axis. Abbreviations: ALS, amyotrophic lateral sclerosis.

4. Discussion

This study summarizes the results of genetic analyses of ALS-patients performed at the Institute of Neurological Sciences-CNR Mangone (CS) during the past 2 decades. A large cohort of patients was investigated using Sanger sequencing analysis of well-established ALS-related genes: *SOD1, C9orf72, TARDBP, FUS, ANG, VAPB, VCP, and ATXN1*. The highest frequency of positive cases was obtained in *C9orf72* (6%), followed by *SOD1* (2%), T*ARDBP*

(1.5%), and *FUS* (1%). In particular, C9orf72 repeat expansion analysis revealed the presence of the pathogenic intronic (GGGGCC)n repeat expansion in 22 of 65 FALS patients (33.8%) and 37 of 913 SALS patients (4%), confirming this mutation as the most frequent alteration in ALS Italian patients (Chiò et al., 2012a). Pathogenic C9orf72 repeat expansion frequencies vary greatly by ethnicity/geographic origin. The highest frequencies are reported in northern European countries (FALS 40% and SALS 8%), with low frequencies reported in Asian countries (FALS 2.3% and SALS 0.3%) (Cruts et al., 2015). In our cohort, patients carrying the expansion were more likely to be female, with a family history of disease and a bulbar-onset, which is consistent with previous findings (Majounie et al., 2012; Umoh et al., 2016). Moreover, 11 C9orf72-carrier patients showed clinical frontotemporal dementia (FTD) to primary diagnosis, confirming that expansions are commonly observed in patients with FTD/ALS (van Blitterswijk et al., 2013). SOD1 molecular investigation revealed a mutational frequency of 10% for FALS cases and 1.4% for SALS cases. These results are similar to the frequencies observe in population-based studies of ALS in Italy but are slighter lower than those reported in other countries, supporting a different geographic distribution for these mutations (Battistini et al., 2012; Chiò et al., 2012a; Conte et al., 2012; Lattante et al., 2012). We identified the novel heterozygous D83V missense mutation in an SALS patient. Segregation analysis in the patient's family revealed that it was present in the healthy father, suggesting the non-pathogenicity or the incomplete penetrance of this variation. Unfortunately, because the patient was unavailable for further study, it was not possible to confirm the predicted effects of the c.251A>T on the SOD1 protein by functional analysis.

Mutational frequencies observed in TARDBP (10.3% FALS and 0.9% SALS) and FUS (5% FALS and 0.78% SALS) were consistent with previous findings (Chiò et al., 2012b; Lattante et al., 2013; Polymenidou and Cleveland, 2017; Sproviero et al., 2012). We confirmed that FUS mutations are associated with an earlier onset of the disease in comparison with the general mean age of approximately 60 years reported for ALS (Chiò et al., 2012a). None of these patients showed signs of cognitive impairment. Approximately 5% of patients with ALS also develop FTD (Groen et al., 2010) and, to date, only a small number of ALS/dementia patients with FUS mutations have been described, although cognitive dysfunction has been reported to be absent or rare in FUS-mediated ALS (Blair et al., 2010; Lagier-Tourenne and Cleveland, 2009; Yan et al., 2010). In addition, regarding the new variation c.738_740duplAGG (p.

Gly246Dupl) identified in a sporadic male ALS patient, we were unable to demonstrate its segregation with the disease, and we did not perform functional assays but prediction methods suggested a pathogenic role of this variation. However, this new variant should be interpreted cautiously considering that it is located in a region of the FUS gene (exon 6) where many of the mutations detected represent susceptibility factors or variants with incomplete penetrance in FALS (Deng et al., 2014), in contrast with most of the mutations located in exon 12e15 (C-term) of the gene that were shown to be pathogenic in FALS and SALS cases. Considering that ALS is a complex and multigenic disease, it is plausible that multiple variants cooperate in influencing disease onset, severity, or duration. To this regard, the investigation of ATXN1 as a potential disease modifier in C9orf72 expansion carriers revealed a statistically significant association between ALS patients bearing the expanded polyQ ATNX1 and those with the pathological expansion in C9orf72. A similar result was reported in a recent independent study (Lattante et al., 2018). These data suggest that mutant ATNX1 may predispose carriers of C9orf72 expansions to ALS development, therefore influencing their phenotype.

In summary, this report gives a picture of a 2-decade traditional genetic investigations of ALS patients in the south of Italy, confirming not only C9orf72 as the most frequent genetic alteration in this population, but also supporting the role of ATNX1 intermediate expansions in predisposing to development of ALS in C9orf72- related patients. However, due to the complex genetic architecture of ALS, a more accurate genomic characterization of patient's needs to be ensured for the development of new-targeted strategies in clinical practice and personalized medicine.

Disclosure statement

The authors declare no competing interests.

Acknowledgements

We thank all the patients and their families for agreeing to participate in genetic studies. We would also like to extend our thanks to the technicians and researchers of the Molecular Genetics laboratory of the Institute of Neurological Sciences, CNR, Mangone, for their help in collaborating with us in the last 20 years.

Table 4. Mutations in ALS patients and clinical phenotypes

Patient	Amino acid change	Variant	MAF (ExAC/GnomAD)	ClinVar (last reviewed)[a]	fALS/sALS	Gender (M/F)	Site of onset (S/B)	Age of onset (y)	Disease duration (y)	Reference
SOD1 NM_000454.4										
1	N19S	c.59A>G	0.00008/0.0084	Uncertain significance[A] (June 15, 2018)	sALS	Male	Bulbar	45	1.1	Andersen et al. (2003)
2	N19S	c.59A>G	0.00008/0.00008	Uncertain significance[A] (June 15, 2018)	sALS	Male	Spinal	85	1.2	Andersen et al. (2003)
3	Q22L	c.68A>T	-/-	Likely pathogenic[B] (March 31, 2020)	sALS	Female	Spinal	48	2	Andersen et al. (2003)
4	G61R	c.184G>C	-/-	Not reported	sALS	Female	Spinal	56	N/A	Conforti et al. (2011)
5	D83V	c.251A>T	-/-	Not reported	sALS	Male	Spinal	44	2	This report
6	D90A	c.272A>C	0.00112/0.00143	Likely pathogenic[B] (December 12, 2016)	sALS	Female	Spinal	55	Years	Andersen et al. (1995)
7	D90A	c.272A>C	0.00112/0.00143	Likely pathogenic[B] (December 12, 2016)	fALS	Female	Spinal	46	N/A	Andersen et al. (1995)
8	D90A	c.272A>C	0.00112/0.00143	Likely pathogenic[B] (December 12, 2016)	fALS	Male	Spinal	49	N/A	Andersen et al. (1995)
9	D90 A	c.272 A>C	0.00112/0.00143	Likely pathogenic[B] (December 12, 2016)	fALS	Female	Spinal	55	Years	Andersen et al. (1995)
10	D90A	c.272A>C	0.00112/0.00143	Likely pathogenic[B] (December 12, 2016)	fALS	Male	Spinal	63	N/A	Andersen et al. (1995)
11	D90A	c.272A>C	0.00112/0.00143	Likely pathogenic[B] (December 12, 2016)	sALS	Male	Spinal	33	24	Andersen et al. (1995)
12	D90A/hete	c.272A>C	0.00112/0.00143	Likely pathogenic[B] (December 12, 2016)	sALS	Female	Spinal	54	2	Andersen et al. (1995)
13	D90A/hete	c.272A>C	0.00112/0.00143	Likely pathogenic[B] (December 12, 2016)	sALS	Male	Spinal	52	2.4	Andersen et al. (1995)
14	G93D	c.281G>C	-/-	Not reported	fALS	Female	Spinal	63	1.8	Esteban et al. (1994)
15	G93D	c.281G>C	-/-	Not reported	sALS	Female	Spinal	36	N/A	Esteban et al. (1994)
16	L106P	c.320T>C	-/-	Not reported	sALS	Male	Spinal	77	3	Valentino et al. (2005)
17	R115C	c.346C>T	-/-	Not reported	sALS	Male	Spinal	73	N/A	Tortelli et al. (2013)
18	L144F	c.435G>T	-/0.00001	Pathogenic[B] (March 31, 2020)	sALS	Female	Spinal	N/A	N/A	Weber et al. (2012)
19	I149T	c.449T>C	-/0.00008	Not reported	fALS	Male	Spinal	41	2	Pramatarova et al. (1995)
TARDBP NM_007375.3										
20	G294V	c.881G>T	-/-	Pathogenic[C] (April 23, 2009)	sALS	Female	Bulbar	63	1.3	Corrado et al. (2009)
21	G294V	c.881G>T	-/-	Pathogenic[C] (April 23, 2009)	sALS	Male	Bulbar	68	1.2	Corrado et al. (2009)
22	G294V	c.881G>T	-/-	Pathogenic[C] (April 23, 2009)	sALS	Male	Spinal	48	1.6	Corrado et al. (2009)
23	G294V	c.881G>T	-/-	Pathogenic[C] (April 23, 2009)	fALS	Male	Spinal	61	N/A	Corrado et al. (2009)
24	G294V	c.881G>T	-/-	Pathogenic[C] (April 23, 2009)	fALS	Female	Spinal	58	0.3	Corrado et al. (2009)
25	G294V	c.881G>T	-/-	Pathogenic[C] (April 23, 2009)	fALS	Male	Bulbar	59	3.3	Corrado et al. (2009)
26	G294 V	c.881 G>T	-/-	Pathogenic[C] (April 23, 2009)	fALS	Female	Spinal	63	1	Corrado et al. (2009)
27	G294V/homo	c.881G>T	-/-	Pathogenic[C] (April 23, 2009)	sALS	Female	Spinal	78	1.2	Corrado et al. (2009)
28	G295R	c.883G>A	-/-	Pathogenic[C] (March 12, 2015)	sALS	Female	Spinal	58	19	Corrado et al. (2009)
29	G295R	c.883G>A	-/-	Pathogenic[C] (March 12, 2015)	sALS	Female	Bulbar	51	N/A	Corrado et al. (2009)
30	G376D	c.1127G>A	-/-	Not reported	fALS	Female	Spinal	58	N/A	Conforti et al. (2011)
31	S379A	c.1135T>G	-/0.00003	Not reported	sALS	Female	Spinal	79	0.7	Sprovieri et al. (2019)
32	A382T	c.1144G/A	-/0.00003	Likely pathogenic[B] (March 31, 2020)	sALS	Male	Spinal	52	6	Kabashi et al. (2008)
33	A382T	c.1144G/A	-/0.00003	Likely pathogenic[B] (March 31, 2020)	fALS	Male	Spinal	36	3	Kabashi et al. (2008)
FUS NM_004960.3 34	Gly174_Gly175del	c.518_523del	-/-	Not reported	sALS	Male	Spinal	65	N/A	Kwiatkowski et al. (2009)

35	R216C	c.646C>T	0.00002/0.00012	Pathogenic[C] (August 10, 2012)	sALS	Male	N/A	60	N/A	Kwiatkowski et al. (2009)
36	Gly246DUPL	c.738_740duplAGG	-/-	Not reported	sALS	Male	Spinal	75	2.2	This report
37	R521G	c.1561C>G	-/0.00001	Pathogenic[A] (March 5, 2018)	fALS	Male	Spinal	31	N/A	Kwiatkowski et al. (2009)
38	R521C	c.1561C>T	-/0.000012	Pathogenic[C] (August 31, 2010)	sALS	Female	Spinal	35	N/A	Kwiatkowski et al. (2009)
39	R521C	c.1561C>T	-/0.000012	Pathogenic[C] (August 31, 2010)	sALS	Female	Spinal	53	5	Kwiatkowski et al. (2009)
40	R521C	c.1561C>T	-/0.000012	Pathogenic[C] (August 31, 2010)	fALS	Female	Spinal	26	5	Kwiatkowski et al. (2009)
41	P525L	c.1574C>T	-/0.000004	Pathogenic[B] (March 31, 2020)	sALS	Female	Spinal	45	3	Kwiatkowski et al. (2009)
42	P525L	c.1574C>T	-/0.000004	Pathogenic[B] (March 31, 2020)	sALS	Male	Spinal	26	1	Kwiatkowski et al. (2009)
43	P525L	c.1574 C>T	-/0.000004	Pathogenic[B] (March 31, 2020)	fALS	Female	Bulbar	21	1.6	Kwiatkowski et al. (2009)
ANG NM_001145.4										
44	M-1I	c.3G>A	0.000199/0.000216	Uncertain significance[B] (October 1, 2018)	sALS	Male	Spinal	63	N/A	Conforti et al. (2008)

Key: ExAC, Exome Aggregation Consortium; fALS, familial amyotrophic lateral sclerosis; FTD, frontotemporal dementia; GnomAD, genomes aggregation database; MAF, minor allele frequency; N/A, data not available; sALS, sporadic amyotrophic lateral sclerosis.

[a] ClinVar: Variant interpretation and assertion criteria according to the following: A, Nykamp K et al. (Genet Med, 2017); B, ACMG Guidelines by Richards et al. (2015); C, no assertion criteria provided.

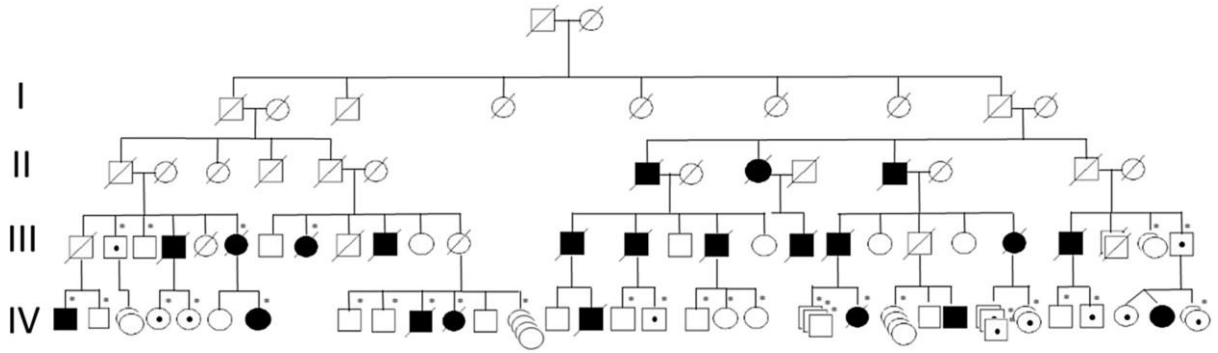


Fig. 3. Pedigree of the TARDBP-G376D family with a history of ALS showing an autosomal- dominant pattern of inheritance. Square indicates male; circle female; slash deceased; black symbols patients affected by ALS. empty symbols with black dot are asymptomatic carriers. * Indicates DNA available for the study. Abbreviations: ALS, amyotrophic lateral sclerosis.

Table 5. SNPs detected in ALS patients and controls.

Overall patients, n	dbSNP	cDNA alteration	Amino acid change	Func. refGene	ALS patients (%)	Control patients (%)	Allele frequency ExAC	Allele frequency GnomAD	ClinVar[a]
SOD1 NM_000454.4									
997	rs1804447	c.*2C>T	-	3-UTR	0.1	-	-	0.25	Benign
	-	-	-	5-UTR	0.1	-	-	-	-
	rs112510394	c.423T>A	A141A	Exonic	0.1	-	0.00046	0.00022	Benign
	rs143100660	c.180T>C	S60S	Exonic	0.1	-	0.00001	0.00006	Benign
	rs373888553	c.239þ34A>C	-	Intronic	5.3	-	0.03857	0.04007	Benign
	rs2234694	c.357þ42del [TACA]	-	Intronic	0.1	-	-	-	-
	-								
TARDBP NM_007375.3									
928	rs61730366	c.198T>A	A66A	Exonic	0.1	-	0.00640	0.00178	Benign

FUS NM_004960.3 953	rs80301724	c.*41G>A	-	3-UTR	1.0	-	0.00691	0.00405	Benign
	rs757651881	c.G230_G231del	-	Inframe deletion	0.1	-	-	-	Uncertain significance
	rs13331793	c.1393þ34G>T	-	Intronic	0.1	-	-	1.2	-
	-	c.669_671del GGcGGc	Gly226_ Gly227del	Inframe deletion	-	-	-	-	Benign
	-	c.335-15del [TTTT]	E	Intronic	0.1	-	-	-	-
	rs138901914	c.1566G>A	R522R	Exonic	0.2	-	0.00123	0.00081	Benign
ANG NM_001145.4 390	rs11701	c.330T>A	G110G	Exonic	28	22	-	0.12600	Benign
	rs2228653	c.363A>G	T121T	Exonic	0.25	0.3	0.00488	0.01278	Benign
	rs121909541	c.208A>G	I70V	Exonic	0.5	0	0.000609	0.00080	Uncertain significance
	rs17560	c.250A>G	K84E	Exonic	0.25	0.3	0.0015	0.0038	Benign
VCP NM:007126.5 300	rs10972300	c.129þ47G>A	-	Intronic	18.3	13	0.1458	0.1634	Benign
	rs757728490	c.576þ10C>G	-	Intronic	0.3	0	-	-	Likely benign
	-	c.1194þ38T>C	-	Intronic	4.6	0	-	-	-
	-	c.1082-21INS [TTGTGTACTGT]	-	Intronic	16.6	5	-	-	-
	rs142577424	c.1704A>G	Q568Q	Exonic	1.3	0	0.0026	0.0024	Benign
VAPB NM_004738.4 154	rs563516701	c.1722A>G	L574L	Exonic	0.3	0	$8\ 10^{6}$	-	-
	rs2234487	c.315þ35C>T	-	Intronic	54	43	0.41674	0.45475	Benign
	rs2234488	c.315þ138A>G	-	Intronic	43	27	-	0.36118	–
	rs374376908	c.547C>T	L183L	Exonic	0.6	0	$8\ 10^{6}$	-	-
	rs146459055	c.390T>G	D130E	Exonic	2	2.4	0.00135	0.00051	Benign

Key: e, not present or zero; cDNA, complementary deoxyribonucleic acid; dbSNP, database of single nucleotide polymorphism; ExAC, Exome Aggregation Consortium; GnomAD, genomes aggregation database; n, number; SNPs, single nucleotide polymorphisms.

a According to ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/).

References

Andersen, P.M., Nilsson, P., Ala-Hurula, V., Keränen, M.L., Tarvainen, I., Haltia, T., Nilsson, L., Binzer, M., Forsgren, L., Marklund, S.L., 1995. Amyotrophic lateral sclerosis associated with homozygosity for an Asp90Ala mutation in CuZnsuperoxide dismutase. Nat. Genet. 10, 61e66.

Andersen, P.M., Sims, K.B., Xin, W.W., Kiely, R., O'Neill, G., Ravits, J., Pioro, E., Harati, Y., Brower, R.D., Levine, J.S., Heinicke, H.U., Seltzer, W., Boss, M., Brown Jr., R.H., 2003. Sixteen novel mutations in the Cu/Zn superoxide dismutase gene in amyotrophic lateral sclerosis: a decade of discoveries, defects and disputes. Amyotroph. Lateral Scler. Other Mot. Neuron Disord 4, 62e73.

Battistini, S., Benigni, M., Ricci, C., Rossi, A., 2012. SOD1 mutations in amyotrophic lateral sclerosis. Eur. Neurol. J. 4, 33e43.

Blair, I.P., Williams, K.L., Warraich, S.T., Durnall, J.C., Thoeng, A.D., Manavis, J., Blumbergs, P.C., Vucic, S., Kiernan, M.C., Nicholson, G.A., 2010. FUS mutations in amyotrophic lateral sclerosis: clinical, pathological, neurophysiological and genetic analysis. J. Neurol. Neurosurg. Psychiatry 81, 639e645.

Brooks, B.R., Miller, R.G., Swash, M., Munsat, T.L., 2000. World Federation of Neurology Research Group on motor neuron diseases. El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. Amyotroph. Lateral Scler. Other Mot. Neuron Disord 1, 293e299.

Byrne, S., Elamin, M., Bede, P., Hardiman, O., 2012. Absence of consensus in diagnostic criteria for familial neurodegenerative diseases. J. Neurol. Neurosurg. Psychiatry 83, 365e367.

Chiò, A., Calvo, A., Mazzini, L., Cantello, R., Mora, G., Moglia, C., Corrado, L., D'Alfonso, S., Majounie, E., Renton, A., Pisano, F., Ossola, I., Brunetti, M., Traynor, B.J., Restagno, G., PARALS, 2012a. Extensive genetics of ALS: a population-based study in Italy. Neurology 79, 1983e1989.

Chiò, A., Logroscino, G., Traynor, B.J., Collins, J., Simeone, J.C., Goldstein, L.A., White, L.A., 2013. Global epidemiology of amyotrophic lateral sclerosis: a systematic review of the published literature. Neuroepidemiology 41, 118e130.

Chiò, A., Restagno, G., Brunetti, M., Ossola, I., Calvo, A., Canosa, A., Moglia, C., Floris, G., Tacconi, P., Marrosu, F., Marrosu, M.G., Murru, M.R., Majounie, E., Renton, A.E., Abramzon, Y., Pugliatti, M., Sotgiu, M.A., Traynor, B.J., Borghero, G., SARDINIALS Consortium, 2012b. ALS/FTD phenotype in two Sardinian families carrying both C9ORF72 and TARDBP mutations. J. Neurol. Neurosurg. Psychiatry 83, 730e733.

Conforti, F.L., Spataro, R., Sproviero, W., Mazzei, R., Cavalcanti, F., Condino, F., Simone, I.L., Logroscino, G., Patitucci, A., Magariello, A., Muglia, M., Rodolico, C., Valentino, P., Bono, F., Colletti, T., Monsurrò, M.R., Gambardella, A., La Bella, V., 2012. Ataxin-1 and ataxin-2 intermediate-length PolyQ expansions in amyotrophic lateral sclerosis. Neurology 79, 2315e2320.

Conforti, F.L., Sprovieri, T., Mazzei, R., Ungaro, C., La Bella, V., Tessitore, A., Patitucci, A., Magariello, A., Gabriele, A.L., Tedeschi, G., Simone, I.L., Majorana, G., Valentino, P., Condino, F., Bono, F., Monsurrò, M.R., Muglia, M., Quattrone, A., 2008. A novel angiogenin gene mutation in a sporadic patient with amyotrophic lateral sclerosis from southern Italy. Neuromuscul. Disord. 18, 68e70.

Conforti, F.L., Sproviero, W., Simone, I.L., Mazzei, R., Valentino, P., Ungaro, C., Magariello, A., Patitucci, A., La Bella, V., Sprovieri, T., Tedeschi, G., Citrigno, L., Gabriele, A.L., Bono, F., Monsurrò, M.R., Muglia, M., Gambardella, A., Quattrone, A., 2011. TARDBP gene mutations in south Italian patients with amyotrophic lateral sclerosis. J. Neurol. Neurosurg. Psychiatry 82, 587e588.

Conte, A., Lattante, S., Zollino, M., Marangi, G., Luigetti, M., Del Grande, A., Servidei, S., Trombetta, F., Sabatelli, M., 2012. P525L FUS mutation is consistently associated with a severe form of juvenile amyotrophic lateral sclerosis. Neuromuscul. Disord. 22, 73e75.

Corrado, L., Ratti, A., Gellera, C., Buratti, E., Castellotti, B., Carlomagno, Y., Ticozzi, N., Mazzini, L., Testa, L., Taroni, F., Baralle, F.E., Silani, V., D'Alfonso, S., 2009. High frequency of TARDBP gene mutations in Italian patients with amyotrophic lateral sclerosis. Hum. Mutat. 30, 688e694.

Cruts, M., Engelborghs, S., van der Zee, J., Van Broeckhoven, C., 2015. C9orf72-related

amyotrophic lateral sclerosis and frontotemporal dementia. In: Adam, M.P., Ardinger, H.H., Pagon, R.A., Wallace, S.E., Bean, L.J.H., Stephens, K., Amemiya, A. (Eds.), SourceGeneReviews_ [Internet]. University of Washington, Seattle, Seattle (WA), pp. 1993e2020.

Deng, H., Gao, K., Jankovic, J., 2014. The role of FUS gene variants in neurodegenerative diseases. Nat. Rev. Neurol. 10, 337e348.

Elden, A.C., Kim, H.J., Hart, M.P., Chen-Plotkin, A.S., Johnson, B.S., Fang, X., Armakola, M., Geser, F., Greene, R., Lu, M.M., Padmanabhan, A., Clay-Falcone, D., McCluskey, L., Elman, L., Juhr, D., Gruber, P.J., Rüb, U., Auburger, G., Trojanowski, J.Q., Lee, V.M., Gitler, A.D., 2010. Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. Nature 466, 1069e1075.

Esteban, J., Rosen, D.R., Bowling, A.C., Sapp, P., McKenna-Yasek, D., O'Regan, J.P., Beal, M.F., Horvitz, H.R., Brown Jr., R.H., 1994. Identification of two novel mutations and a new polymorphism in the gene for Cu/Zn superoxide dismutase in patients with amyotrophic lateral sclerosis. Hum. Mol. Genet. 3, 997e998.

Gibson, S.B., Downie, J.M., Tsetsou, S., Feusier, J.E., Figueroa, K.P., Bromberg, M.B., Jorde, L.B., Pulst, S.M., 2017. The evolving genetic risk for sporadic ALS. Neurology 89, 226e233.

Groen, E.J., van Es, M.A., van Vught, P.W., Spliet, W.G., van Engelen-Lee, J., de Visser, M., Wokke, J.H., Schelhaas, H.J., Ophoff, R.A., Fumoto, K., Pasterkamp, R.J., Dooijes, D., Cuppen, E., Veldink, J.H., van den Berg, L.H., 2010. FUS mutations in familial amyotrophic lateral sclerosis in The Netherlands. Arch. Neurol. 67, 224e230.

Kabashi, E., Valdmanis, P.N., Dion, P., Spiegelman, D., McConkey, B.J., Vande Velde, C., Bouchard, J.P., Lacomblez, L., Pochigaeva, K., Salachas, F., Pradat, P.F., Camu, W., Meininger, V., Dupre, N., Rouleau, G.A., 2008. TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. Nat. Genet. 40, 572e574.

Kwiatkowski Jr., T.J., Bosco, D.A., Leclerc, A.L., Tamrazian, E., Vanderburg, C.R., Russ, C., Davis, A., Gilchrist, J., Kasarskis, E.J., Munsat, T., Valdmanis, P., Rouleau, G.A., Hosler, B.A., Cortelli, P., de Jong, P.J., Yoshinaga, Y., Haines, J.L., Pericak-Vance, M.A., Yan, J., Ticozzi, N., Siddique, T., McKenna-Yasek, D., Sapp, P.C., Horvitz, H.R., Landers, J.E., Brown Jr., R.H., 2009. Mutations in the FUS/ TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. Science 323, 1205e1208.

Lagier-Tourenne, C., Cleveland, D.W., 2009. Rethinking ALS: the FUS about TDP-43. Rev. Cell 36, 1001e1004.

Lamp, M., Origone, P., Geroldi, A., Verdiani, S., Gotta, F., Caponnetto, C., Devigili, G., Verriello, L., Scialò, C., Cabona Canosa, A., Vanni, I., Bellone, E., Eleopra, R., Mandich, P., 2018. Twenty years of molecular analyses in amyotrophic lateral sclerosis: genetic landscape of Italian patients. Neurobiol. Aging 66, 179.e5-179.e16.

Lattante, S., Conte, A., Zollino, M., Luigetti, M., Del Grande, A., Marangi, G., Romano, A., Marcaccio, A., Meleo, E., Bisogni, G., Rossini, P.M., Sabatelli, M., 2012. Contribution of major amyotrophic lateral sclerosis genes to the etiology of sporadic disease. Neurology 79, 66e72.

Lattante, S., Pomponi, M.G., Conte, A., Marangi, G., Bisogni, G., Patanella, A.K., Meleo, E., Lunetta, C., Riva, N., Mosca, L., Carrera, P., Bee, M., Zollino, M., Sabatelli, M., 2018. ATXN1 intermediate-length polyglutamine expansions are associated with amyotrophic lateral sclerosis. Neurobiol. Aging 64, 157.e1-157.e5.

Lattante, S., Rouleau, G.A., Kabashi, E., 2013. TARDBP and FUS mutations associated with amyotrophic lateral sclerosis: summary and update. Hum. Mutat. 34, 812e826.

Majounie, E., Renton, A.E., Mok, K., Dopper, E.G., Waite, A., Rollinson, S., Chiò, A., Restagno, G., Nicolaou, N., Simon-Sanchez, J., van Swieten, J.C., Abramzon, Y., Johnson, J.O., Sendtner, M., Pamphlett, R., Orrell, R.W., Mead, S., Sidle, K.C., Houlden, H., Rohrer, J.D., Morrison, K.E., Pall, H., Talbot, K., Ansorge, O. Chromosome 9-ALS/FTD Consortium; French research network on FTLD/FTLD/ALS; ITALSGEN Consortium, Hernandez, D.G., Arepalli, S., Sabatelli, M., Mora, G., Corbo, M., Giannini, F., Calvo, A., Englund, E., Borghero, G., Floris, G.L., Remes, A.M., Laaksovirta, H., McCluskey, L., Trojanowski, J.Q., Van Deerlin, V.M., Schellenberg, G.D., Nalls, M.A., Drory, V.E., Lu, C.S., Yeh, T.H., Ishiura, H., Takahashi, Y., Tsuji, S., Le Ber, I., Brice, A., Drepper, C., Williams, N., Kirby, J., Shaw, P., Hardy, J., Tienari, P.J., Heutink, P., Morris, H.R.,

Pickering-Brown, S., Traynor, B.J., 2012. Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. Lancet Neurol. 11, 323e330.

McCauley, M.E., Baloh, R.H., 2019. Inflammation in ALS/FTD pathogenesis. Acta Neuropathol. 137, 715e730.

Millecamps, S., Boillée, S., Le Ber, I., Seilhean, D., Teyssou, E., Giraudeau, M., Moigneu, C., Vandenberghe, N., Danel-Brunaud, V., Corcia, P., Pradat, P.F., Le Forestier, N., Lacomblez, L., Bruneteau, G., Camu, W., Brice, A., Cazeneuve, C., Leguern, E., Meininger, V., Salachas, F., 2012. Phenotype difference between ALS patients with expanded repeats in C9ORF72 and patients with mutations in other ALS-related genes. J. Med. Genet. 49, 258e263.

Müller, K., Brenner, D., Weydt, P., Meyer, T., Grehl, T., Petri, S., Grosskreutz, J., Schuster, J., Volk, A.E., Borck, G., Kubisch, C., Klopstock, T., Zeller, D., Jablonka, S., Sendtner, M., Klebe, S., Knehr, A., Günther, K., Weis, J., Claeys, K.G., Schrank, B.,
Sperfeld, A.D., Hübers, A., Otto, M., Dorst, J., Meitinger, T., Strom, T.M., Andersen, P.M., Ludolph, A.C., Weishaupt, J.H., German ALS network MND-NET, 2018. Comprehensive analysis of the mutation spectrum in 301 German ALS families. J. Neurol. Neurosurg. Psychiatry 89, 817e827.

Polymenidou, M., Cleveland, D.W., 2017. Biological spectrum of amyotrophic lateral sclerosis prions. Cold Spring Harb Perspect. Med. 7, a024133.

Pramatarova, A., Figlewicz, D.A., Krizus, A., Han, F.Y., Ceballos-Picot, I., Nicole, A., Dib, M., Meininger, V., Brown, R.H., Rouleau, G.A., 1995. Identification of new mutations in the Cu/Zn superoxide dismutase gene of patients with familial amyotrophic lateral sclerosis. Am. J. Hum. Genet. 56, 592e596.

Renton, A.E., Chiò, A., Traynor, B.J., 2014. State of play in amyotrophic lateral sclerosis genetics. Nat. Neurosci. 17, 17e23.

Rowland, L.P., Shneider, N.A., 2001. Amyotrophic lateral sclerosis. N. Engl. J. Med. 344, 1688e1700.

Sprovieri, T., Ungaro, C., Perrone, B., Naimo, G.D., Spataro, R., Cavallaro, S., La Bella, V., Conforti, F.L., 2019. A novel S379A TARDBP mutation associated to lateonset sporadic ALS. Neurol. Sci. 40, 2111e2118.

Sproviero, W., La Bella, V., Mazzei, R., Valentino, P., Rodolico, C., Simone, I.L., Logroscino, G., Ungaro, C., Magariello, A., Patitucci, A., Tedeschi, G., Spataro, R., Condino, F., Bono, F., Citrigno, L., Monsurrò, M.R., Muglia, M., Gambardella, A., Quattrone, A., Conforti, F.L., 2012. FUS mutations in sporadic amyotrophic lateral sclerosis: clinical and genetic analysis. Neurobiol. Aging 33, 837.e1-5.

Tortelli, R., Conforti, F.L., Cortese, R., D'Errico, E., Distaso, E., Mazzei, R., Ungaro, C., Magariello, A., Gambardella, A., Logroscino, G., Simone, I.L., 2013. Amyotrophic lateral sclerosis: a new missense mutation in the SOD1 gene. Neurobiol. Aging 34, 1709.e3-5.

Turner, M.R., Al-Chalabi, A., Chiò, A., Hardiman, O., Kiernan, M.C., Rohrer, J.D., Rowe, J., Seeley, W., Talbot, K., 2017. Genetic screening in sporadic ALS and FTD. J. Neurol. Neurosurg. Psychiatry 88, 1042e1044.

Umoh, M.E., Fournier, C., Li, Y., Polak, M., Shaw, L., Landers, J.E., Hu, W., Gearing, M., Glass, J.D., 2016. Comparative analysis of C9orf72 and sporadic disease in an ALS clinic population. Neurology 87, 1024e1030.

Valentino, P., Conforti, F.L., Pirritano, D., Nisticò, R., Mazzei, R., Patitucci, A., Sprovieri, T., Gabriele, A.L., Muglia, M., Clodomiro, A., Gambardella, A., Zappia, M., Quattrone, A., 2005. Brachial amyotrophic diplegia associated with a novel SOD1 mutation (L106P). Neurology 64, 1477e1478.

van Blitterswijk, M., DeJesus-Hernandez, M., Niemantsverdriet, E., Murray, M.E., Heckman, M.G., Diehl, N.N., Brown, P.H., Baker, M.C., Finch, N.A., Bauer, P.O., Serrano, G., Beach, T.G., Josephs, K.A., Knopman, D.S., Petersen, R.C., Boeve, B.F., Graff-Radford, N.R., Boylan, K.B., Petrucelli, L., Dickson, D.W., Rademakers, R., 2013. Association between repeat sizes and clinical and pathological characteristics in carriers of C9ORF72 repeat expansions (Xpansize-72): a cross-sectional cohort study. Lancet Neurol. 12, 978e988.

Weber, M., Neuwirth, C., Thierbach, J., Schweikert, K., Czaplinski, A., Petersen, J., Jung, H.H., Birve, A., Marklund, S.L., Andersen, P.M., 2012. ALS patients with SOD1 mutations in Switzerland show very diverse phenotypes and extremely long survival. J. Neurol. Neurosurg. Psychiatry 83, 351e353.

Yan, J., Deng, H.X., Siddique, N., Fecto, F., Chen, W., Yang, Y., Liu, E., Donkervoort, S., Zheng, J.G., Shi, Y., Ahmeti, K.B., Brooks, B.,

Engel, W.K., Siddique, T., 2010. Frameshift and novel mutations in FUS in familial amyotrophic lateral sclerosis and ALS/dementia. Neurology 75, 807e814.

Zou, Z.Y., Zhou, Z.R., Che, C.H., Liu, C.Y., He, R.L., Huang, H.P., 2017. Genetic epidemiology of amyotrophic lateral sclerosis: a systematic review and meta-analysis. J. Neurol. Neurosurg. Psychiatry 88, 540e549.

Chapter 3

Published in Genes 2021, 12, 1843. https://doi.org/10.3390/ genes12121843

Individual Oligogenic Background in p.D91A-SOD1 Amyotrophic Lateral Sclerosis Patients

ABSTRACT

The p.D91A is one of the most common ALS-causing SOD1 mutations and is known to be either recessive or dominant. The homozygous phenotype is characterized by prolonged survival and slow progression of disease, whereas the affected heterozygous phenotypes can vary. To date, no genetic protective factors located close to SOD1 have been associated with the mild progressive homozygous phenotype. Using Next Generation Sequencing (NGS), we characterized a small cohort of sporadic and familial p.D91A-SOD1 heterozygous (n = 2) or homozygous (n = 5) ALS patients, to reveal any additional contributing variant in 39 ALS-related genes. We detected unique sets of non-synonymous variants, four of which were of uncertain significance and several in untranslated regions of ALS-related genes. Our results supported an individual oligogenic background underlying
both sporadic and familial p.D91A cases irrespective of their p.D91A mutant alleles. We suggest that a comprehensive genomic view of p.D91A-SOD1 ALS patients may be useful in identifying emerging variants and improving disease diagnosis as well as guiding precision medicine.

Keywords: p.D91A-SOD1; zygosity; NGS targeted-gene panel; individual oligogenic background

1. Introduction

Amyotrophic lateral sclerosis (ALS) represents the third most common neurodegenerative disease, characterized by the progressive adult-onset degeneration of upper and lower motor neurons [1]. There are two main forms of ALS, familial (FALS) and apparently sporadic (SALS) accounting for about 10% and 90% of cases, respectively [2]. This complex disease is caused by the interplay of causative genetic factors (monogenic or oligogenic) and risk factors (genetic and non-genetic) [3]. The first ALS-related gene described was superoxide dismutase 1 (SOD1), whose mutations affect about 12% of FALS and 1% of SALS [1]. The most common mutation affecting SOD1 and causing ALS is the substitution of alanine for aspartate at position 91 of exon 4, called p.D91A (also known as p.D90A; dbSNP155 ID rs80265967; NM_000454.5 c.272A>C) [4]. Prevalence of p.D91A in ALS cases varies globally and is distinctly absent in some populations (https://gnomad.broadinstitute.org/, accessed on 22 October 2021). Despite the extensive evidence demonstrating the pathogenicity of p.D91A in SOD1 and according to the variant interpretation and assertion criteria of ACMG guidelines [6], this variant is still reported in ClinVar with conflicting interpretations of pathogenicity. However, the p.D91A

variant is reported as a risk allele, resulting in disease when biallelic or in combination with another risk factor, by the more recent ClinGen curation database [7].

Although all the mutations affecting SOD1 are dominant, p.D91A can also be recessive [2] and the disease status may arise from heterozygous or homozygous mutant alleles, respectively. Indeed, in Scandinavia the p.D91A allele is a variant associated with recessive inheritance [8,9], while in many other countries dominant inheritance was also associated to the disease [10–12]. In particular, in Nordic countries p.D91A-SOD1 is reported with a polymorphic frequency of 2.5% [13], rendering this polymorphism a risk factor in those countries. All p.D91A-homozygous ALS patients show a phenotype characterized by a slower course of disease and not always associated with respiratory failure or cognitive issues [1,9,14,15], while the clinically affected p.D91A-heterozygotes present variable clinical signs and disease progression. There are no affected p.D91A-heterozygotes among homozygous pedigrees [4] and there is limited literature showing p.D91A-SOD1 affected heterozygous patients in multiple members of an ALS family [16,17]. Moreover, p.D91A carriers belong not only to FALS but also to SALS cases, and SOD1 haplotypes show a common ancestor with a shared Scandinavian haplotype of rare alleles in both homozygous and heterozygous patterns [13,18]. Thus, two phenotype-explaining hypotheses have been proposed, the presence of a protective genetic modifier on the Scandinavian haplotype or a co-segregating contributing variant together with the p.D91A haplotype outside Scandinavia [19]. An updated literature search for previously reported p.D91A SOD1-related phenotypes [20] revealed the presence of at least three groups of patients with differences in disease progression rate and survival time, without fully identifying potential genetic modifiers or contributing variants in addition to the p.D91A zygosity (Table 1).

Table 1. p.D91A *SOD1*-related phenotypes.

p.D91A SOD1 Zigosity	***Survival time***	***Progression rate***	***Phenotype***	***Contributing Risk Factors Proposed***	***References***
Hom	More than two years	Slow	Spinal ALS*	Contributing variants mitigating the phenotype not yet identified	[1,9,14-16]
Het	More than two years	Slow	Spinal ALS	Heterozygous compound in SOD1	[16,17,21-24]
Het	About two years	Fast	Variable forms of ALS**	Contributing variants on TDP-43 inclusions not yet identified	[4,5,25]

Hom = homozygous; Het = heterozygous; * with or without respiratory failure and/or cognitive issues; ** including bulbar onset.

The first group of patients displays a slowly evolving phenotype linked to the p.D91Ahomozygous genotype [1,9,14–16] with no other identified genetic modifier responsible for the mild phenotype. In addition, a p.D91A-homozygous patient has been recently described having vocal cord impairment, which is not a typical clinical sign associated with this genotype [15]. The second group of patients, concerning the heterozygous pattern of zygosity, includes cases with compound heterozygous mutations (p.D91A/p.D96N; p.D91A/p.D90V), showing a lower limb site of onset and a slow progressive phenotype with a variable disease duration, ranging from 7 to 28 years [21,22,24]. In addition, five clinically affected p.D91A-heterozygous cases with slow progression of disease were reported: one patient with a negative family history for the disease [23], another belonging to a large family carrying the TDP-43 p.G298S mutation [17], and three individuals of the same family [16]. However, it is not clear if other SOD1 variants were investigated in these cases. The third p.D91A SOD1-related phenotype, shown in Table 1, is characterized by an aggressive evolution of the disease and is found in individuals described carrying a dominant p.D91A variant co-segregating with ALS [4]. In this group falls the case of a p.D91A-heterozygous affected carrier showing TDP-43 aggregates, with a family history of ALS and other neurodegenerative diseases [25]. Few other case reports described coexistence of TDP-43 inclusions with dominantly inherited SOD1 variants since these aggregates are neuropathological hallmarks of ALS-FTD and SALS patients [25–28]. Furthermore, the p.D91A-heterozygous mutation plus

the pathogenic C9ORF72 repeat expansion or the variant of uncertain significance (VUS) UBQLN2-Q460R [29] were already described in patients associated with the ALS-FTD phenotype [19,30]. Differences in the genotype–phenotype correlations delineated above may have considerable therapeutic implications. Indeed, recruitment for antisense therapy was recently discouraged in p.D91A-heterozygous affected carriers after finding evidence of one case showing TDP-43 aggregates as autoptic findings [25]. Although no further variants have been identified in the conserved region surrounding SOD1 that may explain the mild progressive phenotype in homozygous mutation carriers, the existence of contributing genetic factors in other DNA regions cannot be ruled out [31]. Based on these premises, in this study we investigated by targeted NGS the presence of additional variants in 39 ALS-related genes in SALS and FALS patients carrying the p.D91A-SOD1 heterozygous or homozygous mutation, with the aim to reveal any other contributing variant able to explain the homozygous mild phenotype.

2. Materials and Methods

2.1. *Patients*

Informed consent was obtained from all the participants included in the study. ALS patients (three familial and four sporadic cases, four women and three men) from southern Italy unrelated families, diagnosed with ALS according to the El Escorial criteria [32], were previously recruited at our institution and genetically defined as carriers of the p.D91A-SOD1 heterozygous or homozygous mutation by Sanger sequencing analysis, as previously reported [14,33]. Clinical features and known genetic background of patients described in this study are shown in Table 2.

Table 2. Clinical and genetic data of p.D91A-SOD1 patients.

fALS or sALS	Sample ID	Mutant Allele	Gender M/F	Site of Onset	Age of onset (yrs)	Disease Duration (yrs)
fALS	P1	Hom	F	LL	68	n/a
fALS	P2	Hom	M	LL	49	8
fALS	P3	Hom	F	LL	46	2.2
sALS	P4	Hom	F	LL	55	8.4
sALS	P5	Hom	M	UL	33	22
sALS	P6	Het	M	LL	52	2.5
ALS	P7	Het	F	LL	54	2

[a] Patient died; [b] alive in April 2021; [c] no more follow-up since 2013. All the data have been inferred from patient medical records and from previously published data [14,33]. FALS = familial amyotrophic lateral sclerosis; SALS = sporadic amyotrophic lateral sclerosis; Hom = homozygous allele variant; Het = heterozygous allele variant; SNV = single nucleotide variation; Gender = male/female; S = UL for upper limbs and LL for lower limbs; n/a = data not available. The allele frequency percentages for both variants are: 0.001432 (gnomAD v2 1.1.exomes), 0.00207 (gnomAD v2 1.1.genomes), and 0.0004 (Genome Project databases).

In our cohort, patients affected by the biallelic p.D91A variant, mainly showed a prolonged survival unlike p.D91A heterozygous affected carriers. ALS patients without SOD1 mutations, together with individuals affected either by a motor neuronal (MN) phenotype or other neurodegenerative diseases not associated to ALS and belonging to our Southern Italian reference population were used as control samples (C1-C8) since we were not interested in pleiotropic effect [34]. Our filtering strategy aimed to identify rare and polymorphic variants able to synergistically act with the already known causative mutation, so data were normalized for two different types of confounding factors, such as genetic background and overlapping phenotypes at the same time.

The study was approved by the local Ethics Committee of Azienda Ospedaliero Universitaria of Bari N 1025.

2.2. *NGS Analysis*

Genomic DNA quality and quantity were evaluated using standard agarose electrophoresis and Qubit™ Fluorometer (Invitrogen, Waltham, MA, USA). Based on quantitative results, samples were normalized to 50 ng/µL to be used as an input in targeted NGS sequencing analysis. Deep sequencing analysis was performed using a custom ALS-related gene panel including genes known to be associated or possibly associated with ALS and

overlapped phenotypes. In detail, we targeted the coding regions of 39 ALS-related genes with at least 25 bp of intronic flanking regions, together with the promoter region of the following subset of genes: *SOD1, TARDBP, FUS, ANG, ALS2, TBK1, SPG11, PFN1, TUBA4A, SETX, VCP, MATR3, VAPB, CCNF, NEK1, HNRPA1, and ERBB4* (Supplementary Table S1) [35]. Libraries were prepared using the custom Ion AmpliSeq kit (Life Technologies, Carlsbad, CA, USA), and sequencing analysis was run on an Ion Torrent Personal Genome Machine™ (PGM™) sequencer (Thermo Fisher Scientific, Carlsbad, CA, USA). To set the bioinformatic pipeline, we followed the best practice consensus recommendations developed by the College of American Pathologists and the American Medical Informatics Association [36]. Primary and secondary data analysis were performed using the Torrent Suite (Thermo Fisher Scientific, Carlsbad, USA), with the Human genome [19] to align sequences and a Germline low stringency variant caller setting. Tertiary level data analysis was carried out using Partek Flow software build version 10.0.21.0302 (Partek Inc., St. Louis, MO, USA. Variant annotation was performed using Ensemble transcript release 75, SnpEff, VEP.84 databases. Variant frequencies in ALS patients and controls belonging to the project MinE database, were annotated using the project MinE data browser (http://databrowser.projectmine.com/, accessed on 20 April 2021), whose current dataset contains WGS data from 4366 ALS cases and 1832 controls [37]. Variants were then pre-filtered for SNP low stringency quality parameters (FAO > 2, FDP > 6, QUAL > 20, STB < 0.9) to filter out false positives and retain >99% of true positives calls (a strategy optimized for amplicon-based semiconductor sequencing) [38]. Variants with a QD < 1 were also filtered out based on our experience on SNV false positive calls and the observation that true positive calls have high mean coverage and quality by depth values [39]. Variant prioritization was carry out filtering for Read Depth_20X (minimum read depth for germline variants calling), and for MAF—European Ancestry Population—Freq< 0.5. We further removed intronic variants and synonymous variants not affecting canonical splice sites. The same NGS and bioinformatic pipelines were applied either on patients (P1-P7) or controls (C1-C8) to compare and filter patients' variants. Comparison between patient and control variants were performed using the Summarize cohort mutations by merging pairs option in Partek Flow. Variant classification was performed querying VarSome (https://varsome.com/, accessed on 20 April 2021), a search engine for human genomic variation freely available and implemented to automatically classify and report the variant classification according to ACMG guidelines [40], and ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/, accessed on 20 April 2021) [41]. We also used CADD GRCh37-v1.6 (https://cadd.gs.washington.edu/snv, accessed

on 24 October 2021) [42] prediction ranking for deleteriousness of variants, without setting an arbitrary cut-off for our disease model, since p.D91A variant is reported having a CADD score of 9.481, which is below the scaled scores of 10 (predicted to correspond to the 10% most deleterious substitutions in the human genome). Inferential statistics was not conducted because of the small sample size. Descriptive statistics for variant frequencies was calculated in our case series, in our South Italian reference population [33], as well as inferred by the related population databases.

2.3. eQTLs

Expression Quantitative Trait Loci (eQTL) analysis, helpful to understand the effect of genetic variations on the transcriptome in healthy post-mortem tissues donors, was performed using Genotype-Tissue Expression GTEx Portal v.8 (www.gtexportal.org, accessed on 24 October 2021) [43]. The eQTL for each of the 19 variants of interest was calculated for five tissues of interest (whole blood, brain cortex, brain frontal cortex, spinal cord, and skeletal muscle) using GTEx eQTL Calculator, generating a p-value for each variant-gene pair T-statistics in an eQTL. T-test results were corrected using Bonferroni correction test.

3. Results

Data obtained from deep sequencing analysis of p.D91A-SOD1 patients are reported in detail in supplementary data. In particular, coverage analysis and alignment quality parameters are shown in Supplementary Tables S1 and S2, respectively. Results obtained after pre-filtering for variant quality and prioritization for germline read depth and frequency in the European population showed only already known variants (Supplementary Table S3). We obtained a p.D91A average read depth across the seven ALS patients of 616X, and a mean quality by depth of 8.4 and 38.3 for heterozygous and homozygous, respectively. The variant prioritization strategy adopted (Figure 1) showed the presence of 19 SNVs in p.D91A-SOD1 patients (Table 3). Using the VarSome tool, four variants were classified as a VUS (rs45488900, rs41266793, rs139334167, rs76708676).

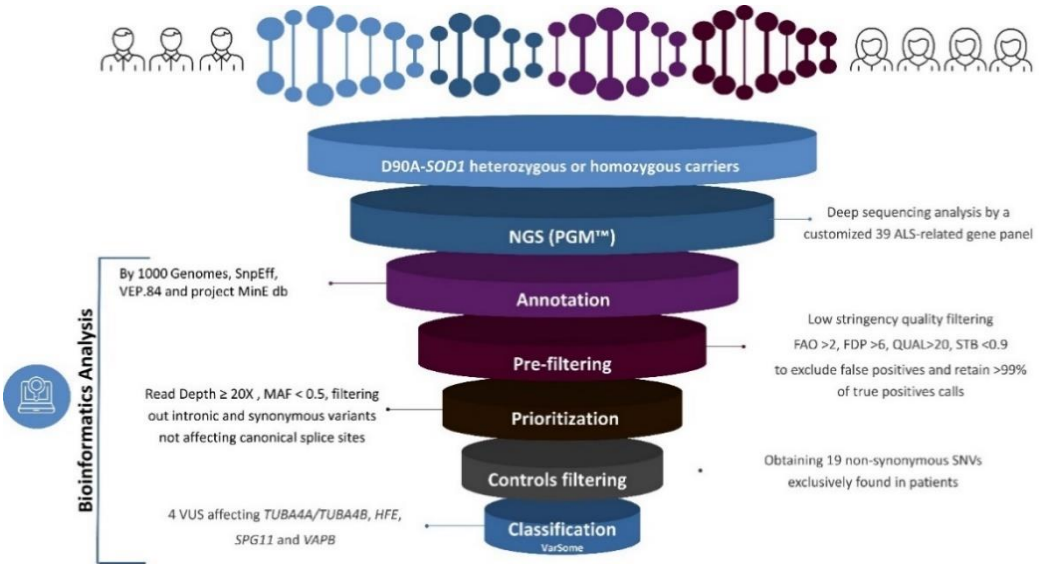


Figure 1. Workflow of the variant prioritization strategy adopted. The figure was created editing funnels and healthcare infographics provided by https://infograpia.com/, accessed on 12 July 2021.

Table 3. Variants identified in p.D91A-*SOD1* patients by targeted NGS are annotated with frequency and classification.

Gene	V	rs ID	VA	VF	MAF	VarSome	ClinVar	Proj. MinE	CADD PHRED score	Ref.
DCTN1	c.*21C>T	rs11555696	3' UTR	0.14	0.022720/ 0.02188	LB	B	n/a	8.217	[45]
TUBA4A	c.227-74C>T	rs45488900	Intron	0.28	-/0.12853	VUS	n/a	n/a	5.309	[46]
TUBA4B	n.-1456G>T		Upstream		n/a	n/a	n/a	n/a	5.309	n/a
NEK1	c.2255A>G; p.Glu752Gly	rs34099167	Missense	0.4	0.14032/ 0.09304	B	B	n/a	25.1	n/a
NEK1	c.1388C>T; p.Ala463Val	rs34540355	Missense	0.14	0.035127/ 0.03250	B	B/LB	0.0593/0.0603	16.28	n/a
HFE	c.-48C>G	rs41266793	5' UTR	0.14	-/-	VUS	n/a	n/a	0.233	n/a
FIG4	c.*29G>A	rs10659	3' UTR	0.14	0.046277/ 0.07829	B	B	n/a	0.408	n/a
SETX	c.*849G>T	rs74975459	3' UTR	0.14	-/0.01150	B	B	n/a	3.561	n/a
SETX	c.59G>A; p.Arg20His	rs79740039	Missense	0.14	0.009062/ 0.00653	B	B	0.00882/0.00764	0.166	[47–49]
SPG11	c.7069C>T; p.Leu2357Phe	rs139334167	Missense	0.14	-/0.00083	VUS	CIoP	0.00183/0.00164	25.8	[50]

SPG11	c.2083G>A; p.Ala695Thr	rs78183930	Missense	0.14	0.012527/ 0.01897	B	B	0.0121/0.0150	26.8	[50–54]
PG11	c.1108G>A; p.Glu370Lys	rs77697105	Missense	0.14	0.016740/ 0.02196	B	B/LB	0.0182/0.0194	21.6	n/a
FUS	c.*910C>T	rs118018900	Downstream	0.14	0.04772/ 0.02543	B	B	n/a	1.015	n/a
PFN1	c.-342T>C	rs148770753	5' UTR	0.14	-/0.00848	B	n/a	n/a	10.19	n/a
VAPB	c.*753C>G	rs6070466	3' UTR	0.14	0.01155/ 0.00452	B	B	n/a	15.82	n/a
VAPB	c.*1265G>C	rs7400	3' UTR	0.14	0.05550/ 0.09272	B	B	n/a	14.85	n/a
VAPB	c.*2819A>G	rs74568509	3' UTR	0.14	0.04862/ 0.07642	B	B	n/a	0.673	n/a
VAPB	c.*4520T>C	rs763514	3' UTR	0.14	0.05098/ 0.10229	B	B	n/a	0.83	n/a
VAPB	c.*6182C>T	rs76708676	3' UTR	0.14	0.02834/ 0.01132	VUS	B	n/a	0.515	n/a
APEX1	c.*2A>T	rs17112002	3' UTR	0.14	0.003300/ 0.00373	LB	n/a	n/a	5.682	[55]

* V, variant; VA, variant annotation; VF, variant frequency out of 7 p.D91A patients; MAF, minor allele frequency in Exome Aggregation Consortium (ExAC) and Genomes Aggregation Database (GnomAD); VarSome, the human genomic variant search engine—(B) benign, (LB) likely benign, (VUS) uncertain significance; Project MinE variant browser, a database providing information on genetic variations found in WGS of ALS patients and controls—AF ALS cases/AF controls; ClinVar, archive of interpretations of clinically relevant variants—(B) benign, (LB) likely benign, (CioP) conflicting interpretations of pathogenicity; n/a, data not available. CADD PHRED score, combined annotation dependent depletion.

In Table 4 we list the co-occurrence of variants in ALS-associated genes in each patient: 19 variants were exclusively found in patients and not in controls; two variants, in PON1 and/or GRN, instead, were detected in both patients and controls (except for P1 and C1). Only two variants were shared by two or three patients, including one segregating with the genotype zygosity. This was the case of TUBA4A/TUBA4B rs45488900, shared by two p.D91A-homozygous SALS patients of our cohort, P4 (Het) and P5 (Hom). The entire list of annotated variants detected in patients and controls is available in Supplementary Table S4 (on web).

Table 4. Co-occurrence of variants in ALS-associated genes in each patient.

Patient ID	FALS or SALS	p.D91A SOD1	Genes	HGVSc/HGVSp	Zygosity	rs ID	Ref.
P1	FALS	Hom	DCTN1	c.*21C>T	Het	rs11555696	
			NEK1	c.2255A>G; p.Glu752Gly	Hom	rs34099167	
P2	FALS	Hom	PON1	c.575A>G; p.Gln192Arg	Het	rs662	[56]
			GRN	c.*78C>T	Het	rs5848	[57]
P3	FALS	Hom	HFE	c.-48C>G	Het	rs41266793	
			FIG4	c.*29G>A	Het	rs10659	
			GRN	c.*78C>T	Het	rs5848	[57]
P4	SALS	Hom	TUBA4A/TUBA4B	c.227-74C>T/ n.-1456G>T	Het	rs45488900	
			SPG11	c.7069C>T; p.Leu2357Phe	Het	rs139334167	
			VAPB	c.*753C>G	Het	rs6070466	
			PON1	c.575A>G; p.Gln192Arg	Het	rs662	[56]
			GRN	c.*78C>T	Hom	rs5848	[57]
P5	SALS	Hom	TUBA4A/TUBA4B	c.227-74C>T/ n.-1456G>T	Hom	rs45488900	
			NEK1	c.2255A>G; p.Glu752Gly	Het	rs34099167	
			PFN1	c.-342T>C	Het	rs148770753	
			VAPB	c.*6182C>T	Het	rs76708676	
			APEX1	c.*2A>T	Het	rs17112002	[55]
			GRN	c.*78C>T	Het	rs5848	[57]
P6	SALS	Het	NEK1	c.1388C>T; p.Ala463Val	Het	rs34540355	
			SPG11	c.2083G>A; p.Ala695Thr	Het	rs78183930	[50,53]
			SPG11	c.1108G>A; p.Glu370Lys	Het	rs77697105	
			PON1	c.575A>G; p.Gln192Arg	Het	rs662	[56]
			GRN	c.*78C>T	Het	rs5848	[57]
P7	SALS	Het	NEK1	c.2255A>G; p.Glu752Gly	Het	rs34099167	
			SETX	c.*849G>T	Het	rs74975459	
			SETX	c.59G>A; p.Arg20His	Het	rs79740039	[48]
			FUS	c.*910C>T	Het	rs118018900	
			VAPB	c.*1265G>C	Het	rs7400	
			VAPB	c.*2819A>G	Het	rs74568509	
			VAPB	c.*4520T>C	Het	rs763514	
			PON1	c.575A>G; p.Gln192Arg	Het	rs662	[56]

To investigate the possible relation between variants detected by our analysis and gene loci affecting gene expression, particularly for untranslated region variants, we also calculated their potential effect on gene expression through their mapping on eQTLs. Data retrieved by GTEx Portal v.8 and corrected by a Bonferroni correction test (Supplementary Table S5) showed a tissue-specific effect for three out of 19 variants queried in non-diseased tissues of interest (rs11555696, 183 rs34099167, and rs118018900).

4. Discussion

In this study, we performed targeted NGS analysis in a small group of south Italian ALS patients, previously genetically characterized as p.D91A carriers, hypothesizing that genetic factors in other ALS-related genes, in combination with the p.D91A-SOD1 variant, may contribute to the different disease phenotypes in homozygous and heterozygous cases. Recently, Sanger sequencing analysis performed in 997 ALS patients from southern Italy by our research group, revealed that 2% of patients had SOD1 mutations [33]. In particular, the frequency of p.D91A affected individuals represented 0.8% of all ALS cases diagnosed (0.6% p.D91A-hom and 0.2% p.D91A-het). These data are in line with the frequency of this mutation reported by other Italian research groups [57]. Previous reports demonstrated the absence of a neuroprotective factor in the genomic region near SOD1 in p.D91A-homozygous ALS patients, suggesting the existence of a putative protective factor modulating the phenotype located elsewhere in the genome [31]. The present investigation showed that p.D91A-heterozygous and -homozygous ALS cases do not contain a genetic modifier near SOD1, nor near ALS-linked genes, highlighting the presence of unique variant gene sets in each patient (Table 4). A similar conclusion was recently published in a study investigating the A90V-SOD1 mutation in SALS patients, suggesting that additional genetic variants could contribute to disease penetrance [24]. We identified 19 non-synonymous variants, including four of uncertain significance, in ALS-D91A carriers. Most substitutions (13/19) exclusively found in p.D91A patients were in non-coding regions of ALS-related genes, while the remaining (6/19) were missense mutations without any clear evidence of pathogenic effects. Although large genes (i.e., SETX or NEK1) have more chance to accumulate rare variants, the 19 variants identified in genes related to ALS were exclusively found in patients and not in controls. However, no clear genotype–phenotype correlation was established due to the small sample size. All but one patient (P1) showed the presence of variants already identified as risks factors for neurodegenerative diseases [56–58], rs662 (PON1) and rs5848 (GRN) (Table 4). Previous studies inconsistently suggested an effect

of PON1 SNPs on ALS susceptibility, and rs662 was associated with bulbar onset and reduced survival in ALS cases very recently [55]. However, all the patients carrying this risk factor in our cohort showed a spinal onset of the disease. The GRN rs5848 polymorphism was reported in Alzheimer's Disease (AD) and Parkinson's Disease (PD) patients as risk factor for ubiquitin- and TDP-43 -positive frontotemporal degeneration [59]. Interestingly, this genetic variant lies in the binding-site for the miR-659 of the 3'UTR of GRN and may alter gene regulation [59]. Two variants classified as VUS were already described, the rs139334167 in SPG11 and the rs45488900 in TUBA4A. The missense variation affecting SPG11 was previously reported in a case of PD [44], while the rs45488900 affects TUBA4A but also the upstream region (n.-1456G>T) of TUBA4B. The encoded protein of the latter, was found differentially over-expressed in post-mortem pre-frontal cortex samples of patients affected by atypical ubiquitin-positive frontotemporal lobar degeneration, characterized by ubiquitin and FUS positive inclusions, while TUBA4A was down-expressed in the cerebellum of the same group of patients when compared to controls [58]. Interestingly, rs45488900 was shared by p.D91A-homozygous SALS patients showing a slow course but with a different clinical picture of the disease (Table 2; Table 4). Due to the small number of homozygous SALS in our cohort, no inference on the possible role of this variant in association with the phenotype could be made, but this aspect remains noteworthy and deserves to be explored in a larger number of cases. We did not find any literature reports describing two other variants classified as VUS. These substitutions, positioned in the 5'UTR of HFE (rs41266793) and in the 3' UTR of VAPB (rs76708676), were detected in the homozygous case P3, and in the P5 patient with the longest observed disease duration (22 years) (Table 3). Further investigating the potential influence of identified variants on gene expression by eQTL analysis, we also observed that two polymorphisms, rs34099167 and rs11555696, were associated with deregulated expression levels of DCTN1 and NEK1 in some of the tissues considered (Supplementary Table S5). Interestingly, the altered expression of these two genes was already found by our research team using unsupervised clustering of gene expression in motor cortex samples, identifying two transcriptome-based SALS subgroups of patients [60]. In particular, NEK1 was found down-regulated in one cluster of patients while the second one was characterized by increased expression of DCTN1 and reduced levels of SOD1 [60]. NEK1 belongs to NIMA-related serine/threonine kinases family and is involved in mitochondrial membrane regulation, DNA damage response, ciliogenesis and maintenance of the cytoskeleton network [29], while DCTN1 is a motor protein involved in dynein-mediated axonal retrograde transport and ciliogenesis [61]. Moreover, dynactin (Dctn1) acting with

overexpressed dynamitin (Dctn2), was shown to produce a late-onset progressive motor neuron disease inhibiting axonal transport in transgenic mice [62]. In our data, rs34099167 and/or rs11555696 were detected in 3 out of 7 patients, both homozygotes and heterozygotes, showing different clinical features, although no correlation between phenotype and genotype was established (Table 4). Recently, the analysis of two different cohorts, with a majority of apparently sporadic cases, showed an oligogenic basis of ALS associated with earlier age of disease [63,64]. Differences in genotype-phenotype correlations would have considerable therapeutic implications. ALS is a devastating pathology in which multiple variants cooperate in influencing disease onset, severity or duration and, until now, no truly effective treatment exists [65]. Thanks to recent efforts to selectively treat SOD1-related ALS patients, ASO therapies designed to knock-down the expression of the gene have emerged [66]. To this regard are of particular interest two ongoing phase 3 studies using intrathecally administered ASO Tofersen, an orphan drug capable of reducing SOD1 protein levels [67]. To ensure the appropriate recruitment of patients to clinical trials it appears evident the importance to establish if p.D91A mutation is pathogenic in the heterozygous state and if other contributing factors influence the phenotype.

5. Conclusions

Our study suggests the possibility that additional genetic factors contribute to the individual oligogenic basis of p.D91A-SOD1 carriers. In particular, all patients, except for P2 carrying only risk factors, showed an oligogenic pattern in line with the model proposed for ALS etiopathogenesis in which mutations in two or more genes are required to develop the disease, but they are not all necessarily truly pathogenic [1,19]. Increasing the number of sequenced p.D91A patients could be useful in identifying emerging genetic factors and improving disease diagnosis, as well as guiding precision medicine.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/ 10.3390/genes12121843/s1, Table S1: Quality metrics and coverage analysis parameters; Table S2: The table illustrates the alignment quality parameters (to hg19) related to the samples analysed in the same run; Table S3: Variants per patient; Table S4: The entire list of annotated variants detected in both patients and controls in available; Table S5: Variants associated with eQTL effects in different tissues.

Author Contributions: Conceptualization, F.L.C., G.G., I.L.S. and S.C.; methodology, G.G. and G.M.; validation, B.P.; formal analysis, G.G.; investigation, G.G; resources, F.L.C., I.L.S. and S.C.; data curation, G.G. and F.L.C.; writing—original draft preparation, G.G., G.M. and F.L.C.; writing— review and editing, S.C., I.L.S., S.A. and F.L.C.; visualization, G.G. and B.P.; supervision, F.L.C.; funding acquisition, F.L.C. and S.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by a special award to the Department of Pharmacy, Health and Nutritional Sciences of University of Calabria (Italy) (Department of Excellence, Italian Law232/2016) from the Italian Ministry of Education, University and Research (MIUR) and by MIUR ex 60% to F.L.C., by MIUR grant CTN01_00177_817708 and the European Social Fund operational program for the Sicily region (Italy) "Development and application of biosensoristic technologies in genomics" PO FESR Sicily 2014–2020 grant to S.C.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of the Azienda Ospedaliero Universitaria of Bari.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study for research purpose.

Data Availability Statement: The data presented in this study are available in the results and supplementary material sections.

Acknowledgments: The authors would like to thank the Project MinE GWAS Consortium. We also thank the Genotype-Tissue Expression (GTEx) Project that was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, Genes 2021, 12, 1843 10 of 12 NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from the GTEx Portal release V8 (dbGaP accession phs000424.v8.p2) on 04/30/21. Moreover, G.G. addresses special thanks to Martina Michalikova and the colleagues of the course of Writing Scientist Pro 2021 for their suggestions to improve the microstructure of Section 1.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interests.

References

1. Renton, A.E.; Chiò, A.; Traynor, B.J. State of play in amyotrophic lateral sclerosis genetics. Nat. Neurosci. **2014**, 17, 17–23.

2. Rowland, L.P.; Shneider, N.A. Amyotrophic Lateral Sclerosis. N. Engl. J. Med. **2001**, 344, 1688–1700.

3. Bandres-Ciga, S.; Noyce, A.J.; Hemani, G.; Nicolas, A.; Calvo, A.; Mora, G.; Arosio, A.; Barberis, M.; Bartolomei, I.; Battistini, S.; et al. Shared polygenic risk and causal inferences in amyotrophic lateral sclerosis. Ann. Neurol. **2019**, 85, 470–481.

4. Felbecker, A.; Camu, W.; Valdmanis, P.N.; Sperfeld, A.D.; Waibel, S.; Steinbach, P.; Rouleau, G.A.; Ludolph, A.C.; Andersen, P.M. Four familial ALS pedigrees discordant for two SOD1 mutations: Are all SOD1 mutations pathogenic? J. Neurol. Neurosurg. Psychiatry **2010**, 81, 572–577.

5. Farrugia Wismayer, M.; Farrugia Wismayer, A.; Pace, A.; Vassallo, N.; Cauchi, R.J. SOD1 D91A variant in the southernmost tip of Europe: A heterozygous ALS patient resident on the island of Gozo. Eur. J. Hum. Genet. **2021**, 27, 27–30.

6. Richards, S.; Aziz, N.; Bale, S.; Bick, D.; Das, S.; Gastier-Foster, J.; Grody,W.W.; Hegde, M.; Lyon, E.; Spector, E.; et al. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet. Med. **2015**, 17, 405–424.

7. Rehm, H.L.; Berg, J.S.; Brooks, L.D.; Bustamante, C.D.; Evans, J.P.; Landrum, M.J.; Ledbetter, D.H.; Maglott, D.R.; Martin, C.L.;Nussbaum, R.L.; et al. ClinGen — The Clinical Genome Resource. N. Engl. J. Med. **2015**, 372, 2235–2242.

8. Andersen, P.M.; Nilsson, P.; Ala-Hurula, V.; Keränen, M.L.; Tarvainen, I.; Haltia, T.; Nilsson, L.; Binzer, M.; Forsgren, L.; Marklund, S.L. Amyotrophic lateral sclerosis associated with homozygosity for an Asp90Ala mutation in CuZn-superoxide dismutase. Nat. Genet. **1995**, 10, 61–66.

9. Andersen, P.M.; Forsgren, L.; Binzer, M.; Nilsson, P.; Ala-Hurula, V.; Keränen, M.L.; Bergmark, L.; Saarinen, A.; Haltia, T.; Tarvainen, I.; et al. Autosomal recessive adult-onset amyohophic lateral sclerosis associated with homozygosity for Asp90A1a CuZn-superoxide dismutase mutation A clinical and genealogical study of 36 patients. Brain **1996**, 119, 1153–1172.

10. Robberecht,W.; Aguirre, T.; Van Den Bosch, L.; Tilkin, P.; Cassiman, J.J.; Matthijs, G. D90A heterozygosity in the SOD1 gene is associated with familial and apparently sporadic amyotrophic lateral sclerosis. Neurology **1996**, 47, 1336–1339.

11. Skvortsova, V.I.; Limborska, S.A.; Slominsky, P.A.; Levitskaya, N.I.; Levitsky, G.N.; Shadrina, M.I.; Kondratyeva, E.A. Sporadic ALS associated with the D90A CU,ZN superoxide dismutase mutation in Russia. Eur. J. Neurol. **2001**, 8, 167–172.

12. Giannini, F.; Battistini, S.; Mancuso, M.; Greco, G.; Ricci, C.; Volpi, N.; Del Corona, A.; Piazza, S.; Siciliano, G. D90A-SOD1 mutation in ALS: The first report of heterozygous Italian patients and unusual findings. Amyotroph. Lateral Scler. **2010**, 11, 216–219.

13. Parton, M.J.; Broom, W.; Andersen, P.M.; Al-Chalabi, A.; Nigel Leigh, P.; Powell, J.F.; Shaw, C.E. D90A-SOD1 mediated amyotrophic lateral sclerosis: A single founder for all cases with evidence for a Cis-acting disease modifier in the recessive haplotype. Hum. Mutat. **2002**, 20, 473.

14. Luisa Conforti, F.; Sprovieri, T.; Mazzei, R.; Patitucci, A.; Ungaro, C.; Zoccolella, S.; Magariello, A.; Bella, V.L.; Tessitore, A.; Tedeschi, G.; et al. Further evidence that D90A-SOD1 mutation is recessively inherited in ALS patients in Italy. Amyotroph. Lateral Scler. **2009**, 10, 58–60.

15. Bernard, E.; Pegat, A.; Svahn, J.; Bouhour, F.; Leblanc, P.; Millecamps, S.; Thobois, S.; Guissart, C.; Lumbroso, S.; Mouzat, K. Clinical and Molecular Landscape of ALS Patients with SOD1 Mutations: Novel Pathogenic Variants and Novel Phenotypes. A Single ALS Center Study. Int. J. Mol. Sci. **2020**, 21, 6807.

16. Khoris, J.; Moulard, B.; Briolotti, V.; Hayer, M.; Durieux, A.; Clavelou, P.; Malafosse, A.; Rouleau, G.A.; Camu,W. Coexistence of dominant and recessive familial amyotrophic lateral sclerosis with the D90A Cu,Zn superoxide dismutase mutation within the same country. Eur. J. Neurol. **2000**, 7, 207–211.

17. Lowry, J.L.; Ryan, É.B.; Esengul, Y.T.; Siddique, N.; Siddique, T. Intricacies of aetiology in intrafamilial degenerative disease. Brain Commun. **2020**, 2.

18. Battistini, S.; Giannini, F.; Greco, G.; Bibbò, G.; Ferrera, L.; Marini, V.; Causarano, R.; Casula, M.; Lando, G.; Patrosso, M.C.; et al. SOD1 mutations in amyotrophic lateral sclerosis. J. Neurol. **2005**, 252, 782–788.

19. van Blitterswijk, M.; van Es, M.A.; Hennekam, E.A.M.; Dooijes, D.; van Rheenen, W.; Medic, J.; Bourque, P.R.; Schelhaas, H.J.; van
der Kooi, A.J.; de Visser, M.; et al. Evidence for an oligogenic basis of amyotrophic lateral sclerosis. Hum. Mol. G enet. **2012**, 21, 3776–3784.

20. Andersen, P.M. Amyotrophic lateral sclerosis associated with mutations in the CuZn superoxide dismutase gene. Curr. Neurol. Neurosci. Rep. **2006**, 6, 37–46.

21. Hand, C.K.; Mayeux-Portas, V.; Khoris, J.; Briolotti, V.; Clavelou, P.; Camu,W.; Rouleau, G.A. Compound heterozygous D90A and D96N SOD1 mutations in a recessive amyotrophic lateral sclerosis family. Ann. Neurol. **2001**, 49, 267–271.

22. Parton, M.J.; Andersen, P.M.; Broom,W.J.; Shaw, C.E. Compound heterozygosity and variable penetrance inSOD1 amyotrophic lateral sclerosis pedigrees. Ann. Neurol. **2001**, 50, 553.

23. Luigetti, M.; Conte, A.; Madia, F.; Marangi, G.; Zollino, M.; Mancuso, I.; Dileone, M.; Del Grande, A.; Di Lazzaro, V.; Tonali, P.A.; et al. Heterozygous SOD1 D90A mutation presenting as slowly progressive predominant upper motor neuron amyotrophic lateral sclerosis. Neurol. Sci. **2009**, 30, 517–520.

24. Kuuluvainen, L.; Kaivola, K.; Mönkäre, S.; Laaksovirta, H.; Jokela, M.; Udd, B.; Valori, M.; Pasanen, P.; Paetau, A.; Traynor, B.J.; et al. Oligogenic basis of sporadic ALS The example of SOD1 p.Ala90Val mutation. Neurol Genet **2019**, 5, 335.

25. Feneberg, E.; Turner, M.R.; Ansorge, O.; Talbot, K. Amyotrophic lateral sclerosis with a heterozygous D91A SOD1 variant and classical ALS-TDP neuropathology. Neurology **2020**, 95, 595–596.

26. Mackenzie, I.R.A.; Bigio, E.H.; Ince, P.G.; Geser, F.; Neumann, M.; Cairns, N.J.; Kwong, L.K.; Forman, M.S.; Ravits, J.; Stewart, H.; et al. Pathological TDP-43 distinguishes sporadic amyotrophic lateral sclerosis from amyotrophic lateral sclerosis with SOD1 mutations. Ann. Neurol. **2007**, 61, 427–434.

27. Okamoto, Y.; Ihara, M.; Urushitani, M.; Yamashita, H.; Kondo, T.; Tanigaki, A.; Oono, M.; Kawamata, J.; Ikemoto, A.; Kawamoto, Y.; et al. An autopsy case of SOD1-related ALS with TDP-43 positive inclusions. Neurology **2011**, 77, 1993–1995.

28. Lattante, S.; Ciura, S.; Rouleau, G.A.; Kabashi, E. Defining the genetic connection linking amyotrophic lateral sclerosis (ALS) with frontotemporal dementia (FTD). Trends Genet. **2015**, 31, 263–273.

29. Nguyen, H.P.; Van Broeckhoven, C.; van der Zee, J. ALS Genes in the Genomic Era and their Implications for FTD. Trends Genet. **2018**, 34, 404–423.

30. Morgan, S.; Shatunov, A.; Sproviero, W.; Jones, A.R.; Shoai, M.; Hughes, D.; Al Khleifat, A.; Malaspina, A.; Morrison, K.E.; Shaw, P.J.; et al. A comprehensive analysis of rare genetic variation in amyotrophic lateral sclerosis in the UK. Brain **2017**, 140, 1611–1618.

31. Broom, W.J.; Johnson, D.V.; Garber, M.; Andersen, P.M.; Lennon, N.; Landers, J.; Nusbaum, C.; Russ, C.; Brown, R.H. DNA sequence analysis of the conserved region around the SOD1 gene locus in recessively inherited ALS. Neurosci. Lett. **2009**, 463, 64–69.

32. Brooks, B.R.; Miller, R.G.; Swash, M.; Munsat, T.L. El Escorial revisited: Revised criteria for the diagnosis of amyotrophic lateral sclerosis. Amyotroph. Lateral Scler. Other Mot. Neuron Disord. **2000**, 1, 293–299.

33. Ungaro, C.; Sprovieri, T.; Morello, G.; Perrone, B.; Spampinato, A.G.; Simone, I.L.; Trojsi, F.; Monsurrò, M.R.; Spataro, R.; La Bella, V.; et al. Genetic investigation of amyotrophic lateral sclerosis patients in south Italy: A two-decade analysis. Neurobiol. Aging **2021**, 99, 99.e7–99.e14.

34. Al-Chalabi, A.; Van Den Berg, L.H.; Veldink, J. Gene discovery in amyotrophic lateral sclerosis: Implications for clinical management. Nat. Rev. Neurol. **2017**, 13, 96–104.

35. Ungaro, C.; Citrigno, L.; Trojsi, F.; Sprovieri, T.; Gentile, G.; Muglia, M.; Monsurrò, M.R.; Tedeschi, G.; Cavallaro, S.; Conforti, F.L. ALS and CHARGE syndrome: A clinical and genetic study. Acta Neurol. Belg. **2018**, 4, 629–635.

36. Roy, S.; Coldren, C.; Karunamurthy, A.; Kip, N.S.; Klee, E.W.; Lincoln, S.E.; Leon, A.; Pullambhatla, M.; Temple-Smolkin, R.L.; Voelkerding, K.V.; et al. Standards and Guidelines for Validating Next-Generation

Sequencing Bioinformatics Pipelines. J. Mol. Diagnostics **2018**, 20, 4–27.

37. van der Spek, R.A.A.; van Rheenen, W.; Pulit, S.L.; Kenna, K.P.; van den Berg, L.H.; Veldink, J.H. The project MinE databrowser: Bringing large-scale whole-genome sequencing in ALS to researchers and the public. Amyotroph. Lateral Scler. Front. Degener. **2019**, 20, 432–440.

38. Damiati, E.; Borsani, G.; Giacopuzzi, E. Amplicon-based semiconductor sequencing of human exomes: Performance evaluation and optimization strategies. Hum. Genet. **2016**, 135, 499–511

39. Summa, S.D.; Malerba, G.; Mori, A.; Mijatovic, V.; Pinto, R.; Tommasi, S. Quality measures to improve variant calling of Ion Torrent data. In Proceedings of the NETTAB & IB, Bari, Italy, 14–16 October 2015.

40. Kopanos, C.; Tsiolkas, V.; Kouris, A.; Chapple, C.E.; Albarca Aguilera, M.; Meyer, R.; Massouras, A. VarSome: The human genomic variant search engine. Bioinformatics **2019**, 35, 1978–1980.

41. Landrum, M.J.; Lee, J.M.; Benson, M.; Brown, G.R.; Chao, C.; Chitipiralla, S.; Gu, B.; Hart, J.; Hoffman, D.; Jang, W.; et al. ClinVar: Improving access to variant interpretations and supporting evidence. Nucleic Acids Res. **2018**, 46, D1062–D1067.

42. Rentzsch, P.; Witten, D.; Cooper, G.M.; Shendure, J.; Kircher, M. CADD: Predicting the deleteriousness of variants throughout the
human genome. Nucleic Acids Res. **2019**, 47, D886–D894.

43. Aguet, F.; Barbeira, A.N.; Bonazzola, R.; Jo, B.; Kasela, S.; Liang, Y.; Parsana, P.; Aguet, F.; Battle, A.; Brown, A.; et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science **2020**, 369, 1318–1330.

44. Hysi, P.G.; Choquet, H.; Khawaja, A.P.;Wojciechowski, R.; Tedja, M.S.; Yin, J.; Simcoe, M.J.; Patasova, K.; Mahroo, O.A.; Thai, K.K.; et al. Meta-analysis of 542,934 subjects of European ancestry identifies new genes and mechanisms predisposing to refractive error and myopia. Nat. Genet. **2020**, 52, 401–407.

45. Vuckovic, D.; Bao, E.L.; Akbari, P.; Lareau, C.A.; Mousas, A.; Jiang, T.; Chen, M.-H.; Raffield, L.M.; Tardaguila, M.; Huffman, J.E.; et al. The Polygenic and Monogenic Basis of Blood Traits and Diseases. Cell **2020**, 182, 1214–1231.e11.

46. Bernard, V.; Stricker, S.; Kreuz, F.; Minnerop, M.; Gillessen-Kaesbach, G.; Zühlke, C. Ataxia with oculomotor apraxia type 2: Novel mutations in six patients with juvenile age of onset and elevated serum _-fetoprotein. Neuropediatrics **2008**, 39, 347–350.

47. Arning, L.; Epplen, J.T.; Rahikkala, E.; Hendrich, C.; Ludolph, A.C.; Sperfeld, A.-D. The SETX missense variation spectrum as evaluated in patients with ALS4-like motor neuron diseases. Neurogenetics **2013**, 14, 53–61.

48. Høyer, H.; Braathen, G.J.; Busk, Ø.L.; Holla, Ø.L.; Svendsen, M.; Hilmarsen, H.T.; Strand, L.; Skjelbred, C.F.; Russell, M.B. Genetic Diagnosis of Charcot-Marie-Tooth Disease in a Population by Next-Generation Sequencing. Biomed Res. Int. **2014**, 2014, 1–13.

49. Ghani, M.; Lang, A.E.; Zinman, L.; Nacmias, B.; Sorbi, S.; Bessi, V.; Tedde, A.; Tartaglia, M.C.; Surace, E.I.; Sato, C.; et al. Mutation analysis of patients with neurodegenerative disorders using NeuroX array. Neurobiol. Aging **2015**, 36, 545.e9–545.e14.

50. Crimella, C.; Arnoldi, A.; Crippa, F.; Mostacciuolo, M.L.; Boaretto, F.; Sironi, M.; D'Angelo, M.G.; Manzoni, S.; Piccinini, L.; Turconi, A.C.; et al. Point mutations and a large intragenic deletion in SPG11 in complicated spastic paraplegia without thin corpus callosum. J. Med. Genet. **2009**, 46, 345–351.

51. Pensato, V.; Castellotti, B.; Gellera, C.; Pareyson, D.; Ciano, C.; Nanetti, L.; Salsano, E.; Piscosquito, G.; Sarto, E.; Eoli, M.; et al. Overlapping phenotypes in complex spastic paraplegias SPG11, SPG15, SPG35 and SPG48. Brain **2014**, 137, 1907–1920.

52. Morgan, S.; Shoai, M.; Fratta, P.; Sidle, K.; Orrell, R.; Sweeney, M.G.; Shatunov, A.; Sproviero,W.; Jones, A.; Al-Chalabi, A.; et al. Investigation of next-generation sequencing technologies as a diagnostic tool for amyotrophic lateral sclerosis. Neurobiol. Aging **2015**, 36, 1600.e5–1600.e8.

53. Chrestian, N.; Dupré, N.; Gan-Or, Z.; Szuto, A.; Chen, S.; Venkitachalam, A.; Brisson, J.-D.; Warman-Chardon, J.; Ahmed, S.; Ashtiani, S.; et al. Clinical and genetic study of hereditary spastic paraplegia in Canada. Neurol. Genet. **2017**, 3, e122.

54. Krüger, S.; Battke, F.; Sprecher, A.; Munz, M.; Synofzik, M.; Schöls, L.; Gasser, T.; Grehl, T.; Prudlo, J.; Biskup, S. Rare Variants in Neurodegeneration Associated Genes Revealed

by Targeted Panel Sequencing in a German ALS Cohort. Front. Mol. Neurosci. **2016**, 9, 92.

55. Verde, F.; Tiloca, C.; Morelli, C.; Doretti, A.; Poletti, B.; Maderna, L.; Messina, S.; Gentilini, D.; Fogh, I.; Ratti, A.; et al. PON1 is a disease modifier gene in amyotrophic lateral sclerosis: Association of the Q192R polymorphism with bulbar onset and reduced survival. Neurol. Sci. **2019**, 40, 1469–1473.

56. Chen, Y.; Li, S.; Su, L.; Sheng, J.; Lv,W.; Chen, G.; Xu, Z. Association of progranulin polymorphism rs5848 with neurodegenerative diseases: A meta analysis. J. Neurol. **2015**, 262, 814–822.

57. Perrone, B.; Conforti, F.L. Common mutations of interest in the diagnosis of amyotrophic lateral sclerosis: How common are common mutations in ALS genes? Expert Rev. Mol. Diagn. **2020**, 20, 703–714.

58. Martins-De-Souza, D.; Guest, P.C.; Mann, D.M.; Roeber, S.; Rahmoune, H.; Bauder, C.; Kretzschmar, H.; Volk, B.; Baborie, A.; Bahn, S. Proteomic analysis identifies dysfunction in cellular transport, energy, and protein metabolism in different brain regions of atypical frontotemporal lobar degeneration. J. Proteome Res. **2012**, 11, 2533–2543.

59. Rademakers, R.; Eriksen, J.L.; Baker, M.; Robinson, T.; Ahmed, Z.; Lincoln, S.J.; Finch, N.; Rutherford, N.J.; Crook, R.J.; Josephs, K.A.; et al. Common variation in the miR-659 binding-site of GRN is a major risk factor for TDP43-positive frontotemporal dementia. Hum. Mol. Genet. **2008**, 17, 3631–3642.

60. Aronica, E.; Baas, F.; Iyer, A.; ten Asbroek, A.L.M.A.; Morello, G.; Cavallaro, S. Molecular classification of amyotrophic lateral sclerosis by unsupervised clustering of gene expression in motor cortex. Neurobiol. Dis. **2015**, 74, 359–376.

61. Chen, T.Y.; Syu, J.S.; Han, T.Y.; Cheng, H.L.; Lu, F.I.;Wang, C.Y. Cell Cycle-Dependent Localization of Dynactin Subunit p150glued at Centrosome. J. Cell. Biochem. **2015**, 116, 2049–2060.

62. LaMonte, B.H.; Wallace, K.E.; Holloway, B.A.; Shelly, S.S.; Ascaño, J.; Tokito, M.; Van Winkle, T.; Howland, D.S.; Holzbaur, E.L.F. Disruption of dynein/dynactin inhibits axonal transport in motor neurons causing late-onset progressive degeneration. Neuron **2002**, 34, 715–727.

63. McCann, E.P.; Henden, L.; Fifita, J.A.; Zhang, K.Y.; Grima, N.; Bauer, D.C.; Chan Moi Fat, S.; Twine, N.A.; Pamphlett, R.; Kiernan, M.C.; et al. Evidence for polygenic and oligogenic basis of Australian sporadic amyotrophic lateral sclerosis. J. Med. Genet. **2021**, 58, 87–95.

64. Shepheard, S.R.; Parker, M.D.; Cooper-Knock, J.; Verber, N.S.; Tuddenham, L.; Heath, P.; Beauchamp, N.; Place, E.; Sollars, E.S.A.; Turner, M.R.; et al. Value of systematic genetic screening of patients with amyotrophic lateral sclerosis. J. Neurol. Neurosurg. Psychiatry **2021**, 92, 510–518.

65. Taylor, J.P.; Brown, R.H.; Cleveland, D.W. Decoding ALS: From genes to mechanism. Nature **2016**, 539, 197–206.

66. Abati, E.; Bresolin, N.; Comi, G.; Corti, S. Silence superoxide dismutase 1 (SOD1): A promising therapeutic target for amyotrophic lateral sclerosis (ALS). Expert Opin. Ther. Targets **2020**, 24, 295–310.

67. Amado, D.A.; Davidson, B.L. Gene therapy for ALS: A review. Mol. Ther. **2021.**

Supplementary data

Supplementary Table S1. Quality metrics and coverage analysis parameters are showed in the table.

Sample ID	Bases	≥ Q20	Reads	Mapped reads	On target	Mean depth
P1	177,643,157	165,654,399	670,34	669,185	98.08%	772.1
P2	102,241,575	94,962,476	394,631	393,957	98.27%	444.8
P3	186,869,242	173,718,879	710,203	709,223	97.65%	809.3
P4	140,082,743	129,756,787	530,358	529,626	97.76%	606.5
P5	278,633,241	260,071,606	1,090,256	1,088,487	98.31%	1,212
P6	188,166,608	174,162,787	707,28	706,370	98.00%	816.9
P7	131,761,923	122,366,505	491,362	490,686	98.58%	573.1

Supplementary Table S2. The table illustrates the alignment quality parameters (to hg19) related to the samples analysed in the same run.

	Alignment Quality		
	AQ17	AQ20	Perfect
Total Number of Bases [Mbp]	1.73 G	1.68 G	1.26 G
Mean Length [bp]	261	255	199
Longest Alignment [bp]	513	513	492
Mean Coverage Depth	0.6	0.5	0.4

Supplementary Table S3. The table below shows the values of average variant quality, number of variants either SNV or Indel, number of known variants, transition/transversion ratio, and heterozygous/homozygous ratio after filtering variants for quality parameters, minor allele frequency in European population and minimum value of germline variant read depth.

Sample ID	Average variant quality	Variants*	SNV	Indels	Known variants	Ti/Tv ratio	Het/Hom ratio
P1	4,868.53	85 (of 178)	80	5	85	3 (60/20)	5.07 (71/14)
P2	4,902.22	84 (of 183)	78	0	84	2.25 (54/24)	5.00 (70/14)
P3	5,177.62	82 (of 196)	78	8	82	2.3913 (55/23)	12.67 (76/6)
P4	4,797.13	92 (of 193)	89	5	92	2.56 (64/25)	5.57 (78/14)
P5	5,097.94	78 (of 181)	73	5	78	1.7037 46/27)	8.75 (70/8)
P6	4,920.41	85 (of 180)	83	4	85	3.15 (63/20)	7.50 (75/10)
P7	5,044.23	97 (of195)	91	8	97	2.13793 (62/29)	4.71 (80/17)

* Number of variants after filtering out of number of variants called.

Supplementary Table S4 is available online at https://www.mdpi.com/article/ 10.3390/genes12121843/s1.

Supplementary Table S5: Variants associated with eQTL effects in different tissues.

Gene	SNP	Alternative Allele	GTEx eQTL SNP Gene Associations in multiple tissues (P < 0.002632) *				
			Brain Cortex	Brain Frontal Cortex	Spinal Cord	Muscle Skeletal	Whole Blood
DCTN1	rs11555696	A	n/a	n/a	**5,70E-04**	**3,60E-21**	**1,30E-08**
TUBA4A	rs45488900	T	-	-	-	-	-
NEK1	rs34099167	C	**9,20E-05**	**3,80E-05**	n/a	**1,20E-15**	n/a
NEK1	rs34540355	A	-	-	-	n/a	-
HFE	rs41266793	G	n/a	n/a	n/a	n/a	n/a
FIG4	rs10659	A	-	-	n/a	-	-
SETX	rs74975459	A	n/a	n/a	n/a	n/a	n/a
SETX	rs79740039	T	n/a	n/a	n/a	n/a	n/a
SPG11	rs139334167	A	n/a	n/a	n/a	n/a	n/a
SPG11	rs78183930	T	n/a	n/a	n/a	n/a	n/a
SPG11	rs77697105	T	n/a	n/a	n/a	n/a	n/a
FUS	rs118018900	T	-	-	-	-	**6,50E-05**
PFN1	rs148770753	G	n/a	n/a	n/a	n/a	n/a
VAPB	rs6070466	G	n/a	n/a	n/a	n/a	n/a
VAPB	rs7400	C	n/a	-	n/a	-	-
VAPB	rs74568509	G	n/a	-	n/a	-	-
VAPB	rs763514	C	n/a	-	n/a	-	-
VAPB	rs76708676	T	n/a	n/a	n/a	n/a	n/a
APEX1	rs17112002	T	n/a	n/a	n/a	n/a	n/a

*p<0.05: value indicates the existence of a predicted relationship between a genetic variant and a gene locus affecting gene expression in the queried tissues calculated by a T-statistics. T-test results were then corrected through the Bonferroni multiple testing correction P<0.002632; (n/a): No association was found for SNP; (-) no significant association available (p-value > 0.002632). The normalized effect size (NES) of eQTLs was computed as the effect of the alternative allele (ALT) relative to the reference allele (REF) in the human genome reference GRCh38/hg38 (i.e., the eQTL effect allele is the ALT allele).

Test	Alpha
T-statistics	0,05
Bonferroni correction	0,00263

Chapter 4

High-Throughput Genetic Testing in ALS: targeted next generation sequencing study of an Italian ALS Cohort

1. Introduction

Amyotrophic lateral sclerosis (ALS) is characterized by an oligogenic architecture. The lack of segregation of pathogenic variants in some affected family members, the detection of pathogenic variants (non-penetrant variants) in healthy individuals, and the co-occurrence of mutations in several ALS genes support the idea that the genetic of ALS is very complex (36). During the last years, the development of high- throughput sequencing technologies and the screening of large cohorts of affected patients has led to the identification of approximatively 40 disease-associated genes and a significant number of genetic variants (37). To date, NGS sequencing has been introduced in clinical settings and genetic testing is increasingly being offered to ALS patients, with relevant psychological, social and ethical consequences, which need to be considered. In fact, genetic counselling has become an essential step in the management of ALS patients, not only in familial cases, but also in sporadic ones.

Based on these premises, in this study we investigated a cohort of patients by targeting NGS with the aim of revealing the presence of genetic changes that may cause or contribute to ALS pathogenesis.

2. Patients and Methods

2.1 Patients

Informed consent was obtained from all the participants included in the study. ALS patients (5 familial and 64 sporadic cases, 39 women and 30 men) from southern Italy unrelated families, diagnosed with ALS according to the El Escorial criteria (3), were evaluated by second-generation sequencing. Twenty-eight (28/69; 40%) patients were tested previously negative for mutations in *SOD1*, *FUS*, *C9orf72*, and *TARDBP (Chapter 2)* and forty-one (41/69, 60%) were new recruited at our university. Age at onset, gender distribution, and site of symptom onset (bulbar and/or spinal-onset) are shown in Table 1.

Table 1. Clinical data of ALS patients.

Gender, n (%)	
Female	39 (56,5%)
Male	30 (43,4%)
Family history, n (%)	
fALS	5 (7,2 %)
sALS	64 (92,7 %)
Site of onset, n (%)	
Bulbar	20 (28,9%)
Spinal	48(69,5%)
Mean age at onset (y)	62

Key: fALS, familial Amyotrophic Lateral Sclerosis; sALS: sporadic Amyotrophic Lateral sclerosis.

2.2 Targeted sequencing and bioinformatic analysis

Genomic DNA was extracted from peripheral blood using Wizard Genomic DNA Purification Kit (Promega, Madison, WI, USA). The quantity and quality of DNA were determined by NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer (Thermo Fisher Scientific) and tested on an agarose gel.

Next-generation sequencing (NGS) analysis was carried out using a targeted gene panel, including ALS-related genes, designed in our lab using the Ion Ampliseq Designer (https://ampliseq.com/browse.action). The entire list of potentially causative and susceptibility genes included in the panel is shown in Table 2.

Libraries were prepared using the custom Ion AmpliSeq kit Plus (Thermo Fisher Scientific), quantified with a Qubit™ Fluorometer (Invitrogen), and diluted at 100 pM. Barcoded libraries were sequenced on Ion Torrent sequencing platforms (Thermo Fisher Scientific), using standard kits and following the manufacturer's instructions. Base-calling, pre-processing of the reads, and short read alignment was performed with the Ion S5 Torrent Suite software (Thermo Fisher Scientific).

Variant annotation was mainly performed with Annovar (38) and human genome sequence builds GRCh37/hg19 was used as a genome reference sequence. For the study of the variants, a filtering algorithm was used which takes into consideration variants: (a) exonic, intronic, 5′UTR, 3′UTR, upstream, downstream, ncRNA; (b) non-synonymous variations; (c) minor allele frequency (MAF) <0.01 of the European-derived population.

According to the hypothesis that the causative mutations of this rare disease are not present in the general population, single nucleotide polymorphisms (SNPs) with allele frequencies >0,0001 identified in The Exome Aggregation Consortium (ExAC) database and 1000

Genomes were filtered out. Low-quality variants were also filtered out as described previously (*see Chapter 2*). A positive SOD1 DNA sample, previously genetically confirmed with heterozygous likely pathogenic mutation in SOD1 (p.D91A), was included in this study.

To confirm rare variants detected by NGS, we used we used the SeqStudio™ Genetic Analyzer (SeqA7 software, version 6.0; Applied biosystem). The frequencies were compared with in-house ALS database and, in ALS patients and controls belonging to the project MinE database (http://databrowser.projectmine.com/, accessed on 20 April 2021) (39).

Mutations were classified according to the 2015 American College of Medical Genetics Standards and Guidelines for the interpretation of sequence variants. In particular, SOD1 variants, were classified using the criteria as reported in our recent work (40).

To confirm rare variants detected by NGS, we analyzed the variant sites using a SeqStudio™ Genetic Analyzer and analyzed with SeqA7 software, version 6.0 (Applied biosystem).

The *C9orf72* was not included in this panel because of the nature of mutation (hexanucleotide repeat expansion) and, it was tested separately as above mentioned (*see Chapter 2*).

Table 2. Genes analysed by NGS.

Gene name	Chromosomal location	OMIM	Inheritance	Phenotype
SOD1	21q22.11	147450	AD, (AR)	Amyotrophic Lateral sclerosis 1
TARDBP	1p36.22 32	605078	AD, (AR)	Amyotrophic Lateral sclerosis 10, with or without frontotemporal dementia
FUS	16p11.1	137070	AD, (AR)	Amyotrophic Lateral sclerosis 6, with or without frontotemporal dementia
ANG	14q11.2	105850	AD	Amyotrophic Lateral sclerosis 9
CHCHD10	22q11.23	615903	AD	Frontotemporal dementia and/or Amyotrophic Lateral Sclerosis 2
OPTN	10p13	602432	AD, (AR)	Amyotrophic Lateral Sclerosis 12
ALS2	2q33.1	606352	AR	Amyotrophic Lateral Sclerosis 2, juvenile
TBK1	12q14.2	604834	AD	Frontotemporal dementia and/or Amyotrophic Lateral Sclerosis 4
SPG11	9p21.2	610844	AR	Amyotrophic lateral sclerosis 5, juvenile
PFN1	17p13.2	176610	AD	Amyotrophic lateral sclerosis 18
TUBA4A	2q35	191110	AD	Amyotrophic lateral sclerosis 22 with or without frontotemporal dementia
DAO	12q24.11	124050	AD	
GRN	17q21.31	138945	AD	
SETX	12q3.3 5	608465	AD	Amyotrophic lateral sclerosis 4, juvenile
TAF15	17q12	601574	AD	

VCP	9p13.3	611745		
UBQLN2	Xp11.21	300264	XLD	Amyotrophic lateral sclerosis 15, with or without frontotemporal dementia
SQSTM1	5q35.3	601530	AD	Frontotemporal dementia and/or amyotrophic lateral sclerosis 3
FIG4	6q21	609390	AD	Amyotrophic lateral sclerosis 11
ANO2	12p13.31	610109	-	
APEX	14q11.2	300773	-	
CEP112	17q24.1	618980	-	
CHMP2B	3p11.2	609512	AD	Frontotemporal dementia and/or amyotrophic lateral sclerosis 7
CNTF	11q12.1	118945	-	Amyotrophic Lateral Sclerosis 1; Als1
DCTN1	2p13.1	601143	AD	Amyotrophic lateral sclerosis, susceptibility
HFE	6p22.2	613609	-	
MATR3	5q31.2	164015	AD	Amyotrophic lateral sclerosis 21
NEFH	22q12.2	162230	AD, AR	Amyotrophic lateral sclerosis, susceptibility
PON1	7q21.3	168820	-	
PON2	7q21.3	602447	-	
PRPH	12q13.12	170710	AD, AR	Amyotrophic lateral sclerosis, susceptibility
SMN1	5q13.2	600354	-	Spinal muscular atrophy-1
SMN2	5q13.2	601627	-	Spinal muscular atrophy-2
VAPB	20q13.31	605704	AD	Amyotrophic lateral sclerosis 8
VEGFA	6p21.1	192240	-	
CCNF	16p13.3	600227	AD	Frontotemporal dementia and/or amyotrophic lateral sclerosis 5
NEK1	4q33	604588	AD	Amyotrophic lateral sclerosis, susceptibility
HNRNPA1	12q13.13	615426	AD	Inclusion body myopathy with early-onset Paget disease without frontotemporal dementia 3
ERBB4	2q34	615515	AD	Amyotrophic lateral sclerosis 19

Key: OMIM, Online Mendelian Inheritance, AD, autosomal dominant; AR, autosomal recessive, XLD, X-linked dominant.

3. Results

A total of 69 ALS patients (7,2%, 5/69 fALS, 92,7 % (64/69) sALS; mean age at disease onset: 62 years; age range: 28–81 years) were analyzed in this study. All patients presented UMN and LMN signs, and 16% also presented bulbar signs. We utilized a 39-ALS gene panel NGS-based targeted sequencing and variant prioritization was carried out filtering for Read Depth ≥ 30X (minimum read depth for germline variants calling), and for MAF—European Ancestry Population—Freq < 0.0001. According to NGS data, we identified a total of 139

exonic variants. Among them, 70 and 42 were synonymous and non-synonymous changes, respectively and the remaining 18 were frameshift deletions and substitutions. In this context, we focused our analysis on non-synonymous coding variants because these are considered to account for a significant amount of the functional variations that cause pathology (41).

The workflow adopted for variants identification is shown in Figure 1.

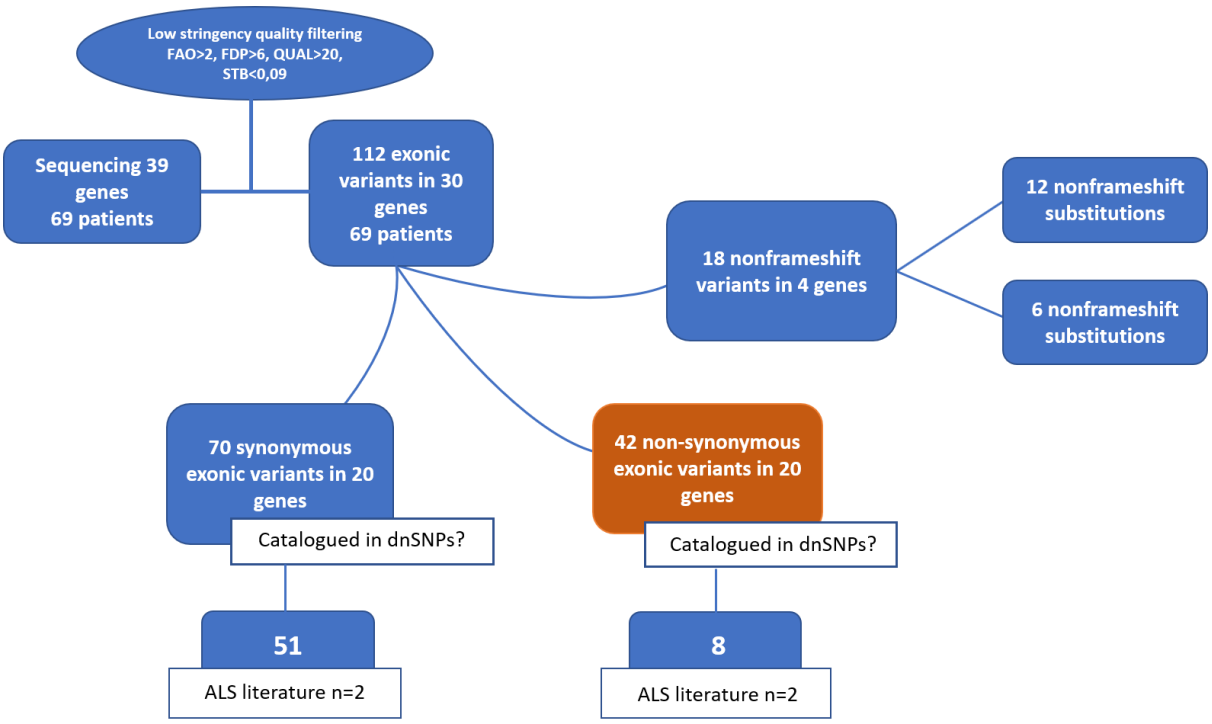


Figure 1. Variants identification workflow. Key: FAO, flow space alternate allele observations; FDP, flow space read depth; QUAL, variant quality; STB, strand bias ratio.

In 27 patients (out of 69 screened) we identified at least one non-synonymous variant (Table 3). On the basis of NGS results, the most frequently mutated genes were *NEFH* (7/69; 11,6%) and *ANO2*, *DCTN1*, *HFE and TAF15* (3/69; 4,34%). Eleven patients (4/69; 5,8%) were carriers of a previously reported variant associated with ALS (Table 3).

A very rare heterozygous mutation, c.137T>G (p.F46C) was detected in SOD1 of exon 2 in an ALS patient with early disease onset (Figure 2).

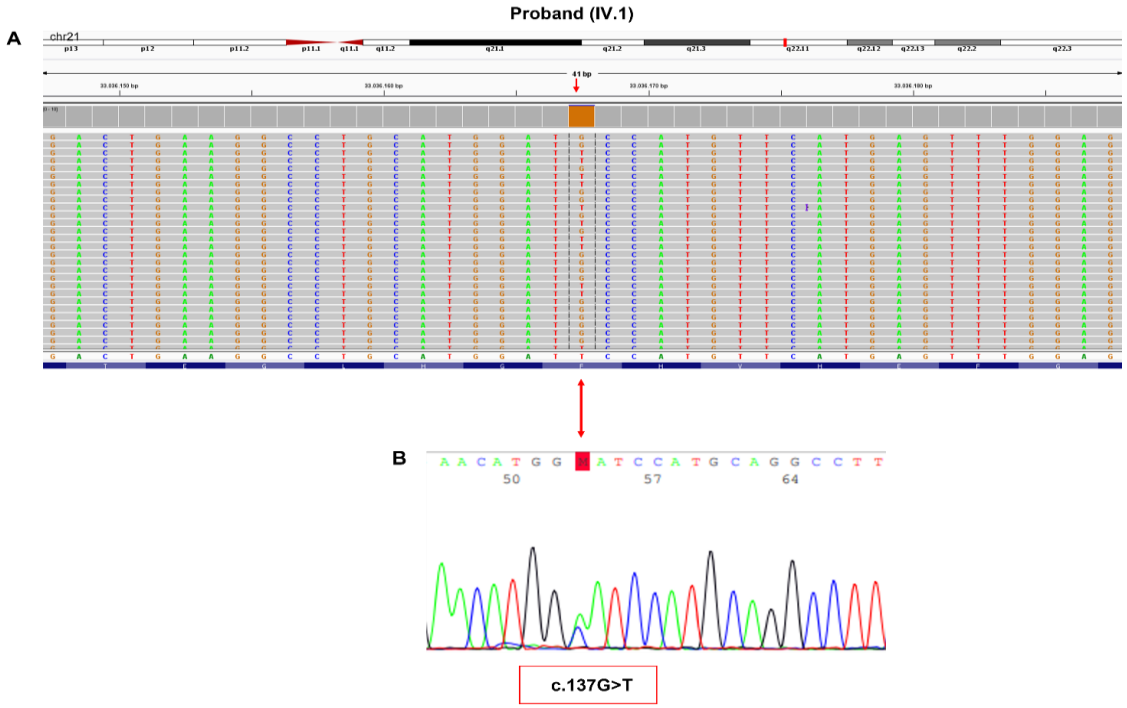


Figure 2. (A) Snapshot of Integrative Genomics Viewer, showing the SOD1 variant c.137G>T (p.Phe46Cys) in the patient. (B) Sanger sequencing showing the c.127G>T pathogenic heterozygous SOD1 variant.

In details, patient ALS-57, a 32-year-old man, developed a left-hand strength deficit with difficulty spreading the fingers and reduced trophic of the hand muscles at age 30 years. He reported two relatives (mother and aunt) died following a cerebral aneurysm and no ALS familial history (Figure 3).

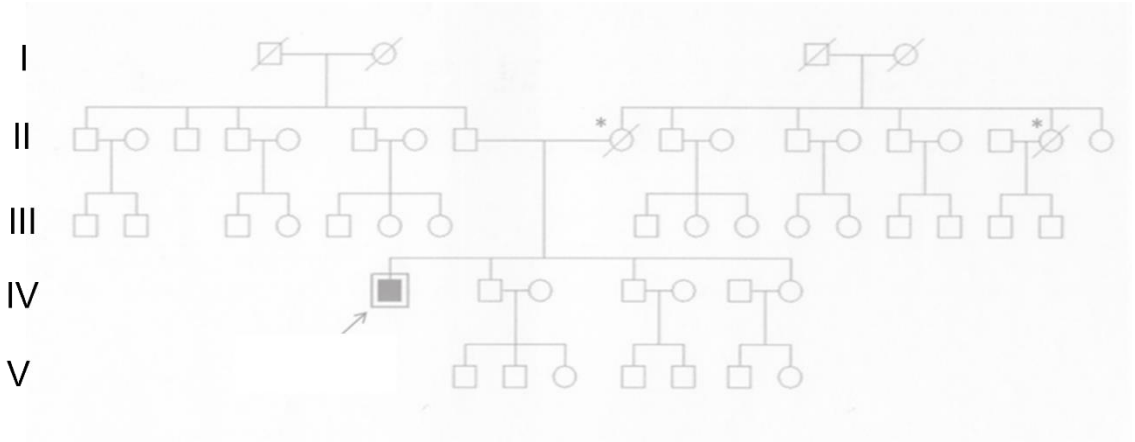


Figure 3. ALS-57 family pedigree. Asterisks indicate the patient's mother and maternal aunt died following a cerebral aneurysm.

The p. Phe46Cys variant was considered for ACMG re-classification in our recent paper (40) and an in-deep functional study is underway on it (data not shown).

Among the novel rare variants, eleven (11/42, 26%) non-synonymous missense variants were consistently predicted to be deleterious by Sorting Intolerant from Tolerant software (SIFT; sift.jcvi.org), Polymorphism Phenotyping (PolyPhen-2; genetics.bwh.harvard.edu/pph2/), and Mutation Taster (https://www.mutationtaster.org/) (Table 3).

Table 3. Rare non-synonymous single-nucleotide variants identified.

	Gene name	N° patients (%)	cDNA change	Protein change	dbSNP[155] ID	ACMG classification	Population MAF 1000 Genomes/ ExAC	ProjectMinE	Sift	Polyphen	mutation Taster	CADD phred	Literature ALS
Non-synonymous variants													
NM_020919	*ALS2*	1 (1,4%)	c.G37A	p.G13R	rs367871772	vus	- /0.000067	*n/a*	0,007 (D)	0,973 (D)	D	26,4	*(42)*
		1 (1,4%)	c.G4069A	p.V1357I	-	vus	-	n/a	0,528 (T)	0,001 (B)	D	20,05	
NM_001278597	*ANO2*	1 (1,4%)	c.T1772	p.I591T	-	vus	-	n/a	0,03 (D)	0,97 (D)	D	27,5	
		1 (1,4%)	c.T1763C	p.V588A	-	vus	-	n/a	0.006 (D)	0,97 (D)	D	17,95	
		1 (1,4%)	c.C1916T	p.A639V	rs749348947	vus	- /0.000009	n/a	0,001(D)	0,69 (P)	D	33	
NM_001199165	*CEP112*	1 (1,4%)	c.G2050A	p.D684N	-	vus	-	n/a	0,006 (D)	0,403 (P)	B	24,3	
		1 (1,4%)	c.A2354T	p.K785I	-	vus	-	n/a	0,008 (D)	0,452 (P)	N	28,4	
NM_004082	*DCTN1*	1 (1,4%)	c.A1445G	p.E482G	-	vus	-	n/a	0,002 (D)	0,99 (D)	D	28,1	
		1 (1,4%)	c.A1403G	p.N468S	-	vus	-	n/a	0,02 (D)	0,784 (P)	D	23,5	
		1 (1,4%)	c.T1508C	p.V503A	-	vus	-	n/a	0,046 (D)	0,95 (D)	D	26,4	
NM_014043	*CHMP2B*	1 (1,4%)	c.A400G	p.K134E	-	vus	-	n/a	0,004(D)	0,965 (D)	D	31	
NM_005235	*ERBB4*	1 (1,4%)	c.C1382T	p.T461I	-	vus	-	n/a	0,318 (T)	0,946 (D)	D	23,6	
		1 (1,4%)	c.T1814C	p.F605S	-	vus	-	n/a	0,002 (D)	1 (D)	D	24,7	
NM_139009	*HFE*	1 (1,4%)	c.A539G	p.D180G	-	vus	-	n/a	0,206 (T)	0,003 (B)	N	7,42	
		1 (1,4%)	c.A269T	p.K90M	-	vus	-	n/a	0,008 (D)	0,631 (P)	D	26	
		1 (1,4%)	c.T860G	p.I287S	-	vus	-	n/a	0,072 (T)	0,004 (B)	N	10,16	
NM_001194956	*MATR3*	1 (1,4%)	c.G10A	p.A4T	-	LB	-	n/a	0 (D)	0,001 (B)	D	17,95	
NM_021076	*NEFH*	2 (2,9%)	c.T2009A	p.V670E	-	LB	_	n/a	0,064 (T)	0,35 (P)	N	6,22	*(43)*
		1 (1,4%)	c.T1780C	p.S594P	-	vus	-	n/a	0,113 (T)	0,002 (B)	N	5,56	
		1 (1,4%)	c.A2100C	p.E700D	-	vus	-	n/a	0,011 (D)	0,878 (D)	N	23,5	
		1 (1,4%)	c.T1631A	p.V544D	-	vus	-	n/a	0,001 (D)	0,287 (P)	N	6,92	
		1 (1,4%)	c.G2140A	p.E714K	-	vus	-	n/a	0,001 (D)	0,968 (D)	N	20,1	
		1 (1,4%)	c.A1907G	p.E636G	-	LB	-	n/a	0,22 (T)	0.002 (B)	N	15,18	
		1 (1,4%)	c.G2146C	p.A716P	-	vus	-	n/a	0,002 (D)	0,542 (P)	N	15,3	
NM_001199397	*NEK1*	1 (1,4%)	c.C577T	p.L193F	rs748415363	vus	- /0.00003	n/a	0,007 (D)	0,099 (B)	D	26	

		1 (1,4%)	c.T1235C	p.V412A	-	vus	-	n/a	0,011 (D)	0,948 (D)	D	24,1	
NM_001008213	*OPTN*	1 (1,4%)	c.G1724A	p.C575Y	-	vus	-	n/a	0	1 (D)	D	26,4	
NM_006262	*PRPH*	1 (1,4%)	c.T23C	p.L8P	-	vus	-	n/a	0,215 (T)	0,202 (P)	D	14,39	
		2 (2,9%)	c.G697A	p.E233K	-	vus	-	*0.00115/0*	0,001 (D)	0,821 (P)	D	35	
NM_015046	*SETX*	1 (1,4%)	c.T2921C	p.I974T	rs765768857	LB	- / 0.000008	n/a	0,185 (T)	0,011 (B)	N	0,113	
		1 (1,4%)	c.A6085G	p.K2029E	rs746525639	VUS	-	n/a	0,024 (D)	0,175 (P)	N	28,3	
NM_0107411	SMN2	3 (4,3%)	c.G859C	p.G287R	-	vus	-	n/a	0,003 (D)	0,995 (D)	D	23,8	
NM_000454	*SOD1*	1 (1,4%, ALS_57)	c.T137G	p.F46C	rs121912457	LP	-	n/a	0 (D)	0,99 (D)	A	25,1	*(44)*
NM_025137	*SPG11*	2 (2,9%)	c.G6166A	p.E2056K	-	LB	-	n/a	0,154 (T)	0,629 (P)	D	27,5	
		1 (1,4%)	c.G7243A	p.A2415T	-	vus	-	n/a	0,04 (D)	0,99 (D)	D	34	
NM_013444	*UBQLN2*	2 (2,9%)	c.C401T	p.T134I	rs764837088	vus	- / 0.00001	n/a	0,4 (T)	0,99 (D)	D	12,48	
		1 (1,4%)	c.A1793G	p.E598G	-	vus		n/a	0,001 (D)	0,67 (P)	D	23,5	
NM_007126	*VCP*	1 (1,4%)	c.T647C	p.I216T	-	vus		n/a	0 (D)	1 (D)	D	29,4	
NM_003376	*VEGFA*	1 (1,4%)	c.T173C	p.L58P	-	vus		n/a	0 (D)	0,94 (D)	D	25,1	
NM_139215	*TAF15*	1 (1,4%)	c.G1637C	p.S546T	rs376749338	likely benign	-/0.000008	n/a	-	-	N	-	
		1 (1,4%)	c.G1652C	p.G551A	-	vus		n/a	0,39 (T)	0,172 (P)	N	14,96	
		1 (1,4%)	c.A1412G	p.Y471C	-	vus		n/a	-	0,001 (B)	N	15,41	

dbSNP, database of single nucleotide polymorphism; ACMG, American College of Medical Genetics and Genomics: the criteria according to our study (P.Ruffo et al., 2022) were applied; LB, likely benign LP, likely pathogenic; vus, variant of uncertain significance; MAF, Minor Allele Frequency; ExAC, Exome Aggregation Consortium; Project MinE variant browser, a database providing information on genetic variations found in WGS of ALS patients and controls; AF ALS cases/AF controls; SIFT: (T) = tolerated and (D)= Deleterious; Polyphen: (B)= benign, (P)=possibly damaging and (D)= deleterious; Mutation Taster: N=polymorphism, A=disease causing automatic and D= disease causing; CADD phred– Combined Annotation Dependent Depletion.

Moreover, in our cohort, eight patients (8/69, 11,5%) carried multiple rare variants in more than one gene. Six of these seven patients (6/8, 85,7%) had changes in two different genes, one patient (1/8, 12,5%) in four distinct genes, one patient (1/8, 12,5%) carried changes in five diverse genes (Table 4).

Table 4. List of patients carrying more than one variant.

ID_pt	fALS/sALS	sex	Site of symptom onset	Age of onset	Variant 1 (ACMG)	Variant 2 (ACMG)	Variant 3 (ACMG)	Variant 4 (ACMG)	Variant 5 (ACMG)
ALS_10	sALS	F	Spinal	28	*SMN2* p.G287R (vus)	*SPG11* p.E2056K (LB)			
ALS_13	sALS	M	Spinal	37	*DCTN1* p.E482G (vus)	*DCTN1* p. N468S (vus)	*NEFH* p.E700D (vus)		
ALS_14	sALS	F	Spinal	80	*VCP* p.I216T (vus)	*PRPH* p.E233K (vus)	*CEP112* p.D684N (vus)	*UBQLN2* p. E598G (vus)	*HFE* p.D180G (vus)
ALS_17	sALS	F	Bulbar	54	*SMN2* p.G287R (vus)	N*EFH* p.V544D (vus)			
ALS_18	sALS	M	Spinal	76	*ERBB4* p.F605S (vus)	*ERBB4* p.T461I (vus)	*CHMP2B* p.K134E (vus)	*HFE* p.K90M (vus)	*SPG11* p.A2415T (vus)
ALS_19	sALS	M	Bulbar	50	*DCTN1* p.V503A (vus)	*NEK1* p.Q911E (vus)			
ALS_25	sALS	F	Spinal	48	*MATR3* p.A4T (LB)	*ANO2* p.V588A (vus)			
ALS_52	sALS	M	Spinal	32	*SOD1 p.F46C (LP)*	*UBQLN2 p.C575T (vus)*			

Key: ID_pt, patient identification code; American College of Medical Genetics and Genomics (ACMG) classification. VUS= uncertain significance variant; LB=likely benign.

4. Discussion

ALS should not be considered as a predominantly sporadic disease with a small percentage of monogenic familial cases, but rather as a pathogenesis shaped by the synergy between rare oligogenic variants. Although gene-by-bene screening by Sanger sequencing is available for testing genes most commonly causative of the disease, the genetic landscape of ALS has been increased by the development of next-generation sequencing.

In this work, by using NGS we analysed 39 ALS-related genes in a southern Italian cohort (69 ALS patients), of which 40% (28/69) tested negative for mutations in *SOD1, FUS, C9orf72, and TARDBP* genes (*see Chapter 2*).

To identify ALS-causative variants that predispose patients to ALS we focused our analysis on very rare (MAF<0.0001) exonic non-synonymous variants. Among all the variants detected, only one was classified as likely pathogenic, the p.F46C in SOD1 (Table 3). This mutation results in an amino acid substitution of phenylalanine with cysteine in a conserved domain of SOD1 protein. It is unclear whether this is a new mutation and/or with variable disease penetrance due to the lack of genetic testing by the parents. The p.F46C variant has previously been reported in one Italian fALS patient (44). Despite the different forms and ages of onset, a comparison of the clinical characteristics of the two patients revealed a similar picture (Table 5).

Table 5. Clinical characteristics of patients with the F46C- SOD1 variant.

sALS/fALS	Age	Genetic Status	Phenotype	Onset	Neurological examination	*References*
sALS	32	Het	Spinal	Hands strenght defecit with reduction of thophism and degenerative disc disease.	Tongue and four limbs fasciculations. Hypostenic tremor.	This work
fALS	59	Het	Spinal	Mild muscle weakness in distal segment of the upper limbs.	Mild dysarthria and dyshagia, muscle weakness and wasting in the four limbs, particularly in the distal districts, and diffuse muscle fasciculations.	(44)

The absence in the literature database of any functional study conducted to assess the pathogenicity of this mutation, prompt us to investigate the role of p.F46C-SOD1 variant by in vitro characterization (data not shown).

In this study, we identified rare variants with uncertain significance and a low frequency (30/42,71,4%), as well as previously described changes (7%) (Table 3) and found that 10% (7/69) of ALS patients showing co-occurrence of at least 2 variants in 2 independent genes (Table 4). It is important to note that the co-occurrence rate of gene variants estimated by our study is in line with the variability of frequencies reported by most previous studies (41). In particular, 75% (6/8) of all patients with sALS carried at least 1 mutation or a VUS (45, 46).

Interestingly, it has been described that oligogenic inheritance is also associated with an earlier age of onset and rapid disease progression (27) (47). In our cohort, the ALS patient carrying both likely pathogenic variant p.F46C in *SOD1* and the VUS p.C575Y in OPTN developed symptoms more years earlier than other individuals. Obviously, the small number of patients examined does not allow us to reveal these observations.

Despite we identified many rare variants with uncertain significance and a low frequency in our cohort, 29% of patients did not show rare variants in any of the 39 ALS genes analyzed, suggesting that more efforts are needed to identify additional genes affecting disease risk.

However, we focused our study on very rare exonic non-synonymous substitutions, but other exonic variants could be involved in the disease such as the frameshift and synonymous variants (48) (see Figure 1). Many variants annotated as synonymous are typically considered as neutral or nearly neutral but are not synonymous at all (49). Indeed, many of these variations could lead to mRNA degradation and missing protein. Thus, there is an important effect on variation interpretation and disease diagnosis and it should be necessary to re-evaluate and sometimes to re-analyze studies based on this assumption.

Chapter 5

Validation of Optical Mapping for the Molecular Diagnosis of C9orf72 and the detection of large repeat expansion

1. Introduction

Specific variant types, which are not commonly detectable by currently used analysis methods, such as repeat expansions (50), variants in non-coding regions (51) and structural variants, are frequent in neurodegenerative diseases (52). Among these, repeat expansions are the prevalent cause of several neurological disorders (53). The pathogenic mechanisms for repeat diseases include a loss of protein function or a gain of function at the RNA or protein level, according to the type and location of the repeat (50). Often, the length of the repeat expansions can be related to the disorder's severity providing important information on the clinical course and genetic counselling (50). Moreover, increasing the number of repeats often involves an earlier onset and more severe disease in successive generations (53). Globally, *C9orf72* -mediated neurodegeneration accounts for approximately one-third of ALS families, 5-10% of sporadic ALS cases, and for about a quarter of familial FTD cases (25%) (13). The frequencies of repeats expansion of the pathogenic *C9orf72* vary according to ethnic/geographical origin and is the most frequent genetic cause of ALS in Europe and North America while lower frequencies in Asian countries is reported (13, 54). GGGGCC (G4C2) hexanucleotide repeat expansions are associated with a mean disease duration of 2.9 ± 2.8 years, and spinal onset (involving limb muscles) is more recurrent than bulbar onset (including involvement of swallowing and speech) (54% vs 39%, respectively) (55). Normal repeats of G4C2 are ~25 units or less, while patients with ALS can have hundreds to thousands of hexanucleotides, with serious consequences (13). Currently, although repeat – primed (RP) - PCR RP and Southern blot (SB) (Figure 1) are methods suited to detect polynucleotide expansions, the first is typically used to detect small expansions (< 80 bp), while the second one is labor intensive, prone to background noise, requires high signal strength, and a trained staff member, limiting *C9orf72* sizing to a few specialized laboratories. In addition to these techniques, implementation of next generation sequencing (NGS) has certainly increased recognition and diagnosis of this group of diseases (53). However, whole genome sequencing, while being able to distinguish between unexpanded and expanded alleles with high sensitivity and specificity by accurately calculating the size of alleles smaller than the read length, would appear to underestimate the size of large expansions, such as *FMR1, DMPK, FXN and C9orf7*2 (53). Moreover, in the ALS case the extensive somatic mosaicism at the locus (*C9orf72*) implies that the estimated repeat lengths from blood may not accurately reflect the repeats in vulnerable brain tissue (56). Properly due to these tissue differences and the difficulty in precisely defining repeat width, C9-mediated disease differs from most other repeat expansion diseases because is still difficult

to establish a clear correlation between repeat length and disease severity. Furthermore, hypermethylation phenomena that affect the expansion and gene expression and thus disease, could explain why some individuals with C9 expansions remain asymptomatic until late in life (57). Therefore, we decided to test a novel imaging and mapping tool, Optical Genome Mapping (OGM), already used in cancer and genetic disease. In particular, OGM developed by Bionano Genomics is a new high-sensitivity technology for the detection of large SVs (35, 58) which, although, to date, only little applied in investigations in undiagnosed individuals, could support difficulties existing in diagnoses (59). In this work we wanted to explore the use of OGM as a diagnostic tool for *C9orf72*. We evaluated a cohort of 15 *C9orf72*-ALS patients suggesting how it might provide more detailed information on the scale of repeat expansion and somatic instability. Although, to date, it is a method rarely applied in the investigation of undiagnosed individuals, it could support the difficulties of existing diagnoses. This part of work was carried out in collaboration with the Department of Neuromuscular Disorders, UCL Queen Square Institute of Neurology, London, UK.

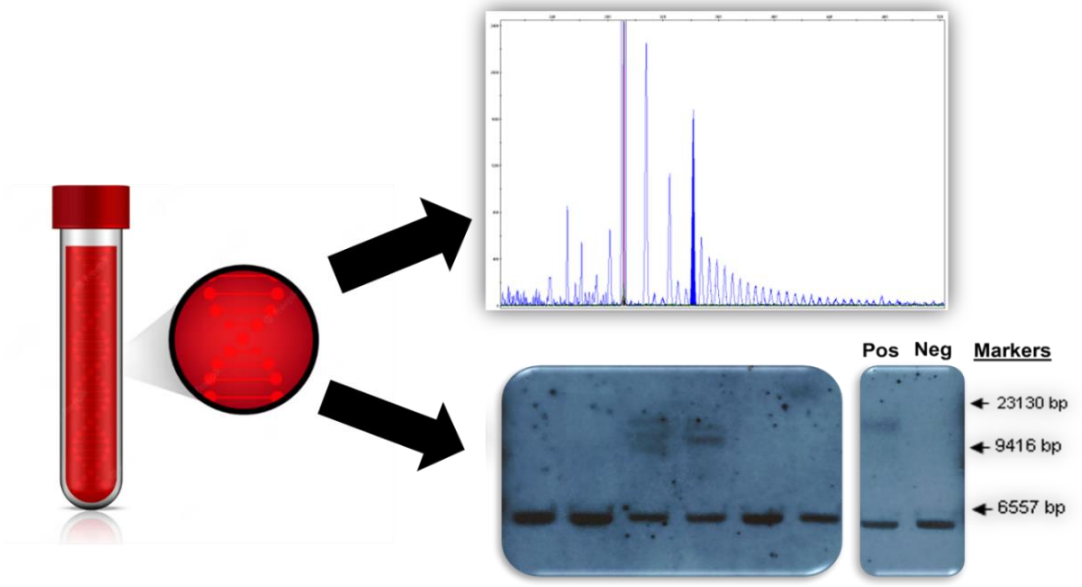


Figure 1. Current molecular diagnostic methods. The figure shows two current diagnostic methods for repeat expansions diagnosis: repeat- primed PCR and Southern blot.

2. Materials and Methods

2.1 Patients

A total of 15 ALS patients (13 sALS and 2 fALS) were recruited at the Institute of Neurology (University College London, London, UK). Clinical characteristics of ALS patients are reported in Table 1. A negative genetic test was obtained for major-genes screened (*SOD1*, *FUS* and *TARDBP*). All patients underwent Southern blot analysis or RP-PCR for *C9orf72*– repeat expansion investigation.

Table 1. Clinical data of ALS patients.

Gender, n (%)	Female	8 (53%)
	Male	7 (47,7%)
Family history, n (%)	fALS	2 (13,3%)
	sALS	13 (86%)
Site of onset	Upper limb	6 (40%)
	Lower limb	9 (60%)
Mean age of onset (y)		55

Key: fALS, familial Amyotrophic Lateral Sclerosis; sALS: sporadic Amyotrophic Lateral sclerosis.

2.2 Southern blot analysis and repeat-primed PCR

Southern hybridization was previously performed by combining genomic DNA digested with 2 frequently cutting restriction endonucleases, whose sites closely flanked the repeat region, and a (GGGGCC) probe targeting multiple sites within the expansion. Expansion size was estimated by interpolation of autoradiographs using a plot of log10 base pair number against migration distance. In this work, to identify more precisely the size of the expansions, the samples were investigated by OGM to clarify issues related to the detectability of the expansion as well as somatic mosaicism. Repeat primed -PCR was performed as described in the Chapter 2.

2.3 Optical Genome Mapping

Peripheral blood samples were immediately frozen at -80 °C after recovery according to the manufacturer's recommendations. Briefly, the blood leukocytes were quantified by a HemoCue WBC Analyzer (Life technologies). Ultra-high molecular weight (UHMW) DNA from approximately 1.5×10^6 cells was extracted using the "Blood and Cell Culture DNA Isolation Kit", and lysed by proteinase K; thereafter, the genomic DNA was washed, recovered, and successively quantified with Qubit Fluoremeter (Invitrogen).

2.4 DNA labeling and further processing for OGM

Based on quantitative results, samples were normalized to 750ng/µL and DNA was fluorescently labelled by using Bionano Prep DLS Labeling Kit (Bionano Genomics). Specifically, DNA was incubated at 37°C for 2 hours with Direct Labeling Enzyme 1, buffer, and DL-Green. Successively, the samples were treated with Proteinase k for 1 hour, and DNA was washed on membranes twice for 60 and 30 minutes. After mixing in Hulamixer for 1 hour and overnight at room temperature, labelled and stained DNA was quantified using the Qubit dsDNA HS assay kit. For optimum molecule assessment only DNA concentrations within the range 4-12 ng/µL was loaded onto Saphyr chips for linearization into and imaging in massively parallel nanochannel arrays. Each chip contains three separate nanochannel arrays for loading three unique samples (Figure 2).

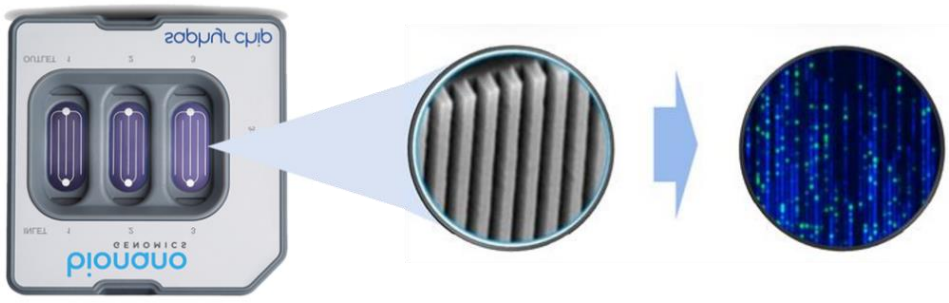


Figure 2. Bionano's chips are fabricated from standard semiconductor silicon wafers.

- Each chip contains up to 120,000 nanochannels, each 50 nanometers wide.
- DNA is deposited into the well on the chip and moved into the nanochannels via electrophoresis.
- DNA passes into the nanochannel, is linearized and then imaged.

2.5 Data Analysis and visualization

Optical mapping analysis was run on Bionano Access (version 1.7), an OGM-specific structural variant analysis software available as a standard web browser application, connects to bioinformatics servers running Bionano Solve (version 3.6). This is an auto-mated analytical pipeline for detecting genomic alterations, used for data processing. In briefly, single molecules were used to generate a de novo assembly of the genome and directly aligned to the reference genome (GRCh38) revealing SVs, CNVs and aneuploidy. All samples were analysed considering pre-analytical and analytical quality control (QC) metrics. At first, the presence of UHMW DNA (viscosity / clarity) was evaluated with a minimum DNA concentration of > 35 ng/µL required for subsequent labelling. Following, the label density of ~15/100 Kbp, an average of N50> 230 Kbp, a map frequency > 70% and an effective coverage of > 80 × for generating an assembly were evaluated among the analytical metrics de novo. All SVs are displayed on a CIRCOS graph, which shows specific chromosomal cytobands in the outer ring, abnormal fusions in the middle, CNVs in the inner ring, other SVs are shown in the next ring (Figure 3).

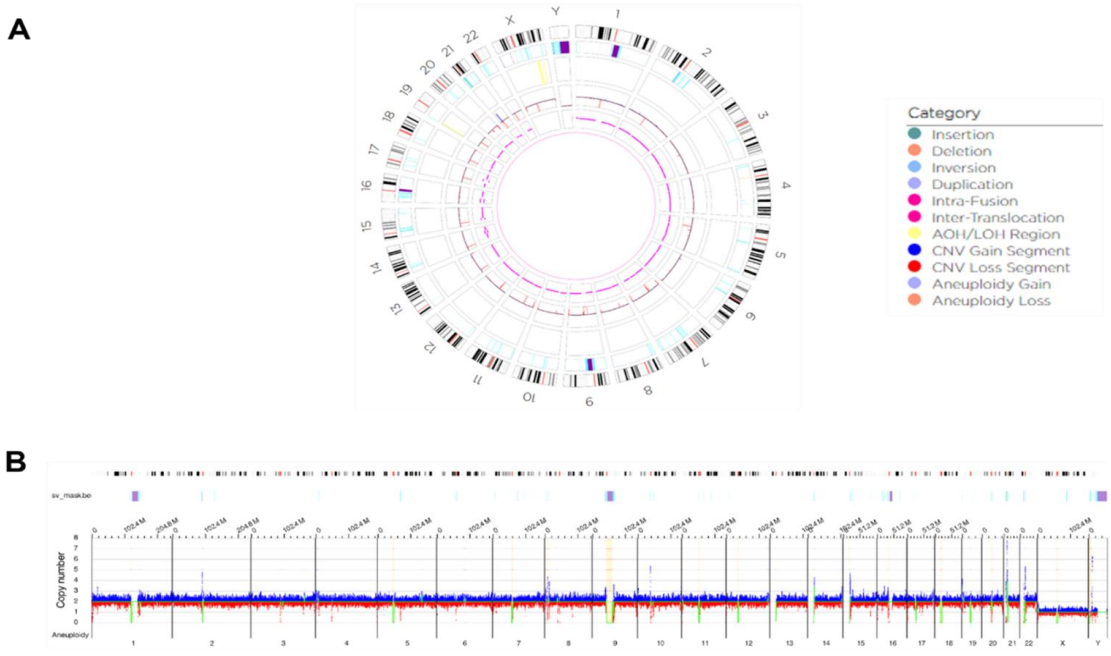


Figure 3. Output results from OGM. **A.** Circle plot showed the presence of different SVs (deletion, duplication, AOH/LOH Region, CNV Gain Segment) in the 22 pairs of autosomes equal in both sexes and a pair of heterochromosomes. **B.** Whole genome CNVs view: Y axis shows copy number range for each of the chromosomes – X axis. Molecules showing regions with increased copy number from the baseline are shown in blue and regions with decreased copy number are shown in red (Bionano Access).

3. Results

15 historical patients with a diagnosis of *C9orf72* based on RP-PCR and SB were analysed. Clinical features of patients are reported in Table 2.

Table 2 . Clinical features of fatients.

Clinical feature	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
fALS/sALS	*fALS*	*sALS*	*sALS*	*sALS*	*sALS*	*sALS*	*sALS*	*sALS*	*sALS*	*sALS*	*sALS*	*sALS*	*fALS*	*sALS*	*sALS*
Gender M/F	*M*	*F*	*M*	*F*	*F*	*F*	*M*	*M*	*M*	*F*	*M*	*M*	*F*	*F*	*F*
Age at onset (yrs)	*48*	*70*	*56*	*66*	*35*	*59*	*48*	*60*	*62*	*69*	*42*	*48*	*46*	*59*	*59*
Family history	*ALS/NMD*	-	-	-	-	-	-	-	-	*PD; NVD*	-	*Neurological disorder*	*ALS/MND/AD*	-	-
Site of Onset	*LL*	*UL*	*LL*	*UL*	*UL*	*LL*	*UL*	*LL*	*LL*	*UL*	*UL*	*UL*	*UL*	*UL*	*LL*
Phenotype	*spinal*	*bulbar*	*Na*	*bulbar*	*spinal*	*spinal*	*spinal*	*spinal*	*spinal*	*bulbar*	*bulbar*	*bulbar*	*bulbar*	*bulbar*	*spinal*

All the data have been referred from patient medical records. LL= for lower limbs and UL for upper limbs; NVD = neurovascular disease; MND = motor neuron disease; AD = Alzheimer disease.

3.1 Confirmation of hexanucleotide repeat expansion

The samples were optically mapped to clarify questions regarding variant detectability and to know the exact size of the otherwise undetectable *C9orf72* expansion. OGM identified repeat expansion in 14 out of 15 patients (repeat sizes ranging from 6 to 25 Kbp). The negative case was repeated with RP-PCR and confirmed to be a *C9orf72* negative case (Figure 4). The presence of the expansion in one of the 15 samples analysed and the comparison between OGM and SB is shown in Figure 5.

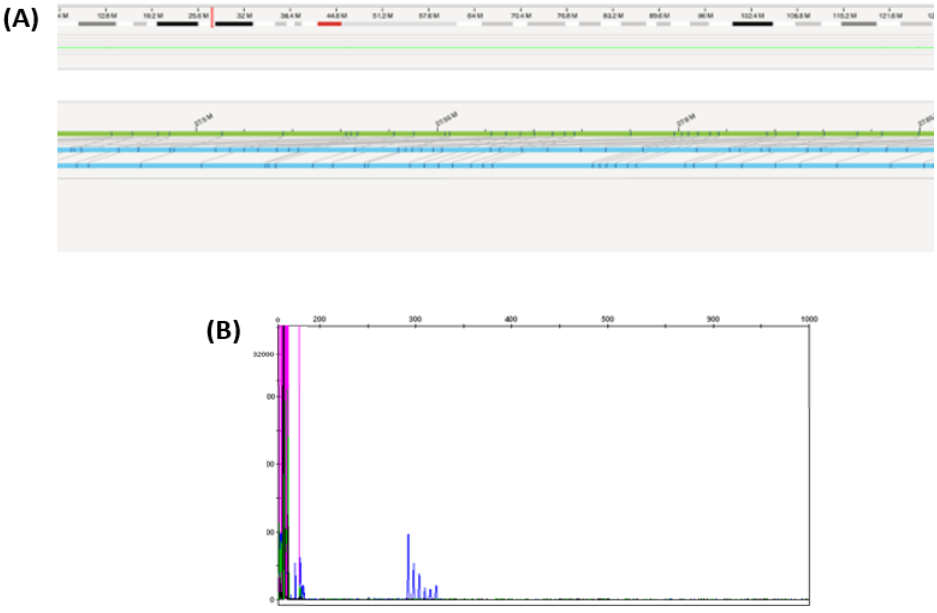


Figure 4. Grafhic OGM output (Bionano Access Software v1.7) and RP-PCR. *C9orf72* negative sample tested by OGM **(A)**; RP-PCR **(B)** detected a normal repeat allele.

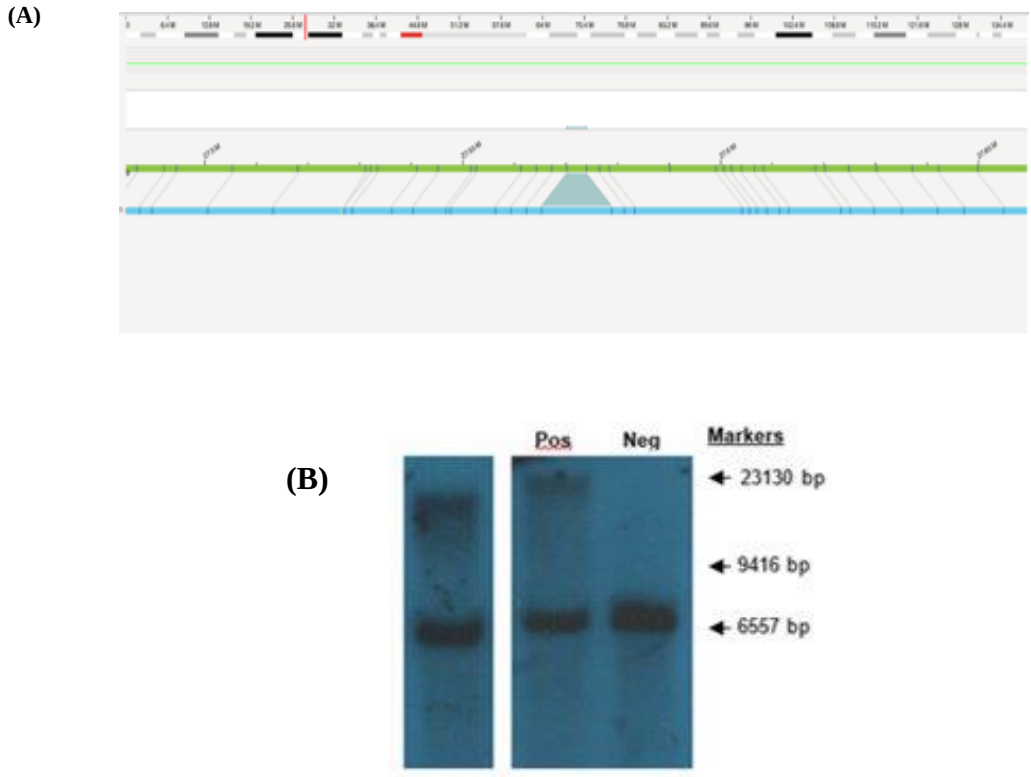


Figure 5. Grafhic OGM output (Bionano Access Software v1.7) and Southern blotting (SB). **(A)** The visualization confirms the presence of the expansion of 9851 bp. **(B)** The patient is characterized by a band within the region between 9 and 23kb.

3.2 OGM quality metrics

The total molecule length N50, which is the point of half mass of the molecule's distribution, was on average 243 Kbp (ranging from 198 Kbp to 369 Kbp). The average label density of the raw molecules was 15.21 labels per 100 Kbp (ranging from 14.9 to 17), leading to an average effective coverage of the reference genome of 117× (ranging from 63× to 268×) with an average diploid genome map length N50 of 73,7 Mbp (ranging from 53 Mbp to 97 Mbp).

3.3. Sizing and somatic instability

The somatic instability phenomena is now known for sporadic late-onset neuro-degenerative diseases (Parkinson's disease, Alzheimer's disease and ALS) (60). In ALS, especially, the repeat expansion in *C9orf72* is somatically unstable, when an expanded allele is already present (18) (61). The variable phenotype caused by *C9orf72* expansions could be due to different somatic expansion patterns. In fact, studies conducted on different tissues for the same patient have been reported and cases have been found with modest expansions in the leukocytes and extremes in the central nervous system. In view of this, OGM has the

potential to assess somatic instability in the tissue examined (Figure 6). Table 2 showed the range (median, min, max,) of individual molecule sizes.

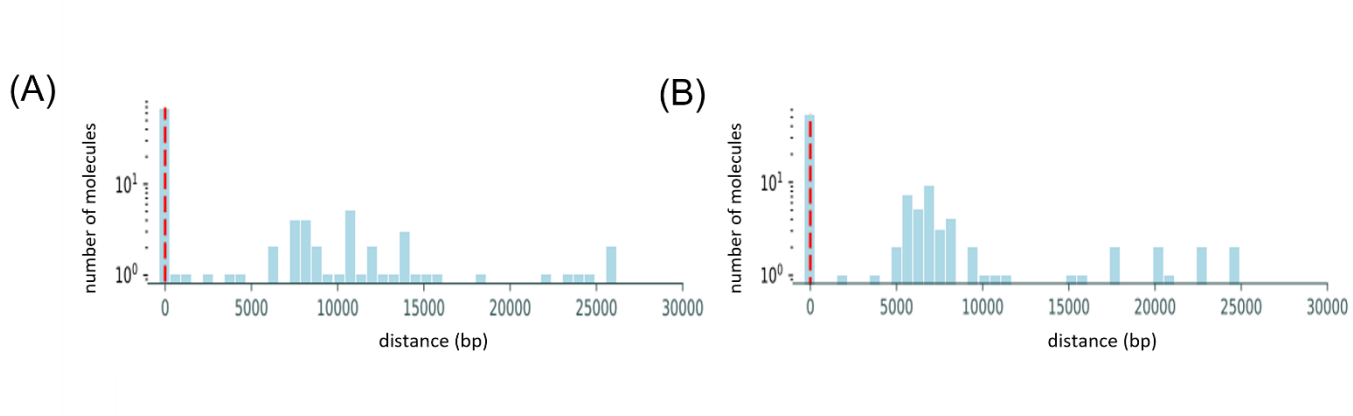


Figure 6. GGCCCC- hexanucleotide repeat size distribution. (A) Sample P9 and (B) Sample P12. The y-axis shows the number of molecules in the distribution, whereas the x-axis shows the length of the GGGGCC repeat expansion in base pairs (scale 0-30kb).

Table 2. Individual molecule sizes for each sample.

sample	mean	min	max
P1	n/a	n/a	n/a
P2	14507,0	2169,0	26844,0
P3	10633,0	3785,0	25389,0
P4	14573,0	2282,0	29158,0
P5	9048,0	4607,0	12547,0
P6	10692,0	1041,0	33180,0
P7	9596,0	1523,0	22912,0
P8	16450,0	2964,0	33949,0
P9	11923,0	1001,0	26020,0
P10	11331,0	4119,0	22987,0
P11	9217,0	2067,0	20345,0
P12	9878,0	1778,0	24903,0
P13	9550,0	4644,0	17384,0
P14	12601,0	3550,0	41810,0
P15	12264,0	2863,0	26240,0

Key: min, minimum; max, maximum; n/a, not available

4. Discussion

Genetic counselling for *C9orf72* is very complex due to the highly variable clinical presentation and technical difficulties in determining the size of the large GGGGCC repeat expansions. Obtaining the exact size of the expanded repeat is difficult due to several characteristics of the *C9orf72* repeat expansion such as somatic mosaicism and repeat

instability, highly repetitive flanking sequence, and 100% GC content (62). For several years, the size of the repeat expansions has been established by RP-PCR and SB, genetic tests currently in use for ALS despite not being able to accurately size repeat expansions. Furthermore, even the use of NGS-based methods does not allow for the correct sizing of long repeats (53). Recently, considerable advances have been carried out in molecular cytogenetics and sequencing technologies to detect genomic variants, i.e. repeat expansions and contractions, related to neurological disorders. To date, there are numerous studies that have used OGM data along with second or third- generation sequencing technologies to improve scaffolding and structural resolution. These include: the diagnosis of facioscapulohumeral muscular dystrophy (FSHD) (63), genetic diseases (64) and cancer and has been suggested for use as a cytogenomic tool for prenatal diagnostics (35, 65). Certain of the difficulty of attributing pathogenicity to the *C9orf72* expansion, due to the somatic instability of the hexanucleotide expansion, we decided to use OGM to correctly identify the expansion size in our cohort of ALS patients. Here, we evaluated for the first time the use of OGM in 14 out of 15 patients to significantly complement already existing *C9orf72* diagnostics. Specifically, we detected the sizing in a broad range of repeat expansion (from 6 Kbps to 25 Kbps and the equivalent repeat size). OGM has shown higher variability of repeat size for most cases and there is the potential to use this technology in future to better define somatic mosaicism. Increasing reading at 400X could provide additional benefits for this purpose. Overall, the advantages of using this technology include a short response time, approximately 10 hours for DNA isolation and library preparation, with approximatively 2 days of run time and 24 hours for automated data collection.

In conclusion, in this study, a non-sequencing-based technique was used for the first time in *C9orf72*-ALS patients with more than 1000 GGGGCC repeats. Overall, the technique allowed to simultaneous evaluate both the large repeat size of the *C9orf72* gene and somatic mosaicism, proving to be a robust alternative approach to the current diagnostic techniques of SB and RP-PCR and paving the way for the use of the optical genome mapping technique as a diagnostic validation method for this gene.

Chapter 6

DISCUSSION AND CONCLUSION

Discussion

In recent years, advances in genomic technology and gene discovery have greatly increased our understanding of the pathogenic mechanisms of multifactorial diseases, in which heterogeneous clinical states may reflect the combined effects of various genes and their interaction with environmental factors. In particular, the complexity of the genetic architecture of ALS, revealed by molecular diagnosis, with a monogenic component of rare high-penetrant variants, an oligogenic component of rare intermediate penetrant variants, and a multifactorial component of common risk variants may be a valuable tool for dissecting out ALS complex heterogeneity and for identifying new molecular mechanisms underlying the characteristic selective degeneration and death of motor neurons. However, many ALS-associated genes and related mechanisms have been identified but this information has not yet been fully translated into clinically useful knowledge. Currently, genetic tests for major ALS-related genes are needed to access therapeutic trials and to identify possible genetic factors that underlie disease development. Furthermore, the genetic complexity of ALS suggests that all forms of genetic variation, from the rare genetic variation of high to intermediate penetrance, as well as common risk variants with small effect sizes, need to be considered.

With these premises in mind, in *Chapter 2* of this thesis, we present the results of genetic analyses performed in a large cohort of Italian ALS patients by using Sanger sequencing. In this study, we focused on novel and already described variations in ALS-linked genes as *SOD1, C9orf72, TARDBP, FUS, ANG, VAPB, VCP, and ATXN*. Our data confirmed *C9orf72* repeat expansion as the most frequent alteration in Italian ALS patients (66). Moreover, in line with other population studies on ALS, the *SOD1* molecular investigation revealed a mutational frequency of 10% for fALS cases and 1.4% for sALS cases (66, 67, 68). Another important result was the statistically significant association between ALS patients carrying the *ATNX1* expanded polyQ and those with the pathological expansion in *C9orf72*, pointing to *ATXN1* as a potential disease modifier in *C9orf72* expansion carriers.

The oligogenic nature of the disease is discussed in *Chapter 3*. In this Chapter, we emphasized how the NGS analysis is important for the full knowledge of the genetic heterogeneity of ALS and for proving a comprehensive characterization of patient-specific molecular signatures that could potentially guide therapeutic decisions. Taking advantage of our custom-made gene panel, designed to uncover alterations in clinically relevant genes for ALS, we were able to reveal the individual oligogenic background underlying both sporadic

and familial *SOD1* p.D91A heterozygous and homozygous cases irrespective of their p.D91A mutant alleles. Homozygous and heterozygous phenotypes have a distinct clinical presentation and this aspect is very interesting, considering that differences in genotype-phenotype correlations may have significant therapeutic implications. Indeed, recruitment for antisense therapy has recently been discouraged in p.D91A-heterozygous carriers (*Chapter 3*).

An important advantage of NGS is well presented in a specific group of patients, where traditional tests have not been successful in making the diagnosis. In this context, in *Chapter 4*, we show the results of NGS-TS investigation performed in a cohort of ALS patients some of whom had previously tested negative (*Chapter 2*). With the aim of investigating the functional variations that could contribute significantly to the pathology, we focused our analysis on non-synonymous coding variants. We evaluated the oligogenic nature of ALS, highlighting that some patients (7/69, 10%) patients showing co-occurrence of at least 2 variants in 2 independent genes (Table 4, Chapter 4). Also, we obtained a significant number of VUS in different genes by expanding the number of target genes and, the co-occurrence of VUS variants had been assessed to be oligogenic inheritance by several studies (46, 69, 70). Our data confirmed the importance of evaluating ALS risk genes in disease aetiology and support the idea that although the role of small effect variants may appear of limited interest, their identification should not be neglected, as they could constitute a potential target for possible tailored therapeutic approaches in the context of the opportunity for personalized medicine.

Although Next-Generation Sequencing technology has revolutionized genetic diagnostics, there are some limitations due, for example, to the incomplete coverage of the regions, limited ability for detection of variation in repetitive elements, causal variants in cases of somatic mosaicism, structural variants and deep intronic variants (71, 72, 73). In particular, mosaicism is technically challenging to detect both in Sanger DNA sequencing and NGS platforms. In our context, the inability to detect the hexanucleotide expansion in *C9orf72*, the most frequent mutation in ALS patients, prompted us to evaluate the new cytogenomic approach, the Optical Genome Mapping platform (OGM). In *Chapter 5*, a cohort of *C9orf72* patients diagnosed with ALS underwent OGM analysis to validate the capability of this technology to accurately identify large insertions. We confirmed the repeat expansion presence in 14 of 15 patients, knowing the sizing in a broad range of repeat expansion (from 6 Kbp to 25Kbp and the equivalent repeat size), indicating the clinical utility of the method as an alternative to current diagnostic techniques (SB and RP-PCR). Furthermore, in this

study, we also focused on the mosaicism phenomena in *C9orf72* carriers, due to somatic instability of the number of repeats between tissues. Indeed, the repeat number varies in blood with age at the collection and over time in subsequent blood collections from *C9orf72* mutation carriers (74, 75). This mosaicism and instability may be part of the mechanisms underlying the clinical heterogeneity observed in patients, but there is no evidence for this. In our study, by using the OGM platform, we demonstrated the diagnostic utility of this approach and its potential future use to better delineate somatic mosaicism.

Overall, the data presented in this thesis provided evidence regarding the usefulness of different diagnostic approaches in identifying genetic drivers for a complex disease like ALS, characterized by different molecular mechanisms and extreme phenotypic heterogeneity.

Concluding remarks and future perspectives

Multigene interactions in ALS could explain the considerable phenotypic variability observed among individuals. This leads us to reconsider the traditional classification of ALS towards a new one in which genetic discoveries allow phenotypes to be grouped by discrete underlying causes. It will be essential to develop and implement new analytical strategies that can facilitate the discovery, and above all, the validation of genetic factors involved in the still-unknown pathogenesis of ALS. It is conceivable that in the not-too-distant future, high-throughput technologies will soon become cheaper than many of the specific genetic tests widely available today. Genetic testing will increasingly be used to guide accurate pathological diagnosis, opening the possibility of a molecular taxonomy for ALS and more precise stratification of patient groups in future drug trials.

References

1. Rosen DR, Siddique T, Patterson D, Figlewicz DA, Sapp P, Hentati A, et al. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. Nature. 1993;362(6415):59-62.
2. Brooks BR. El Escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. Subcommittee on Motor Neuron Diseases/Amyotrophic Lateral Sclerosis of the World Federation of Neurology Research Group on Neuromuscular Diseases and the El Escorial "Clinical limits of amyotrophic lateral sclerosis" workshop contributors. J Neurol Sci. 1994;124 Suppl:96-107.
3. Brooks BR, Miller RG, Swash M, Munsat TL. El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. Amyotroph Lateral Scler Other Motor Neuron Disord. 2000;1(5):293-9.
4. de Carvalho M, Dengler R, Eisen A, England JD, Kaji R, Kimura J, et al. The Awaji criteria for diagnosis of ALS. Muscle Nerve. 2011;44(3):456-7; author reply 7.
5. Carvalho MD, Swash M. Awaji diagnostic algorithm increases sensitivity of El Escorial criteria for ALS diagnosis. Amyotroph Lateral Scler. 2009;10(1):53-7.
6. Statland JM, Barohn RJ, McVey AL, Katz JS, Dimachkie MM. Patterns of Weakness, Classification of Motor Neuron Disease, and Clinical Diagnosis of Sporadic Amyotrophic Lateral Sclerosis. Neurol Clin. 2015;33(4):735-48.
7. Agosta F, Al-Chalabi A, Filippi M, Hardiman O, Kaji R, Meininger V, et al. The El Escorial criteria: strengths and weaknesses. Amyotroph Lateral Scler Frontotemporal Degener. 2015;16(1-2):1-7.
8. Cedarbaum JM, Stambler N, Malta E, Fuller C, Hilt D, Thurmond B, et al. The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. BDNF ALS Study Group (Phase III). J Neurol Sci. 1999;169(1-2):13-21.
9. Ingre C, Roos PM, Piehl F, Kamel F, Fang F. Risk factors for amyotrophic lateral sclerosis. Clinical epidemiology. 2015;7:181-93.
10. Zhong Y, Xu F, Wu J, Schubert J, Li MM. Application of Next Generation Sequencing in Laboratory Medicine. Ann Lab Med. 2021;41(1):25-43.
11. Oskarsson B, Gendron TF, Staff NP. Amyotrophic Lateral Sclerosis: An Update for 2018. Mayo Clin Proc. 2018;93(11):1617-28.
12. Fang T, Je G, Pacut P, Keyhanian K, Gao J, Ghasemi M. Gene Therapy in Amyotrophic Lateral Sclerosis. Cells. 2022;11(13).
13. Perrone B, Conforti FL. Common mutations of interest in the diagnosis of amyotrophic lateral sclerosis: how common are common mutations in ALS genes? Expert Rev Mol Diagn. 2020;20(7):703-14.
14. van Blitterswijk M, DeJesus-Hernandez M, Rademakers R. How do C9ORF72 repeat expansions cause amyotrophic lateral sclerosis and frontotemporal dementia: can we learn from other noncoding repeat expansion disorders? Curr Opin Neurol. 2012;25(6):689-700.
15. Berdyński M, Miszta P, Safranow K, Andersen PM, Morita M, Filipek S, et al. SOD1 mutations associated with amyotrophic lateral sclerosis analysis of variant severity. Scientific Reports. 2022;12(1):103.
16. Gijselinck I, Van Mossevelde S, van der Zee J, Sieben A, Engelborghs S, De Bleecker J, et al. The C9orf72 repeat size correlates with onset age of disease, DNA methylation and transcriptional downregulation of the promoter. Mol Psychiatry. 2016;21(8):1112-24.
17. Waite AJ, Bäumer D, East S, Neal J, Morris HR, Ansorge O, et al. Reduced C9orf72 protein levels in frontal cortex of amyotrophic lateral sclerosis and frontotemporal

degeneration brain with the C9ORF72 hexanucleotide repeat expansion. Neurobiol Aging. 2014;35(7):1779.e5-.e13.
18. Nordin A, Akimoto C, Wuolikainen A, Alstermark H, Jonsson P, Birve A, et al. Extensive size variability of the GGGGCC expansion in C9orf72 in both neuronal and non-neuronal tissues in 18 patients with ALS or FTD. Hum Mol Genet. 2015;24(11):3133-42.
19. Shaw PJ. Molecular and cellular pathways of neurodegeneration in motor neurone disease. J Neurol Neurosurg Psychiatry. 2005;76(8):1046-57.
20. Pramatarova A, Laganière J, Roussel J, Brisebois K, Rouleau GA. Neuron-specific expression of mutant superoxide dismutase 1 in transgenic mice does not lead to motor impairment. J Neurosci. 2001;21(10):3369-74.
21. Philips T, Rothstein JD. Rodent Models of Amyotrophic Lateral Sclerosis. Curr Protoc Pharmacol. 2015;69:5.67.1-5..21.
22. Kaur SJ, McKeown SR, Rashid S. Mutant SOD1 mediated pathogenesis of Amyotrophic Lateral Sclerosis. Gene. 2016;577(2):109-18.
23. Kumar V, Hasan GM, Hassan MI. Unraveling the Role of RNA Mediated Toxicity of C9orf72 Repeats in C9-FTD/ALS. Front Neurosci. 2017;11:711.
24. Volk AE, Weishaupt JH, Andersen PM, Ludolph AC, Kubisch C. Current knowledge and recent insights into the genetic basis of amyotrophic lateral sclerosis. Med Genet. 2018;30(2):252-8.
25. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977;74(12):5463-7.
26. Lamp M, Origone P, Geroldi A, Verdiani S, Gotta F, Caponnetto C, et al. Twenty years of molecular analyses in amyotrophic lateral sclerosis: genetic landscape of Italian patients. Neurobiol Aging. 2018;66:179.e5-.e16.
27. Cady J, Allred P, Bali T, Pestronk A, Goate A, Miller TM, et al. Amyotrophic lateral sclerosis onset is influenced by the burden of rare variants in known amyotrophic lateral sclerosis genes. Ann Neurol. 2015;77(1):100-13.
28. Fumagalli M. Assessing the effect of sequencing depth and sample size in population genetics inferences. PLoS One. 2013;8(11):e79667.
29. Yohe S, Thyagarajan B. Review of Clinical Next-Generation Sequencing. Arch Pathol Lab Med. 2017;141(11):1544-57.
30. Vinkšel M, Writzl K, Maver A, Peterlin B. Improving diagnostics of rare genetic diseases with NGS approaches. J Community Genet. 2021;12(2):247-56.
31. Kenna KP, McLaughlin RL, Byrne S, Elamin M, Heverin M, Kenny EM, et al. Delineating the genetic heterogeneity of ALS using targeted high-throughput sequencing. J Med Genet. 2013;50(11):776-83.
32. Hagemann IS. Chapter 1 - Overview of Technical Aspects and Chemistries of Next-Generation Sequencing. In: Kulkarni S, Pfeifer J, editors. Clinical Genomics. Boston: Academic Press; 2015. p. 3-19.
33. Gijselinck I, Cruts M, Van Broeckhoven C. The Genetics of C9orf72 Expansions. Cold Spring Harbor perspectives in medicine. 2018;8(4).
34. Akimoto C, Volk AE, van Blitterswijk M, Van den Broeck M, Leblond CS, Lumbroso S, et al. A blinded international study on the reliability of genetic testing for GGGGCC-repeat expansions in C9orf72 reveals marked differences in results among 14 laboratories. J Med Genet. 2014;51(6):419-24.
35. Sahajpal NS, Barseghyan H, Kolhe R, Hastie A, Chaubey A. Optical Genome Mapping as a Next-Generation Cytogenomic Tool for Detection of Structural and Copy Number Variations for Prenatal Genomic Analyses. Genes (Basel). 2021;12(3).
36. Marangi G, Traynor BJ. Genetic causes of amyotrophic lateral sclerosis: new genetic analysis methodologies entailing new opportunities and challenges. Brain Res. 2015;1607:75-93.

37. Pecoraro V, Mandrioli J, Carone C, Chiò A, Traynor BJ, Trenti T. The NGS technology for the identification of genes associated with the ALS. A systematic review. Eur J Clin Invest. 2020;50(5):e13228.
38. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.
39. van der Spek RAA, van Rheenen W, Pulit SL, Kenna KP, van den Berg LH, Veldink JH. The project MinE databrowser: bringing large-scale whole-genome sequencing in ALS to researchers and the public. Amyotroph Lateral Scler Frontotemporal Degener. 2019;20(5-6):432-40.
40. Ruffo P, Perrone B, Conforti FL. SOD-1 Variants in Amyotrophic Lateral Sclerosis: Systematic Re-Evaluation According to ACMG-AMP Guidelines. Genes (Basel). 2022;13(3).
41. Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat Genet. 2003;33 Suppl:228-37.
42. Pensato V, Magri S, Bella ED, Tannorella P, Bersano E, Sorarù G, et al. Sorting Rare ALS Genetic Variants by Targeted Re-Sequencing Panel in Italian Patients: OPTN, VCP, and SQSTM1 Variants Account for 3% of Rare Genetic Forms. J Clin Med. 2020;9(2).
43. Borg R, Farrugia Wismayer M, Bonavia K, Farrugia Wismayer A, Vella M, van Vugt J, et al. Genetic analysis of ALS cases in the isolated island population of Malta. Eur J Hum Genet. 2021;29(4):604-14.
44. Gellera C. Genetics of ALS in Italian families. Amyotroph Lateral Scler Other Motor Neuron Disord. 2001;2 Suppl 1:S43-6.
45. Scarlino S, Domi T, Pozzi L, Romano A, Pipitone GB, Falzone YM, et al. Burden of Rare Variants in ALS and Axonal Hereditary Neuropathy Genes Influence Survival in ALS: Insights from a Next Generation Sequencing Study of an Italian ALS Cohort. Int J Mol Sci. 2020;21(9).
46. Pang SY, Hsu JS, Teo KC, Li Y, Kung MHW, Cheah KSE, et al. Burden of rare variants in ALS genes influences survival in familial and sporadic ALS. Neurobiol Aging. 2017;58:238.e9-.e15.
47. Nguyen HP, Van Broeckhoven C, van der Zee J. ALS Genes in the Genomic Era and their Implications for FTD. Trends Genet. 2018;34(6):404-23.
48. Sauna ZE, Kimchi-Sarfaty C. Understanding the contribution of synonymous mutations to human disease. Nature reviews Genetics. 2011;12(10):683-91.
49. Vihinen M. Individual Genetic Heterogeneity. Genes (Basel). 2022;13(9).
50. Paulson H. Repeat expansion diseases. Handb Clin Neurol. 2018;147:105-23.
51. Morgan AP, Gatti DM, Najarian ML, Keane TM, Galante RJ, Pack AI, et al. Structural Variation Shapes the Landscape of Recombination in Mouse. Genetics. 2017;206(2):603-19.
52. Theunissen F, Flynn LL, Anderton RS, Mastaglia F, Pytte J, Jiang L, et al. Structural Variants May Be a Source of Missing Heritability in sALS. Front Neurosci. 2020;14:47.
53. Ibañez K, Polke J, Hagelstrom RT, Dolzhenko E, Pasko D, Thomas ERA, et al. Whole genome sequencing for the diagnosis of neurological repeat expansion disorders in the UK: a retrospective diagnostic accuracy and prospective clinical validation study. Lancet Neurol. 2022;21(3):234-45.
54. Roggenbuck J. C9orf72 and the Care of the Patient With ALS or FTD: Progress and Recommendations After 10 Years. Neurology Genetics. 2021;7(1):e542.
55. Cammack AJ, Atassi N, Hyman T, van den Berg LH, Harms M, Baloh RH, et al. Prospective natural history study of C9orf72 ALS clinical characteristics and biomarkers. Neurology. 2019;93(17):e1605-e17.

56. Costantino I, Nicodemus J, Chun J. Genomic Mosaicism Formed by Somatic Variation in the Aging and Diseased Brain. Genes (Basel). 2021;12(7).
57. Russ J, Liu EY, Wu K, Neal D, Suh E, Irwin DJ, et al. Hypermethylation of repeat expanded C9orf72 is a clinical and molecular disease modifier. Acta neuropathologica. 2015;129(1):39-52.
58. Bocklandt S, Hastie A, Cao H. Bionano Genome Mapping: High-Throughput, Ultra-Long Molecule Genome Analysis System for Precision Genome Assembly and Haploid-Resolved Structural Variation Discovery. Advances in experimental medicine and biology. 2019;1129:97-118.
59. Yuan Y, Chung CY, Chan TF. Advances in optical mapping for genomic research. Computational and structural biotechnology journal. 2020;18:2051-62.
60. Leija-Salazar M, Piette C, Proukakis C. Review: Somatic mutations in neurodegeneration. Neuropathol Appl Neurobiol. 2018;44(3):267-85.
61. van Blitterswijk M, DeJesus-Hernandez M, Niemantsverdriet E, Murray ME, Heckman MG, Diehl NN, et al. Association between repeat sizes and clinical and pathological characteristics in carriers of C9ORF72 repeat expansions (Xpansize-72): a cross-sectional cohort study. Lancet Neurol. 2013;12(10):978-88.
62. Cleary EM, Pal S, Azam T, Moore DJ, Swingler R, Gorrie G, et al. Improved PCR based methods for detecting C9orf72 hexanucleotide repeat expansions. Molecular and cellular probes. 2016;30(4):218-24.
63. Koppikar P, Shenoy S, Guruju N, Hegde M. Testing for Facioscapulohumeral Muscular Dystrophy with Optical Genome Mapping. Curr Protoc. 2023;3(1):e629.
64. Dremsek P, Schwarz T, Weil B, Malashka A, Laccone F, Neesen J. Optical Genome Mapping in Routine Human Genetic Diagnostics-Its Advantages and Limitations. Genes (Basel). 2021;12(12).
65. Dai P, Zhu X, Pei Y, Chen P, Li J, Gao Z, et al. Evaluation of optical genome mapping for detecting chromosomal translocation in clinical cytogenetics. Mol Genet Genomic Med. 2022;10(6):e1936.
66. Chiò A, Restagno G, Brunetti M, Ossola I, Calvo A, Canosa A, et al. ALS/FTD phenotype in two Sardinian families carrying both C9ORF72 and TARDBP mutations. Journal of neurology, neurosurgery, and psychiatry. 2012;83(7):730-3.
67. Conte A, Lattante S, Luigetti M, Del Grande A, Romano A, Marcaccio A, et al. Classification of familial amyotrophic lateral sclerosis by family history: effects on frequency of genes mutation. Journal of neurology, neurosurgery, and psychiatry. 2012;83(12):1201-3.
68. Lattante S, Conte A, Zollino M, Luigetti M, Del Grande A, Marangi G, et al. Contribution of major amyotrophic lateral sclerosis genes to the etiology of sporadic disease. Neurology. 2012;79(1):66-72.
69. Krüger S, Battke F, Sprecher A, Munz M, Synofzik M, Schöls L, et al. Rare Variants in Neurodegeneration Associated Genes Revealed by Targeted Panel Sequencing in a German ALS Cohort. Front Mol Neurosci. 2016;9:92.
70. Nakamura R, Sone J, Atsuta N, Tohnai G, Watanabe H, Yokoi D, et al. Next-generation sequencing of 28 ALS-related genes in a Japanese ALS cohort. Neurobiol Aging. 2016;39:219.e1-8.
71. Xue Y, Ankala A, Wilcox WR, Hegde MR. Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: single-gene, gene panel, or exome/genome sequencing. Genet Med. 2015;17(6):444-51.
72. Lalonde E, Ebrahimzadeh J, Rafferty K, Richards-Yutz J, Grant R, Toorens E, et al. Molecular diagnosis of somatic overgrowth conditions: A single-center experience. Molecular genetics & genomic medicine. 2019;7(3):e536.

73. Caspar SM, Dubacher N, Kopps AM, Meienberg J, Henggeler C, Matyas G. Clinical sequencing: From raw data to diagnosis with lifetime value. Clinical genetics. 2018;93(3):508-19.
74. Suh E, Lee EB, Neal D, Wood EM, Toledo JB, Rennert L, et al. Semi-automated quantification of C9orf72 expansion size reveals inverse correlation between hexanucleotide repeat number and disease duration in frontotemporal degeneration. Acta Neuropathol. 2015;130(3):363-72.
75. Fournier C, Barbier M, Camuzat A, Anquetil V, Lattante S, Clot F, et al. Relations between C9orf72 expansion size in blood, age at onset, age at collection and transmission across generations in patients and presymptomatic carriers. Neurobiol Aging. 2019;74:234.e1-.e8.

ACKNOWLEDGEMENTS

First and foremost, i am grateful to my supervisor, Prof. Francesca Luisa Conforti, introducing me to the wonderful world of genetics and also for motivating, encouraging, inspiring me with her constancy guidance.

I would like to thank all of the staff of Centro Sanitario for being like family to me during these four years. Mostly thanks to my amazing colleagues and all the professors, for teaching me quite everything I actually know and for them support and patience.

I wish to express my gratitude to the Department of Neuromuscular Diseases, UCL Institute of Neurology (London, UK) and to Professor Henry Houlden for inclusion in different projects and for the opportunity to spend a ten-month internship under his direct supervision. Special thanks also go to all the laboratory colleagues for their friendship and teaching.

I gratefully acknowledge the patients and their family for their contribution and collaboration.

Finally, my deepest gratitude goes to my family, and friends for their love and support. Thanks to my father and all my family for their understanding and for always having my back.
Thanks to my mom for believing in me before I learned how to do it.

Without you and your support it would have been more difficult.
Thank you!